EFFICIENT LIKELIHOOD BAYESIAN CONSTRAINED LOCAL MODEL

Anonymous ICME submission

ABSTRACT

The constrained local model (CLM) proposes a paradigm that the locations of a set of local landmark detectors are constrained to lie in a subspace, spanned by a shape point distribution model (PDM). Fitting the model to an object involves two steps. A response map, which represents the likelihood of locations for a landmark, is first computed for each landmark using local-texture detectors. Then, an optimal PDM is determined by jointly maximizing all the response maps simultaneously, with a global-shape constraint. This global optimization can be considered a Bayesian inference problem, where the posterior distribution of the shape parameters, as well as the pose parameters, can be inferred using maximum a posteriori (MAP). In this paper, based on the CLM model, we present a novel CLM variant, which employs random-forest regressors to estimate the location of each landmark, as a likelihood term, efficiently. This novel CLM framework is called efficient likelihood Bayesian constrained local model (elBCLM). Furthermore, in each stage of the regressors, the PDM local non-rigid parameters, i.e. the shape parameters, of the previous stage can work as shape clues for training the regressors for the current stage. Experimental results on benchmark datasets show our approach achieves about 3 to 5 times speed-up, when compared with the existing CLM models, and improves by around 10% on fitting accuracy, when compared with the other regression-based models.

Index Terms— Bayesian Constrained Local Models, Face Alignment; Random Forest; Point Distribution Model;

1. INTRODUCTION

The main goal of face alignment is to locate the semantic structural facial landmarks, such as the eyebrows, eyes, nose, mouth, and face contour, accurately, as illustrated in Fig. 1. This information about facial landmarks is crucial for understanding and analyzing face-related research and applications, such as expression recognition [9], face recognition [10], and face hallucination [11].

The classic active appearance model (AAM) [4] and constrained local models (CLM) [17, 18] for face alignment, which are based on Newton's gradient-descent methods, use facial-appearance information (i.e. the image pixel-intensity patterns in the area around the landmarks), and face-shape information (i.e. the face shape defined by the landmark coordinates) to locate facial landmarks. The shape model, used in active contour model (ASM) [3], AAM [4] and CLM, is usually described by finding the parameters of a statistical shape-distribution model, i.e., PDM. PDM is a linear model, where facial shapes are modeled as a linear combination of the eigen-shapes around the mean shape, which can be generalized to represent unseen facial shapes.

Fig. 1. Facial landmarks fitting located by elBCLM (the Helen dataset [7]).

Recently, a new family of face-alignment algorithms has emerged [1, 2, 5, 8, 13], which directly learns regressors from image-feature descriptors to the target shape increment. These regression-based methods are gaining popularity, due to their excellent performance and high efficiency on face alignment. The regression-based methods do not explicitly learn any shape model, rather, they learn models directly from facial appearance or manually designed features, and then predict landmark locations based on trained models.

In this paper, we propose a novel framework, which enjoys the advantages of both regression-based methods and CLM models. In our framework, the local-response maps, which are replaced with random-forest-based regressors, can be computed efficiently and cover a large area for predicting the landmark locations. Our method can also be considered a regression-based framework, with a PDM shape constraint. Therefore, we devise this algorithm as an efficient likelihood Bayesian constrained local model (elBCLM). With the PDM constraint, regression-based methods [5] with the same setting can achieve 5% improvement, in terms of accuracy. Furthermore, for each stage of the regressors, the PDM local non-rigid parameters from the previous stage can be taken as a shape clue on training each regression model. This can achieve a further 5% improvement.

The remainder of the paper is organized as follows. In Section 2, an overview of existing, related methods for face alignment is given. In Section 3, CLM is presented, and our proposed model, elBCLM, is described in detail in Section 4. A feature-switching strategy to balance the computation and efficiency in a cascaded framework, is proposed in Section 5.

Experiment results are given and discussed in Section 6, and a conclusion is given in Section 7.

