

Enhanced Constrained Local Models (CLM) for Facial Feature Detection

Ayah Alsarayreh* and Fatma Susilawati Mohamad

*Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin,
Besut, Campus, 22200 Besut, Terengganu, Malaysia.*

ORCID: 0000-0002-6390-0220 (Ayah Alsarayreh)

Abstract

Face land-marking, defined as detecting and locating certain distinctive points on the face, is a key goal between a big amounts of ventures that result in face-preparing activities ranging from biometric recognition to mental state comprehension. Despite its applied simplicity, this problem has turned out to be highly tested due to inherent facial variability as well as the enormous amount of puzzling variables such as pose, expression, illumination and occlusions. An integrated fitting model is developed that improve the power of Constrained Local Models (CLM), and obtained an innovative result on the problem of feature detection. The experiment results have indicated that the proposed model is effective in facial features extraction and get result better than Constrained Local Models (CLM).

Keywords: Feature Extraction; Constrained Local Models (CLM); face landmark.

1 INTRODUCTION

The Face feature is defined as detection of certain characteristic points on the face. The points are used to represent crucial information needed to classify an individual; this is achieved by building a model [1]. The model is represented in a predefined number of landmark points, which depends on the complexity of the object's shape and desired level of detailed descriptions of where it is needed. Facial landmarks from the literature are also known as facial feature points, anchor points homologous points and key points. There are many techniques consider the face image and the collection of facial feature points as a whole shape[2]. These techniques employ prior knowledge (from labeled training images) about the face position, and restrict the landmark search by heuristic rules that involve areas, angles, and distances. Consequently the techniques fit the appropriate shape to an unknown face [3]. Examples of such techniques are Active Shape Model (ASM) [4], Active Appearance Model (AAM) [5] and Constrained Local Models (CLM) [6].

In[7], the comparison was made between the AAM and ASM models in terms of speed and achievement of determining the location of the feature point and they found ASM is better than AAM. The AAM uses a model of the appearance of the entire facial region Whereas ASM constrain to models the image texture nearest to each landmark point. ASM is quiker and has an extensive search domain than AAM, whilst AAM gives a better identification to the texture [7]. AAMs can find an

accurate place of the head. Hence, give accurate pose estimation and this consider as advantage of AAMs. However, the accuracy of AAMs depends mostly on the used training set. So, in order to enable AAM to work on different faces it should provide a number of faces with different poses in the training set and this is the main limitation of AAM.

CLM model depend on the express nearby facial appearance and express global facial shape pattern. The regression built strategies utilization holistic or nearby appearance majority of the data furthermore they might implant those global facial state patterns implicitly to joint landmark discovery. The CLM model has brought remarkable developments as far back as a few years. [8] Introduced as a local detector the Deep Constrained Local Model (DCLM) method and the novel Dense Projection Network (DPN). DPN will be a deep neural network consisting of two vital layers: a thick aggravator layer of projection format. For the format Projection layer, patches from claiming facial area should be mapped for a higher dimensional space that also allows the pose to be caught faultlessly by revolutionary varieties. A group of masters will be mimicked in a thick aggravator layer inside one network should aggravate the additional strong point of interest restriction assignment.

[9] uses simple Gabor Filter and Discrete Cosine Transform (DCT) to develop a feature extraction technique. In Gabor Filter method, the images were transformed into Gabor magnitude, then smoothed with the image smoother before the transformation was applied. Using coefficient correlation, the impact of Gabor Filter on the DCT coefficients was assessed. [10] introduced deep learning strategies based on Convolutional Neural Network (CNN) for facial traits elicitation. unlike the preceding models which were developed using CNN, they used traits map gained in distinctive layers. Recently, many techniques have been put to extract facial points or features from the images for enhancing feature extraction [11-13]. [14] developed inadequate reconstruction technique to those face alignment issues. As opposed to a direct regression in the middle of those characteristics and shape space, the idea for shape augment reconstruction is presented. Moreover, a situated of coupled over complete studies termed those shape increase and the local appearance are figured out in a regressive way on select deep characteristics.

