Constrained Local Models (CLM) For Facial Feature Extraction using CLNF and SVR as Patch Experts

Ayah Alsarayreh, Fatma Susilawati Mohamad

Abstract: Methods for detection of facial characteristics have again developed greatly in recent times. However, they also argue in the presence of poor lighting conditions for amazing pose or occlusions. A well-established group of strategies for facial feature extraction is the Constrained Local Model (CLM). Recently, they are bringing cascaded regression-built methodologies out of favor. This is because the failure of presenting nearby CLM detectors to model the highly complex special signature look affected to a small degree by voice, illumination, facial hair and make-up. This paper keeps tabs on execution to collect facial features for the Constrained Local Model (CLM). CLM model relies on patch model to collect facial image demand features. In this paper patch model built using Support Vector Regression (SVR) and Constrained Local Neural Field (CLNF). We show that the CLNF model exceeds SVR by a large margin on the LFPW database to identify facial landmarks.

***Keywords* : Features Extraction, CLNF, SVR, CLM.**

I. INTRODUCTION

Extraction of facial features is a key starting point for an amount of study fields, like facial outflow assessment, face 3d modeling, facial quality assessment, evaluation of multimodal conclusions and individual identity. Furthermore, it is a fully investigated problem for a lot of suspend information has seen a rise in interest over the years.

Up-to-date methodologies for extraction of facial characteristics can be part of two important groups: constructed model and constructed regression. Model constructed methodologies often unambiguously Model both the presence and style of facial landmarks, the first limiting the hunting area as well as a manifestation of regularization. On the other side, regression constructed methodologies do not require an express shape model and feature detection can be done specifically on appearance ahead of time. We demonstrate a brief overview of techniques based on claiming late model and regression.

Model constructed methodologies find a face model's best variables that resemble a picture's appearance. A common model built method is the Constrained Local Model [1, 2] also its diverse properties, for instance, Constrained Local Neural Fields [3] and Discriminative Response Map Fitting [4] that utilize is only the tip of the advanced driven methods of recording neighborhood response maps also inferring landmark regions.

An alternative primary model constructed methodology is the combination of claiming trees model [5] this utilizes a tree built model of deformable components to conduct face detection mutually, posing prediction also facial point of interest. A design for this methodology is the Gauss-Newton Deformable Part Model [6] which, together with a worldwide form using Gauss-Newton optimization, optimizes a component built adaptable presence model. Additional modern suggested 3d heavy face arrangement strategy [7] refreshes the variables of a 3d Morphable model utilizing CNN and requires proven helpful facial point identification of the appearance of the profile.

Regression built models directly anticipates the appearance of the facial characteristics. The dominant part of claiming such methodologies follows a cascaded regression technique where the extraction of the feature is continuously moved forward through the application of a regressor, once the appearance given for the present features is evaluated to express regression of the shape [8]. Sun et al. [9] suggested a cascaded regression methodology developed by CNN to detect meager points of interest. At the end, Trigeorgis et al. [10] suggested Mnemonic Descent strategy using a Recurrent Neural Network to conduct cascaded regression on CNN built apparent characteristic focused near feature fields.

There have been a number of efforts to manage the issue of precise and individual independent facial feature extraction for fluctuating achievement. The Constrained Local Model (CLM) proposed via Cristinacce and Cootes [1], as well as various extensions that accompanied [2-4, 11], is one of the major guarantees. However, CLM techniques still struggle against bad lighting circumstances in the area of occlusion, even when identifying landmarks in invisible datasets.

In this article, we develop the Constrained Local Neural Field (CLNF) and Support Vector Machine (SVR) as a CLM patch expert that arrangements for facial trait detection problems in complicated scenes, allowing us both to capture more unpredictable majority of information as well as misuse spatial connections among pixels. We demonstrate that in identifying facial characteristics for the LFPW database, CLNF output outperforms SVR. The database contains 1287 images, we selected 100 as a test group to test the algorithm proposed.

Revised Manuscript Received on July 22, 2020.

***** Correspondence Author

Aya Sarayra*, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Besut Campus, 22200 Besut, Terengganu, Malaysia. E-mail: Aya_izzat_89@yahoo.com

Fatma Susilawati Mohamad, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Besut Campus, 22200 Besut, Terengganu, Malaysia. E-mail: Fatma@unisza.edu.my

II. CONSTRAINED LOCAL MODEL (CLM)

Figure 1 CLM Basic Components.