2. PREVIOUS WORKS

Face-alignment methods can be classified into two major categories, the classic Newton's gradient-descent-based methods and the regression-based methods. The Newton's gradient-descent-based methods are used in the active appearance model (AAM) [4], which learns the holistic model parameters by updating the Jacobian and Hessian matrices in the fitting stage, and the constrained local model (CLM) [6, 17, 18], which determines each landmark independently, by using local-appearance information and learning local-patch response maps. The CLM models also embed the face-shape model, i.e. the point distribution model (PDM) as the shape constraint. The PDM is a linear model with parameters to represent the shapes. It is used to estimate the likelihood of the landmark potential locations, which is important for model fitting, as it can act as a prior in the Bayesian framework.

Recently, regression-based models [1, 2, 13, 15] have shown significantly better performances than the AAM and CLM frameworks. These models do not learn any shape model explicitly, but only learn models for predicting the landmarks directly from facial-appearance feature patterns. The supervised descent method (SDM) [2] is the pioneer work on regression methods. SDM formulates the face-alignment task as a general optimization problem, which is approximately solved by learning successive mapping functions from local-appearance feature patterns, such as histogram of oriented gradients (HOG) or scale-invariant feature transform (SIFT) features, to the shape updates using linear-regression models. In [1], the authors presented highly efficient local binary features (LBF), which are derived from local pixel shape-indexed features, with a linear-regression framework to achieve faster speed with comparable quality. In [5], a local lightweight feature, namely intimacy definition feature (IDF), was proposed. This feature can achieve about twice the speed-up and more than 20% improvement in terms of alignment error, when compared to the LBF [1] method.

The regression-based methods take advantage of shape information in a limited sense, such that all the points are updated jointly, i.e. each point is described by linear regression with features from all other points. Thus, the shape-pattern constraint is implicitly embedded in the model. Since the shape prior is not employed explicitly, current regression-based methods have the inherent limitation of ignorance or ineffective usage of the shape information, which is the main issue this paper will tackle.

3. CONSTRAINED LOCAL MODELS

3.1. The Shape Model - PDM model

Generally, the shape $\boldsymbol{X}$ of a point distribution model (PDM) is represented by the 2D vertex locations of a mesh, with a $2n$ dimensional vector: $\boldsymbol{X} = (x_1, y_1, .., x_n, y_n)^T$. Traditional way of building a PDM requires a set of shape-annotated images that have been aligned in scale, rotation, and translation by the Procrustes analysis. Applying principal component analysis (PCA) to a set of aligned training examples, the shape can be expressed by a linear parametric model, as follows:

$$\boldsymbol{X}_i = s\mathbf{R}(\bar{\boldsymbol{X}}_i + \boldsymbol{\Phi}_i\mathbf{q}) + \boldsymbol{t}, \quad (1)$$

where $\boldsymbol{X}_i$ denotes the 2D landmark locations of the PDM's i^{th} landmark, and $\mathbf{p} = \{s, \mathbf{R}, \mathbf{t}, \mathbf{q}\}$ denotes the PDM parameters, which consist of a global scaling s, a rotation $\mathbf{R}$, a translation $\mathbf{t}$, and a set of non-rigid parameters $\mathbf{q}$. Here, $\bar{\boldsymbol{X}}_i$ denotes the mean location of the i^{th} PDM landmark in the reference frame (i.e. $\bar{\boldsymbol{X}}_i = [\bar{x}_i, \bar{y}_i]$ for a 2D model), and $\boldsymbol{\Phi}$ is the shape-subspace matrix formed by the leading n eigenvectors (retaining a user-defined variance, e.g. 99%). Therefore, $\mathbf{q}$ can be regarded as a vector of shape parameters. From the probabilistic point of view, the non-rigid shape parameters $\mathbf{q}$ show a Gaussian distribution, leading to the following prior:

$$p(\mathbf{q}) \propto \mathcal{N}(\mathbf{q}; \mathbf{0}, \boldsymbol{\Lambda}), \ \boldsymbol{\Lambda} = \text{diag}\{[\lambda_1; \lambda_2; \ldots; \lambda_m]\} \quad (2)$$

where λ_i denotes the PCA eigenvalue of the i^{th} mode of deformation. $\boldsymbol{\Lambda}$ is constructed from the training set, based on how much shape variation in the training is explained by the i^{th} parameter, with λ_i corresponding to the $\mathbf{q}_i$ parameter.