In this article, we proposed an improved CLM That gathers whole quality of both depth and strength data to identify and monitor facial expressions in pictures. The depth data helps us to minimize the impact of lighting conditions. Moreover, it let

us to decrease the influence of the problem of aperture, which emerge on account of the strong patch response over the edges but not across them. On the other hand, our method has the option using depth reaction CLM only when there is no strength signal or when lighting conditions are inadequate.

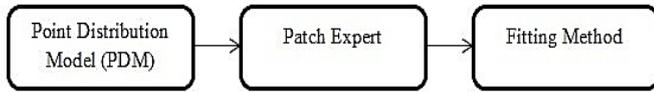


Figure 1: The Enhanced CLM Basic Components.

2 FEATURE EXTRACTION METHOD

The methods and techniques used in this research are explained in this section. Section one provides a description of the technique proposed. Section two describes the used database, while section three shows PDM. Section four describes patch experts, and section five summarizes fitting method.

2.1 Overview of the proposed approach

Our paper's key contribution is the A-CLM, a version of the Constrained Local Model that integrates strength and depth knowledge to map facial feature points. Our strategy utilizes the structure of the Constrained Local Model (CLM), so this model viewed here with its information. As showed in figure 1. CLM has three basic factors. Namely, a point distribution model (PDM), patch expert, and the technique of fitting.

2.2 Face Database

In this paper, we used LFPW database [15], which is a facial picture database connected with detailed data in regards to the gender and age for every picture inside the database. LFPW picture database utilized to check the efficiency and accuracy for the proposed approach. 100 face pictures are decided to be test group, all pictures contain 68 landmarks. The amount of looking locations has the practically grade impact on the position result. The achievement for the proposed model and the current models are compared.

2.3 Point Distribution Model (PDM)

Our improved CLM model can be defined with variables $p = [s, R, q, t]$ which can be diverted to obtain specific model. s denote to the scale factor, R stand to object rotation, t is 2D translation, and a vector representing non-rigid form q variety. The configuration used in A-CLM in point distribution (PDM) is:

$$x_i = sR_{2D}(\bar{x} + \Phi_i q) + t \quad (1)$$

Here $\bar{x}_i = [\bar{x}_l, \bar{y}_l, \bar{z}_l]$ represents the mean quality of the i^{th} characteristic, Φ_i is the $3 \times m$ principal component matrix, furthermore q represents the m dimensional vector for variables monitoring the non-rigid shape.

Instead of perspective projection, this method uses scaled orthographic camera model, which is consider as a weak-perspective, that is because the linearity gives more accurate optimization. For a weak-perspective model, the scaling factor s may acting as the opposite of the average width, while the translation vector t consider the center point. This is a reasonable approximation regarding the distance to camera because of the comparatively limited variations of depth over the face plane.

In the proposed model, the following formula is devoted to calculate the maximum a posteriori probability (MAP) of the face model parameters p :

$$p(P\{\{l_i = 1\}_{i=1}^n, I\}) \propto p(P) \prod_{i=1}^n p(l_i = 1 | x_i, D) \quad (2)$$

Here l_i takes the values ± 1 and it represents the discrete arbitrary variable demonstrating if the i^{th} characteristic point will be adjusted or misaligned, $p(p)$ is the former likelihood of the model parameters p , furthermore $\prod_{i=1}^n p(l_i = 1 | x_i, D)$ will be those joint likelihood of the characteristic focuses x continuously adjusted at a specific point x_i , provided for an force intensity i . Patch experts are utilized to calculate $p(l_i = 1 | x_i, D)$, which is the likelihood that a function is aligned at x_i .