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 components: a point distribution model (PDM), patch expert, and the technique of fitting. The PDM model uses the conversion variables to locate facial characteristic focuses in the picture. Utilizing patch specialists can model the appearance of neighborhood patches around monuments of concern. Non-Uniform Regularized Landmark Mean Shift (NU-RLMS) [3] is the appropriate methodologies used in CLMs. When preparing these models on labeled examples, a suitable methodology utilized to evaluate the variables p that best suit the implied picture:

$$P^{*} = \log \min_{P} \left[\mathcal{R}(P) + \sum_{i=1}^{n} D_i(x_i, \mathrm{I}) \right] \qquad (1)$$

Where R is to the regularization expression that punishes excessively complicated or farfetched forms and D is the measure for misalignment the i^{th} point of interest is encountering toward x_i area in the picture I. The worth of $x_i = [x_i, y_i, z_i]^T$ (the area of the i^{th} feature) is regulated toward the parameters P through the PDM.

A. Point Distribution Model (PDM)

Point Distribution Models [2] are utilized to both control the landmarks areas also for adjust the form to CE-CLM schema. Unpredictable forms to last distinguished landmarks are punished utilizing the term R (p) in the mathematical equation 1. Landmarks areas $x_i = [x_i, y_i]^T$ are parameterized utilizing p = [s, t, w, q] in the following PDM Equation:

$$x_i = s.R_{2D}.(\overline{x_i} + \Phi_i q) + t \qquad (2)$$

Here $\overline{x_i} = [\overline{x_i}, \overline{y_i}, \overline{z_i}]^T$ represents the mean quality of the i^{th} characteristic, Φ_i is the 3 × m principal component matrix, furthermore q represents the m dimensional vector for variables monitoring the non-rigid shape. The rigid shape variables could a chance to be parameterized utilizing 6 scalars: the scaling term s, an interpretation $t = [t_x, t_y]$ T, and orientation $w = [w_x, w_y, w_z]^T$. Rotation parameters w control the rotation grid R_{2D} (start with two rows of 3 × 3 rotation grid R), and would to axis-angle form, because of simplicity for linearizing it. The entire shape might be portrayed via p = [s, t, w, q].

B. Patch Experts

Patch experts (also known as local detectors), are an imperative and only those Constrained Local Model. They assess the likelihood of a landmark being adjusted toward a specific pixel area. Those reaction from the i^{th} patch expert π_{xi} at the picture location x_i dependent upon the encompassing backing district is characterized as:

$$\pi_i = C_i(x_i; I) \qquad (3)$$

Where C_i is a regressor outcome for the i^{th} characteristic. A regressor that provides values from 0 (no alignment) to 1 (ideal alignment) can then model the misalignment. There bring been a number about distinctive technique suggested similarly as patch experts: many SVR models and logistic regressors, or considerably basic format matching methods. Example of patch expert response maps in Figure 2. This paper compare between SVR and CLNF as patch experts.

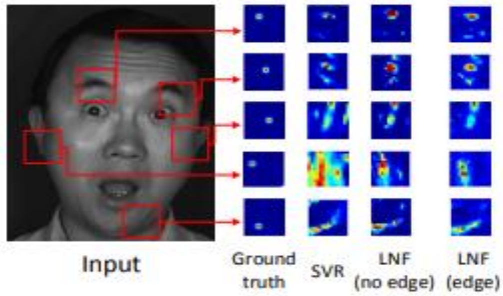

Figure 2 Example of patch expert response maps

In Figure 2 the Perfect reaction is demonstrated over ground truth column. SVR alludes of the standard patch expert utilized via CLM methodologies. We indicate two instances of CLNF model: you quit offering on that one for spatial features, (g_k and l_k) and one without.

1) Constrained Local Neural Field (CLNF)

The CLNF parameterized the utilize of p = [s, R, q, t] terms, which could be differentiated by obtaining distinct model cases: those scale element s; object rotation R (first two rows of a 3d reversal matrix); 2D interpretation t; and a vector describing non-rigid character shape q. The template for point distribution model (PDM) is:

$$x_i = s.R(\overline{x_i} + \Phi_i q) + t \qquad (4)$$

Here xi = (x, y) means the 2D area of the i^{th} characteristic perspective for an picture, xi = (X, Y, Z) will be the mean quality of the i^{th} component of the PDM in the 3D reference frame, and the vector Φ_i will be those i^{th} eigenvector gotten from the preparing set that portrays the straight varieties from claiming non-rigid shape of this characteristic purpose. In CLNF we calculate the maximum a posteriori probability (MAP) of the face model parameters p:

$$p(\mathrm{P}|\{l_i = 1\}_{i=1}^{n}, I)\,\alpha\, p(\mathrm{P}) \prod_{i=1}^{n} p(l_i = 1|x_i, I) \qquad (5)$$

Here $l_i \in \{1, -1\}$ represent 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 $\prod_{i=1}^{n} p(l_i = 1|x_i, I)$ 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 $p(l_i = 1|x_i, I)$, which is the likelihood that a function is aligned at x_i.