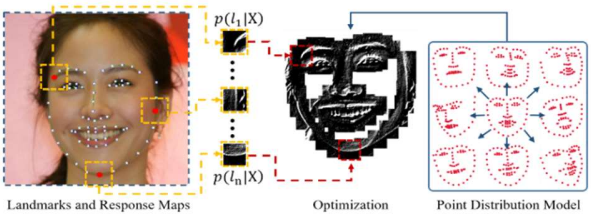

Fig. 2 Overview of the CLM fitting process.

The CLM models have attracted some interest, as they solve many of the drawbacks of holistic approaches. A CLM model normally consists of two parts: a statistical shape model and patch experts (also called local detectors). Both the shape model and patch experts can be trained offline, and then used for online landmark detection, which is achieved by fitting the CLM to a given image. The deformable model is controlled by the parameters in $\mathbf{p}$, and the instance of a model can be described by the locations of its feature points $\boldsymbol{X}_i$ in an image $\boldsymbol{I}$, as illustrated in Fig. 2.

3.2. CLM model and fitting process

The CLM fitting is generally composed of searching the PDM parameters $\mathbf{p}$, which jointly minimizes misalignment error over all the landmarks, regularized properly , as follows:

$$\mathcal{E}(\mathbf{p}) = R(\mathbf{p}) + \sum_{i=1}^{n} D_i(\boldsymbol{X}_i; I), \quad (3)$$

where R penalizes the complex deformations (i.e. the regularization term) and D_i denotes the measure of misalignment for the i^{th} landmark $\boldsymbol{X}_i$ in the image $\boldsymbol{I}$ (i.e. the data term). The form of regularization, which describes plausible object shapes, is related to the assumed distribution of the PDM parameters.

Two steps of CLM fitting process in probability thinking:

(1) an exhaustive local search for feature locations to generate the response maps:

$$\{p(l_i = aligned|\boldsymbol{I}, \boldsymbol{X})\}_{i=1}^{n}, \quad (4)$$

(2) an optimization strategy to maximize the responses of the PDM constrained landmarks.

Most innovations made to the CLM model are to replace the distribution of landmark locations, obtained from each patch-based local detector, with a simpler and more accurate predictor. For the optimization step, once the response maps for each landmark have been computed, by assuming conditional independence, optimization can proceed by maximizing the following function:

$$p(\{l_i = aligned\ _{i=1}^{n}|p) = \prod_{i=1}^{n} p(l_i = aligned|\boldsymbol{X}_i) \quad (5)$$

with respect to the PDM parameters $\mathbf{p}$, where $\boldsymbol{X}_i$ is parameterized as in (1), and dependence on the image $\boldsymbol{I}$ is dropped for succinctness. It should be noted that some forms of CLMs use (4) as minimizing the total local energy responses. In (5), l_i is a discrete random variable, denoting whether the i^{th} landmark is correctly aligned or not.