2.4 Patch experts

We determine how the new feature points are matched depending on the regional support area by using local patch experts who calculate the likelihood of alignment $p(l_i = 1 | x_i, I)$. We use equation (3) as a probabilistic patch expert; the mean value of two logistic regressors given by equations 4 and 5.

$$p(l_i | x_i, I, Z) = 0.5(p(l_i | x_i, I) + p(l_i | x_i, Z)) \quad (3)$$

$$p(l_i | x_i, I) = \frac{1}{e^{dC_{I,i}(x_i; I)} + c} \quad (4)$$

$$p(l_i | x_i, I) = \frac{1}{e^{dC_{z,i}(x_i; z)} + c} \quad (5)$$

Where $C(Z, i)$ is the result of strength and $C(I, i)$ are the results of the depth patch classifiers, c is the logistic regressor intercept for the i^{th} traits, and d is the regression coefficient. We use the support vector regression in short (SVR) that is because it has easy calculations, and its effective implementation on convolution images. The SVR help us to find the hyper plane "close" to most pieces of the training landmarks as potential [16]. Let N training points $\{(x_i, y_i) : 1 \leq i \leq N, x_i \in R^n, y_i \in R\}$. We have to build certain hyper planes and qualities from w and b . The hyper plane w can be selected to a low standard when reducing the aggregate distances from the hyper plane training points. Essentially, use Vapnik's π insensitive loss function:

$$|y_i - (wx_i + b)|_\epsilon = \begin{cases} 0 & |y_i - (wx_i + b)| \leq \epsilon \\ |y_i - (wx_i + b)| & \text{otherwise} \end{cases} \quad (6)$$

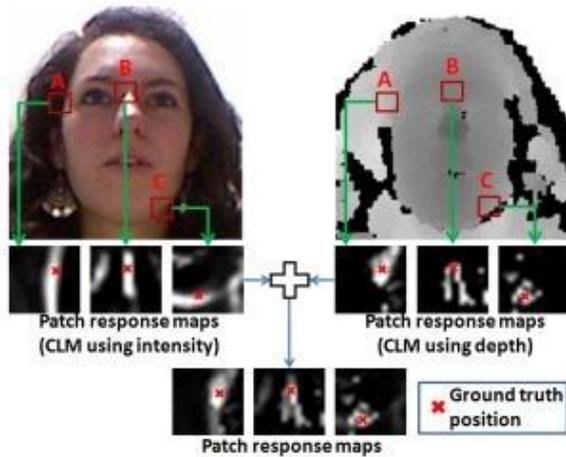


Figure 2: Patch expert response maps.

Figure 2 shows three patch expert response maps: face contour, nose ridge, and chin part. Logistic response maps of regressors that use strength which has strong responses around the edges, making it difficult to locate the precise location of the element. Our solution mitigates the aperture issue by incorporating response maps from both the strength and depth images. The weight of may be chosen via user, and the trade-off between the discoveries of a hyper plane with a large regression execution is controlled by the regularization c . The QP problem related to SVR can be presented according to the following:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} w^T w + c \left(\sum_{i=1}^L \xi_i + \sum_{i=1}^L \xi_i^* \right) \quad (7)$$

such that

$$w^T \phi(x_i) + b - y_i \leq \epsilon + \xi_i$$

and

$$y_i - w^T \phi(x_i) - b \leq \epsilon + \xi_i^*$$

Example pictures of pressure, depth and combined response maps (figure 2 shows the examination of the patch expert method near the pixels of an initial estimate); A major issue faced by CLM is the aperture dilemma, where confidence in identification around the edge is higher than along it, which if the strength response maps is clear for the nose ridge and face outline. having the depth information helps handle this problem, because the strong edges in general is not correlate in both images, allowing more disambiguation of points around strong edges.

2.5 Fitting

We are going to apply a standard two-step CLM fitting technique [6, 17, 18] carry out entire local search around the current feature point estimation gives a response map around each feature point, and then iteratively the parameters of the model will update to optimize Eq(2) to reach the metric convergency. For the fitting we utilize Non-Uniform

Regularized Landmark Mean Shift (NU-RLMS) [2]. Provided for a starting CLM parameter evaluates p , NU-RLMS iteratively finds an overhaul parameter P , We assume that the non-rigid shape parameters q differ by Gaussian distribution with the variance of the i 'th parameter corresponding to the non-rigid deformation mode's own value; the rigid parameters $s, R,$ and t obey a non-informative uniform distribution. Treating the positions of the actual landmarks as unknown variables will marginalize them out of the probability of aligning the landmarks:

$$p(l_i|x_i, I, Z) = \sum_{y_i \in \Psi_i} p(l_i|x_i, I, Z) p(y_i|x_i) \quad (8)$$