2) Support Vector Regression (SVR)

Throughout the training, the aim of the regression-based local appearance model may be to anticipate the uprooting vector $\Delta x_d^* = x_d^* - x$ from the nearby appearance data around x using the regression models, which will be those contrasting the middle of at any pixel area

The aim of the SVR problem is to do the hyper plane "close" to most pieces of the training landmarks as potential [12]. Provided N training points $(x_1, y_1), (x_2, y_2),..., (x_N, y_N)$ with $x_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$, i=1,…, N 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:

$$|y_i - (wx_i + b)|_\varepsilon = \begin{cases} 0 & if\ |y_i - (wx_i + b)| \le \varepsilon \\ |y_i - (wx_i + b)| - \varepsilon & otherwise \end{cases} \quad (6)$$

The worth of ε 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:

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

$$s.t. w^T \emptyset(x_i) + \mathrm{b} - y_i \le \epsilon + \xi_i$$

$$y_i - w^T \emptyset(x_i) - \mathrm{b} \le \epsilon + \xi_i^*$$

C. NU-RLMS

Non-Uniform Regularized Landmark Mean Shift (NU-RLMS) [2]. Provided for a starting CE-CLM parameter evaluates p, NU-RLMS iteratively finds an overhaul parameter Δp such-and-such $p^* = p_0 + \Delta p$. NU-RLMS update the result of the following issue:

$$\underset{\Delta p}{\operatorname{argmin}}(\|P_0 + \Delta P\|^2_{\Lambda-1} + \|J\Delta P_0 - v\|^2_w) \quad (8)$$

Here J will be those Jacobian of the Landmarks areas for admiration to parameters p. Λ^{-1} is those grid of priors for p for gaussian prior N (q; 0,Λ) for non-rigid shape also regular to shape variables. W will be a weighting grid for weighting mean shift vectors: w = w.diag(c_1;…; c_n; c_1;…; c_n) furthermore c_i will be those points of interest identifier precision ascertained throughout model preparation dependent upon relationship coefficient. V = [v_i] will be the mean shift vector ascertained utilizing a Gaussian Kernel Density Estimator utilizing response maps of CEN:

$$V_i = \sum_{y_i \epsilon \Psi_i} \frac{\pi^i_{y_i} \mathcal{N}(x_i^c; y_i; \mathcal{P}I)}{\sum_{z_i \epsilon \Psi_i} \pi^i_{z_i} \mathcal{N}(x_i^c; z_i; \mathcal{P}I)} \quad (9)$$

x_i^c is the present landmark zone forecast and p is a hyper-parameter. This guides to the NU-RLMS update rule:

$$\Delta P = -(J^T WJ + r\Lambda_{-1})(r\Lambda_{-1}P - J^T W_V) \quad (10)$$

III. RESULT AND DISCUSSION

LFPW picture database utilized to check the efficiency for the two methods. 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 two methods are compared . The test results would indicate in figure 3.

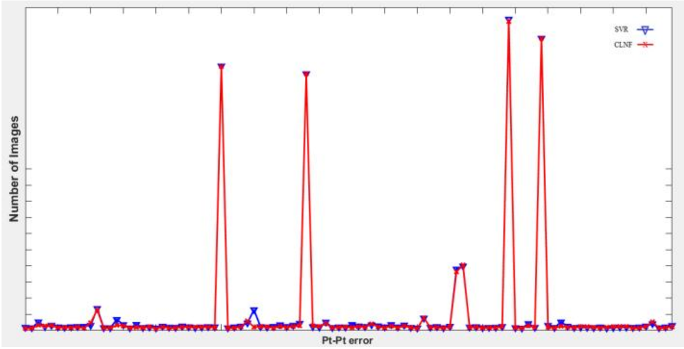

Figure 3 Point to Point Errors for CLM Model using SVR and CLNF.

Figure 3 shows result about point to point errors (pt-pt) (the range among feature landmark and the comparing landmark denoted in pictures), The X axis represent the pt-pt error among the both methods, and the Y axis represent the amount of images, the line represent the pt-pt error on characteristic point. From over results, it could make seen that, under constantly on conditions, the results of the CLNF methods will be more exact than SVR method.

Table 1 Mean Point to Point Error for CLM Model utilizing CLNF and SVR.

Method	Mean Point to Point Error
SVR	0.1040
CLNF	0.0999

Table 1 Comparing pt-pt error average for 68 facial features of test group utilizing the two methods, the outcome indicates that the variation among CLNF and SVR is equivalent to 0.0041 so that the CLNF method locates the points more correctly than the SVR method.