3.3. CLM in Bayesian formulation

The objective of (5) can be interpreted as maximizing the likelihood of the model parameters, such that all its landmarks are aligned with the corresponding locations of the object in an image. The specific form of the objective implicitly assumes conditional independence between the detected landmarks. The probability of correct alignment can be expressed as follows:

$$p(\mathbf{p}|\{l_i = 1\}_{i=1}^{n}, \boldsymbol{I}) \propto p(\mathbf{p}) \prod_{i=1}^{n} p(l_i = 1|\boldsymbol{X}_i, \boldsymbol{I}) \quad (6)$$

i.e. $$\ln\{p(\mathbf{p}|\{l_i = 1\}_{i=1}^{n}, \boldsymbol{I})\} \propto \ln\{p(\mathbf{p})\} + \sum_{i=1}^{n} \mathbf{ln}\{p(l_i = 1|\boldsymbol{X}_i, \boldsymbol{I})\}. \quad (7)$$

Based on equation (2), we have

$$\mathcal{E}(\mathbf{p}) = -\ln\{p(\mathbf{p})\} \quad (8)$$

$$D_i(\boldsymbol{X}_i; I) = -\mathbf{ln}\{p(l_i = 1|\boldsymbol{X}_i, \boldsymbol{I}). \quad (9)$$

If a non-informative (uniform) prior over the PDM parameters is assumed, the formulation in (6) leads to a maximum likelihood (ML) estimate, otherwise it leads to a maximum a posterior (MAP) estimate.

The method first finds the location within each response map for which the maximum was attained, and the locations of the n landmarks are denoted as $\boldsymbol{\mu} = [\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n]^T$. The objective of the optimization procedure is then to minimize the weighted least squares difference between the PDM and the coordinates of the peak responses in the map window, regularized appropriately as follows:

$$\mathcal{E}(\mathbf{p}) = ||\mathbf{q}||^2_{\boldsymbol{\Lambda}^{-1}} + \sum_{i=1}^{n} w_i ||\boldsymbol{X}_i - \boldsymbol{\mu}_i||^2, \quad (10)$$

where the weights $\{w_i\}_{i=1}^{n}$ reflect the confidence on the landmark locations, making it more resistant towards partial occlusion, where occluded landmarks will be more weakly weighted. (10) is iteratively minimized by taking the first order Taylor expansion of the PDM's landmarks:

$$\boldsymbol{X}_i \approx X_i^c + \mathbf{J}_i \boldsymbol{\Delta p} \quad (11)$$

and then solving the parameter update as follows:

$$\boldsymbol{\Delta p} = -\mathbf{H}^{-1} \boldsymbol{X}_i (\boldsymbol{\Lambda}^{-1} \mathbf{p} + \sum_{i=1}^{n} w_i \mathbf{J}_i (X_i^c - \boldsymbol{\mu}_i)), \quad (12)$$

which updates the current parameters: $\mathbf{p} \leftarrow \mathbf{p} + \boldsymbol{\Delta p}$. Here, $\boldsymbol{\Lambda} = \boldsymbol{diag}\{[\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}, \ldots, \boldsymbol{\lambda_m}]\}$, $\mathbf{J} = [\mathbf{J_1}, \mathbf{J_2}, \ldots, \mathbf{J_n}]$ is the PDM's Jacobian matrix, $\boldsymbol{X^c} = [X_1^c; X_2^c; \ldots; X_n^c]$ is the current shape estimate, and

$$\mathbf{H} = \boldsymbol{\Lambda}^{-1} + \sum_{i=1}^{n} w_i \mathbf{J}_i^T \mathbf{J}_i \quad (13)$$

is the Gauss-Newton Hessian matrix. The Jacobian matrix contains the partial derivatives of the n landmarks with respect to the PDM parameters. i.e. the 4 global rigid parameters (s, θ, t_x, t_y) and the m local non-rigid parameters in $\mathbf{q}$, where θ is the angle of rotation matrix $\mathbf{R}$.