Where $p(y_i|x_i) = N(y_i, x_i, pI)$, with p signify the contrast of noise in landmark locations resulting from PCA truncation in PDM construction, and ψ_i signify all integer locations within the patch area. By substituting Equation (8) into Equation (2) we get:

$$p(P) \prod_{i=1}^n \sum_{y_i \in \psi_i} p(l_i|x_i, I, Z) N(y_i; x_i, pI) \quad (9)$$

The term of MAP in Equation 9 can be maximized by Expectation Maximization. The discrepancy between the available implementation and the original algorithm is attributed, in addition to frontal, to the use of qualified patches using profile face pictures. This conduct three groups of classifiers (frontal, left, right), with no answer functions for the occluded landmarks for the left and right sets. This allows us to work with self occlusion since the invisible points for the fitting procedure are not evaluated.

3 EXPERIMENTAL RESULTS

In this section, results of the selected algorithms are displayed. Section 1 summarizes the accuracy and the proposed model while section 2 demonstrates the efficiency of the proposed model.

3.1 Experimental Setup

The proposed model has been used to choose various features beside a hierarchical structure. The constructed structure may take hierarchical and unique features. The proposed model showed very optimistic results in many applications. This approach find that the proposed model is great toward characterizing the high-level facial features. In this section, the proposed model is utilized for facial features extraction.

The performance metrics are used to determine the best facial feature extraction model based on results. The point to point error ($pt-pt$) is used to assess the accuracy of the fitting. The $pt-pt$ error is determined by the distance between the estimated landmarks and the hand labels.

$$E(S, S') = \frac{1}{n} \sum_{i=1}^n \sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2} \quad (10)$$

Where S is the shape ground truth and where S^0 is the shape estimated. If the $pt - pt$ error is within three pixels, the fitting is considered successful. Table 1 shows the average of point to point error for 68 landmarks of testing images from LFPW database using the proposed model, and CLM model. Which show that the proposed approach produces the smallest fitting error.

Table 1: The average of point-to-point error for LFPW Database.

Model	point to point error	No of test sample
CLM	0.1040	100
A-CLM	0.1012	100

Table 1 compare the point to point error for A-CLM model, and CLM, each model performed separately and applied for 100 image from LFPW database, the result shows that A-CLM error rate is 0.1012, and CLM model error rate is 0.1040. From this comparison we can see that the A-CLM error rate is the lowest compared with ASM and AAM.

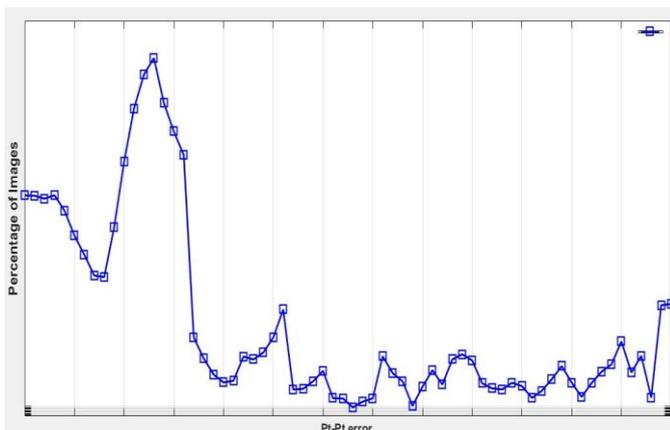


Figure 3: Patch expert response maps.

Figure3 shows point-to-point errors (the distance between model points and the corresponding points marked in images) in LFPW database. The vertical axis is the point-to-point error, and the image number is the Horizontal axis, the line is the point-to-point error of all feature points. It can be seen that the proposed approach (A-CLM) line is more accurate under all conditions than the CLM model except for some cases.

3.2 Efficiency

Table 2 gives the average landmark detection time for images from LFPW database using CLM, and A-CLM models. The average detection time is the taken for the model to detect face and fit the model on the facial features. In the table, the result is given in respect of two different models CLM, and ACLM.