A. Efficiency

Time-measuring effectiveness was done to run each method sequentially on 100 pictures. Table 2 indicates CLNF and SVR methods processing time:

Table 2 Processing time for SVR and CLNF Methods.

Method	Processing Time
SVR	56.344239
CLNF	59.665334

Table 2 show the processing time for CLNF and SVR methods, each method ran individually and implemented sequentially for 100 pictures, The outcome demonstrates that CLNF method requires 59.665334 seconds to run complete operation, SVR method requires 56.344239 seconds, so SVR more efficient than CLNF.

IV. CONCLUSION

In this article, the facial characteristics extracted utilizing CLNF and SVR as patch experts for CLM model with LFPW face databases. We discovered that the CLM model will be quicker in the experimental outcomes and achieve more accurate placement of point's characteristics using CLNF than the SVR.

CLNF patch expert takes advantage of spatial relationships between patch reaction values and learns nonlinear interactions between pixel values and patch reactions. We have proved the advantage of the CLNF strategy on a number of publicly accessible datasets. In future work, these findings can be contrasted with other models that are concerned with face recognition, as well as allowing scientists to improve and construct a combined model using CLNF and SVR as patch specialists to collectively optimize an accurate place of the function point and better match the picture texture.

ACKNOWLEDGMENTS

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

REFERENCES

1. Cristinacce, D. and T.F. Cootes. *Feature detection and tracking with constrained local models*. in *Bmvc*. 2006. Citeseer.
2. 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.
3. Baltrusaitis, T., P. Robinson, and L.-P. Morency. *Constrained local neural fields for robust facial landmark detection in the wild*. in *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2013.
4. Asthana, A., et al. *Robust discriminative response map fitting with constrained local models*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013.
5. Ramanan, D. and X. Zhu. *Face detection, pose estimation, and landmark localization in the wild*. in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012. Citeseer.
6. Tzimiropoulos, G. and M. Pantic. *Gauss-newton deformable part models for face alignment in-the-wild*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.
7. Zhu, X., et al. *Face alignment across large poses: A 3d solution*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
8. Cao, X., et al., *Face alignment by explicit shape regression.* International Journal of Computer Vision, 2014. **107**(2): p. 177-190.
9. Sun, Y., X. Wang, and X. Tang. *Deep convolutional network cascade for facial point detection*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013.
10. Trigeorgis, G., et al. *Mnemonic descent method: A recurrent process applied for end-to-end face alignment*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
11. Baltrušaitis, T., P. Robinson, and L.-P. Morency. *3D constrained local model for rigid and non-rigid facial tracking*. in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. 2012. IEEE.
12. Smola, A.J. and B. Schölkopf, *A tutorial on support vector regression.* Statistics and computing, 2004. **14**(3): p. 199-222.

AUTHORS PROFILE

Fatma Susilawati Mohamad is an Associate Professor in the Department of Information Technology, Faculty of Informatics and Computing, UniSZA. She services more than 20 years in academic field and has been actively involved in teaching, research, publications, professional services, consulting, and administration. She was also been appointed for holding various administrative positions in faculty and university level. Now she serves as a Deputy Dean at the Graduate School, UniSZA. She has graduated a Bachelor of Science (Information System) from Oklahoma City University, U.S.A in 1997. She has continued his studies at Universiti Kebangsaan Malaysia and obtained her Master of Information Technology (Computer Science) in 2004. She received his Ph.D. in Computer Science from Universiti Teknologi Malaysia in 2012. During her tenure as an academic staff, she has taught various courses for undergraduate studies. (Diploma and Bachelor Degree). Her experience gained not only from the aspect of academic, but also from the aspect of research, leadership, management and administration. She has published in more than 30 articles in various conferences and refereed journals where most of her publications came from Scopus and ISI indexed journals. She has led several research grants and her work is recognized in both national and international level. She was also invited to give talk in several conferences and technical workshops national and internationally. Currently, she has more than 10 postgraduate students under supervision where 5 of them were already graduated. Her specialization is in Computer Science focusing on Pattern Recognition and Image Processing. Associate Professor Dr Fatma Susilawati Mohamad, Department of Information Technology, Faculty of Informatics and Computing Universiti Sultan Zainal Abidin (UniSZA), 21300 Gong Badak, Kuala Nerus, Terengganu, Telephone: 09- 6688796, E-mail: fatma@unisza.edu.my, Website: http:/fik.unisza.edu.my/fatma

Ayah Alsarayreh is currently a PhD Candidate at Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin Malaysia. Her research interest is in the field of Human Biometric especially face recognition.