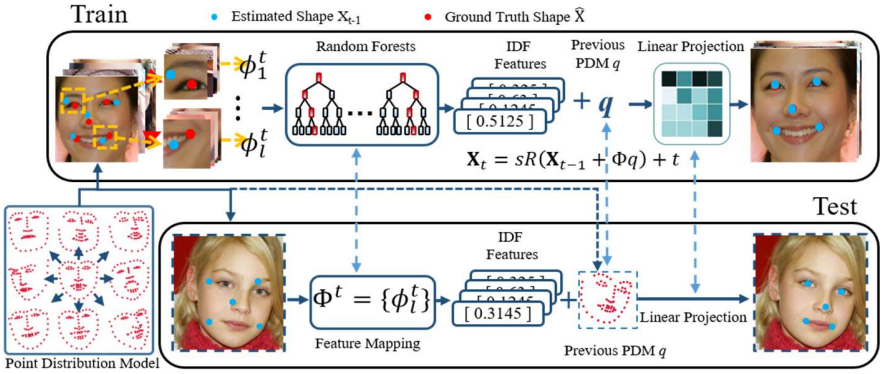

Fig. 3: An overview of the workflow for elBCM cascaded regression face alignment

4. PROPOSED MODEL

In recent years, random forests [14] have emerged as a very effective approach to learning classifiers, for a large variety of computer-vision tasks. This method is relatively simple and has many merits that make it particularly interesting for many of these computer-vision tasks as detailed in [5].

4.1. Random-Forests-based cascaded shape regression

Many face-alignment methods work under a cascaded framework, where an ensemble of N regressors operates in a stage-by-stage manner, which are referred to as stage regressors. This approach was first explored in [8]. At the fitting stage, the input to the regressor R_t, at stage t, is a tuple $(\boldsymbol{I}, \boldsymbol{X}_{t-1})$, where $\boldsymbol{I}$ is an image and $\boldsymbol{X}_{t-1}$ is the shape estimated from the previous stage (the initial shape $\boldsymbol{X}_0$ is typically the mean shape of the training set). The regressors work on features, i.e. information about the current shape estimate, or other features with respect to the current shape estimate, and predict a vector of shape increment as follows:

$$\boldsymbol{X}_t = \boldsymbol{X}_{t-1} + \boldsymbol{R}_t\big(\boldsymbol{\phi}_t(I, \boldsymbol{X}_{t-1})\big), \tag{14}$$

where $\boldsymbol{\phi}_t(\boldsymbol{I}, \boldsymbol{X}_{t-1})$ can be referred to as the shape-indexed features, or features, which are derived from shape-indexed features, such as LBF [1] and IDF [5]. The cascade progressively infers the shape in a coarse-to-fine manner, the early regressors handle large variations in shape, while the later ones perform small refinements. After each stage, the shape estimate resembles the true shape closer and closer.

In our proposed algorithm, the feature-mapping function $\boldsymbol{\phi}_t(\boldsymbol{I}, \boldsymbol{X}_{t-1})$ generates local IDF features, which are derived from the shape-indexed feature or HOG features at the estimated landmark positions at later stages. With the assumption, proved by intensive experimental results, that the shape increments have close correlation with the local features of the landmarks, which define the face shape, given the features and the target shape increments $\{\Delta\boldsymbol{X}_t = \boldsymbol{X}_t - \boldsymbol{X}_{t-1}\}$, a linear projection matrix $\boldsymbol{R}_t$ can be learned. Most regression models [1, 2, 5, 8] share a similar workflow.

4.2. Motivation and proposed method

If the weights are all the same and only one iteration is performed, then (12) can be simplified as follows:

$$\Delta\mathbf{p} = -\mathbf{H}^{-1}\mathbf{X}(\boldsymbol{\Lambda}^{-1}\mathbf{p} + \mathbf{J}\boldsymbol{v}), \tag{15}$$

where $\boldsymbol{v}$ is the shape shift vector obtained by any response-map algorithm. Recently, researchers have worked for using different response maps [18]. Directly obtaining the shape shift vector $\boldsymbol{v}$ can overcome the limitation of local response-map filters based CLM algorithms, which can obtain the shift vector by setting $\boldsymbol{v} = \Delta\mathbf{X}$, while the accuracy can be improved by iterations from the cascaded framework:

$$\Delta\mathbf{p} = -\mathbf{H}^{-1}\mathbf{X}(\boldsymbol{\Lambda}^{-1}\mathbf{p} + \mathbf{J}\Delta\mathbf{X}). \tag{16}$$