Table 2: Comparison of Time measurements for Models using LFPW database.

Model	Avg.time	No of test sample
CLM	138.36826 seconds	100
A-CLM	242.95939 seconds	100

Table 2 compare the time measurements for CLM model, and A-CLM, each model performed separately and applied for 100 image from LFPW database, The result shows that CLM model takes 138.36826 seconds to complete full process, and the proposed approach A-CLM takes 232.86938 seconds.

4 CONCLUSION

This paper proposes an improved facial feature extraction model using the CLM model through using the strong points of the model. The research was conducted using LFPW database. The enhanced section aimed at making the new model work in diversified areas and minimizing complexity of systems in integrated models. The proposed model is proved to have shown the best results, and thereby compensates shape and texture variations.

ACKNOWLEDGMENTS

This work is supported by the Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA), Terengganu, Malaysia.

REFERENCES

- [1] Chen, L., et al., Face age classification based on a deep hybrid model. *Signal, Image and Video Processing*, 2018. 12(8): p. 1531-1539.
- [2] Hassan, A. and S. Viriri. Invariant feature extraction for facial recognition: A survey of the state-of-the-art. in *2018 Conference on Information Communications Technology and Society (ICTAS)*. 2018. IEEE.
- [3] Johnston, B. and P. de Chazal, A review of image-based automatic facial landmark identification techniques. *EURASIP Journal on Image and Video Processing*, 2018. 2018(1): p. 86.
- [4] Cootes, T.F., et al., Active shape models-their training and application. *Computer vision and image understanding*, 1995. 61(1): p. 38-59.
- [5] Cootes, T.F., G.J. Edwards, and C.J. Taylor, Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 2001. 23(6): p. 681-685.
- [6] Cristinacce, D. and T.F. Cootes. Feature detection and tracking with constrained local models. in *Bmvc*. 2006. Citeseer.

- [7] Iqtait, M., F. Mohamad, and M. Mamat. Feature extraction for face recognition via Active Shape Model (ASM) and Active Appearance Model (AAM). in IOP Conference Series: Materials Science and Engineering. 2018. IOP Publishing.
- [8] Zadeh, A., T. Baltrušaitis, and L.-P. Morency, Deep constrained local models for facial landmark detection. arXiv preprint arXiv:1611.08657, 2016. 3(5): p. 6.
- [9] Zahraddeen, S., S. Fatma, and A. Abdulganiyu. An efficient discrete cosine transform and gabor filter based feature extraction for face recognition. in Proceedings of the 6 th International Conference on Postgraduate Education. 2014.
- [10] Geng, X., Q. Wang, and Y. Xia. Facial age estimation by adaptive label distribution learning. in Pattern Recognition (ICPR), 2014 22nd International Conference on. 2014. IEEE.
- [11] Wang, H., J. Hu, and W. Deng, Face Feature Extraction: A Complete Review. IEEE Access, 2018. 6: p. 6001-6039.
- [12] Sharma, A.K., et al. A Survey on Feature Extraction Technique for Facial Expression Recognition System. in 2018 4th International Conference on Computing Communication and Automation (ICCCA). 2018. IEEE.
- [13] Wei, W., et al., Multi-modal facial expression feature based on deep-neural networks. Journal on Multimodal User Interfaces, 2019: p. 1-7.
- [14] Tan, S., et al., A Robust Shape Reconstruction Method for Facial Feature Point Detection. Computational intelligence and neuroscience, 2017. 2017.
- [15] Sagonas, C., et al., 300 faces in-the-wild challenge: Database and results. Image and vision computing, 2016. 47: p. 3-18.
- [16] Smola, A.J. and B. Schölkopf, A tutorial on support vector regression. Statistics and computing, 2004. 14(3): p. 199-222.
- [17] Gu, L. and T. Kanade. A generative shape regularization model for robust face alignment. in european conference on computer vision. 2008. Springer.
- [18] Saragih, J.M., S. Lucey, and J.F. Cohn, Deformable model fitting by regularized landmark mean-shift. International journal of computer vision, 2011. 91(2): p. 200-215.