This equation interprets the main idea of our proposed method, in which random-forest-based regressors are trained to obtain $\Delta\mathbf{X}$ efficiently. $\Delta\mathbf{p}$ is updated for refining the current fitted shape with the PDM model constraint. Since the response-map filters mainly work as a convolutional filter, the drawback of convolutional filters significantly hinders the CLM algorithm. If the size of a patch-based response-map window is set too small, the response-map filter cannot cover a sufficiently large area to handle large posed faces, on the other hand, if it is set too big, the computation requirement may greatly reduce algorithm's efficiency. The PDM-based prior term, as in (2), can be approximated as follows:

$$p(\mathbf{q}_k|\mathbf{q}_{k-1}) \propto \mathcal{N}\big(\mathbf{q}_k|\mu_q, \boldsymbol{\Sigma}_q\big), \tag{17}$$

where $\mu_q = \mathbf{q}_{k-1}$ and $\boldsymbol{\Sigma}_q = \boldsymbol{\Lambda}$. This form of prior assumption can be greatly improved in cascaded framework.

Inspired by this analysis, we propose a novel CLM variant, by replacing the response-map filters with regressors to find the shift vector $\Delta\mathbf{X}$. The refinement of the shift vector $\Delta\mathbf{X}$ is implicitly realized in the cascade framework. For CLM, random-forest-based regressors replace the local response maps, so likelihood can be calculated efficiently under the Bayesian framework. Because of these, we name our proposed algorithm as an efficient likelihood Bayesian constrained local model (elBCLM).

Our proposed algorithm, using regressors to replace local response maps has the following advantages: (1) This method can circumvent the hypothesis that the local detectors are assumed conditionally independent, (2) Our method is able to avoid local optimal location, which may be caused by typical, local noise and ambiguities, since small image patches often contain limited structure, (3) This method is capable of extending the response-map mode with more efficient methods. The whole workflow of our elBCLM algorithm is described in Fig. 3, and the detail of random-forest-based regression can reference to [1, 5].

4.3. PDM as prior for regression features

The PDM is a linear model, which parametrizes a class of shapes. It can also be used to estimate the likelihood of landmark location, given a set of feature points. This is important for model fitting, as it can act as a prior, which works as a guiding feature in the cascade framework.

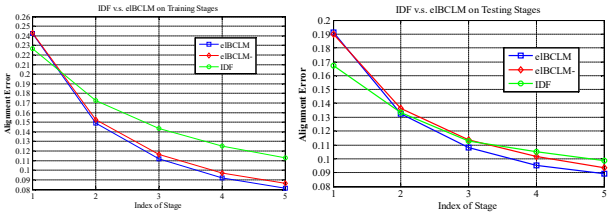

Fig. 4: Comparison of regression-based method IDF [5], elBCLM with PDM shape constraint only, and elBCLM.

As the local non-rigid parameters **q** in the PDM **p** = {*s*, **R**, **t**, **q**} have a closer relationship to the consecutive fitting shapes, so the previous local parameters **q** can work as a guiding feature for fitting the shape at each stage. In our experiments, for the 68 facial landmarks in a face, the dimension of **q** is 136 (68×2), which can be reduced to around a dimension of 32 by using PCA, with 99% of energy retained. Therefore, in our algorithm, the dimension of **q** is set at 32. As **q** is used in regression, this means that the features in $\boldsymbol{\phi}_t(\boldsymbol{I}, \boldsymbol{X}_{t-1})$, from (14), are refined.

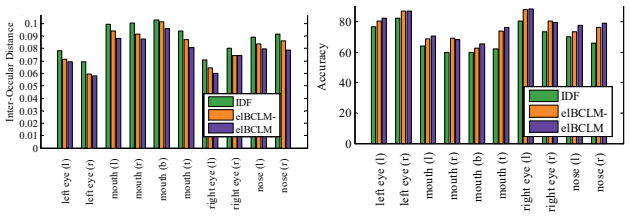

Fig. 5: Comparison of regression-based method IDF[5], elBCLM- (with PDM shape constraint only), and elBCLM, for 10 facial landmarks.

Fig. 4 shows that, both on the training and fitting stages, the regression algorithm with PDM shape constraint only (denoted as elBCLM- in Fig. 4) can reduce the alignment error by more than 5%, when compared to IDF [5]. When the PDM's local parameters **q** are also used as a guiding feature for training the linear-regression models in each stage, our algorithm, elBCLM, is able to achieve an additional 5% improvement in terms of alignment error in the fitting stage. The performances of IDF, elBCLM- (with PDM constraint only), and elBCLM, in terms of accuracy and the Inter-Occular distance criterion, for different facial landmarks are shown in Fig. 5, which demonstrates that the three methods have similar relative performances for different landmarks.

5 FEATURE SWITCHING SCHEME

In the cascade framework, the performance of the regressors can be fine-tuned. Adaptive techniques can be used. For example, adaptive window size is used, where the window size decreases when moving along the stages to balance the computation [1, 2, 5]. In [19], to enhance the capability of handling large variations, global regression is first employed, then part regression, and finally local regression.

From the experiment results, we found that both the simple and complex features can converge relatively faster in the earlier stages, as can be seen in Fig. 6. After the earlier stages, the convergence is insensitive to the features being used. Therefore, to balance speed and accuracy, a simple feature is first used, then after a certain number of stages, it is switched to a more complex, discriminative feature in later stages. We call this a feature-switching scheme. In Fig. 6, we show the performance of the feature-switching scheme with 6 stages. In the first 4 stages, IDF feature is used to achieve a faster speed, and it is switched to HOG feature on the later 2 stages to get higher accuracy, since the computation requirement of one HOG stage is higher than four IDF stages.

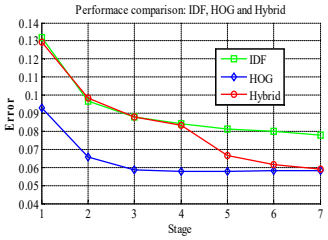

Fig. 6: Alignment comparison: IDF, HOG and Hybrid (Helen dataset [7]).

6. WORKFLOW AND EXPERIMENTAL RESULTS

6.1. Algorithm workflow

The two stages of our proposed algorithm are described in Algorithm 1 and Algorithm 2, respectively.

Algorithm 1: elBCLM Training Stage:

Input: PDM $(\boldsymbol{X}, \boldsymbol{\Phi})$ model, training data$(\boldsymbol{I}_i, \boldsymbol{X}_i, \bar{\boldsymbol{X}}_i)$, for i=1, …, N, where $\boldsymbol{I}$ are face images, and $\boldsymbol{X}$ are shapes; N is the number of samples.
Output: regressors: $\boldsymbol{R} = (R_1, \ldots, R_T)$, T: stage count.
for t=1 to T **do**
  for all $i \in (1 \ldots N)$ **do**
    $\Delta \boldsymbol{X}_t^i = \boldsymbol{X}_t^i - \bar{\boldsymbol{X}}_t^i$ ⇒ calculate $\Delta \boldsymbol{X}_t^i$
    $f_t^i = \phi_t(\boldsymbol{I}^i, S_{t-1}^i)$ ⇒ IDF features + PMD's **q**
  end for
  $R_t = \arg min_R \sum_i |R(f_t^i) - \Delta \boldsymbol{X}_t^i|$
  for all $i \in (1 \ldots N)$ **do**
    $\bar{\boldsymbol{X}}_t^i = \bar{\boldsymbol{X}}_t^i + R(f_t^i)$ ⇒ update shape {Eqn. (14)}
    update Δp with ΔS_t^i ⇒ {Eqn. (16)}
    $\bar{\boldsymbol{X}}_t^i \approx \bar{\boldsymbol{X}}_t^i + \mathbf{J}_t \Delta \mathbf{p}$ ⇒ update shape w.r.t: Eqn. (10)
  end for
end for

Algorithm 2: elBCLM Fitting Stage:

Input: PDM $(\boldsymbol{X}, \boldsymbol{\Phi})$ model, testing image $\boldsymbol{I}$, initial (mean) shape $\boldsymbol{X}^0$, trained regressors: $\boldsymbol{R} = (\boldsymbol{R}_1, \ldots, \boldsymbol{R}_T)$
Output: Estimated pose $\boldsymbol{X}^T$
for t=1 to T **do**
  $f_t = \phi_t(\boldsymbol{I}, \boldsymbol{X}_{t-1})$ ⇒ IDF features + PMD's **q**
  $\Delta X = R_t(\boldsymbol{\phi}_t(I, \boldsymbol{X}_{t-1}))$ ⇒ apply regressor R_t
  $\boldsymbol{X}_t = \boldsymbol{X}_t + \Delta \boldsymbol{X}$ ⇒ update shape {Eqn. (14)}
  update Δ**p** with $\Delta \boldsymbol{X}$ ⇒ {Eqn. (16)}
  $\boldsymbol{X}_t \approx \boldsymbol{X}_t + \mathbf{J}_t \Delta \mathbf{p}$ ⇒ update shape w.r.t: Eqn. (10)
end for

6.2. Experimental Results

Experiments on the Helen dataset [7] show that elBCLM achieves an alignment error of 5.88, at 150 FPS on 68 facial landmarks. Table 1 tabulates the fitting accuracy, as well as other state-of-the-art algorithms. We can see that elBCLM outperforms the state-of-the-art algorithms. Fig. 7 illustrates fitting results based on elBCLM, LBF [1], and CLM [18], which show elBCLM can locate landmarks more accurately.

Method	Error (68 landmarks)
Zhu et. al [16]	8.16*
DRMF [12]	6.70*
RCPR [6]	5.93*
Tadas et. al [18]	6.75
elBCLM	5.88

Table-1: Alignment comparision, results from original papers with "*".

Fig. 7: Fitting results comparison (68 points, Helen dataset [7]), Row-1:LBF[1]: Row-2:CLM[18], Row-3: elBCLM

7. CONCLUSIONS

In this paper, we propose a more accurate and efficient face alignment algorithm. There are two main contributions in our algorithm. The first one is that we connect two schools of face alignments into a single framework. The second one is that the PDM non-rigid local parameters are used as a guiding, discriminative feature for the cascade alignment framework. Experiment results show that our proposed algorithm outperforms the state-of-the-art algorithms in terms of accuracy and efficiency.

8. REFERENCES

[1] Shaoqing Ren, Xudong Cao, Yichen Wei, Jian Sun, Face Alignment at 3000 FPS via Regressing Local Binary Features, CVPR, 2014.

[2] Xiong, X., De la Torre, F, Supervised descent method and its applications to face alignment. CVPR-2013.

[3] T. F. Cootes and C. J. Taylor. Active shape models — "smart snakes". In BMVC, pp. 266–275, 1992.

[4] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. ECCV-1998.

[5] Hailiang Li, Kin-Man Lam, Man-Yau Chiu, Kangheng Wu, Zhibin Lei. Cascaded Face Alignment via Intimacy Definition CoRR abs/1611.06642, (2016), https://arxiv.org/abs/1611.06642

[6] X. P. Burgos-Artizzu, P. Perona, and P. Dollar. Robust face landmark estimation under occlusion. ICCV-2013

[7] V. Le, J. Brandt, Z. Lin, L. Boudev and T. S. Huang, Interactive Facial Feature Localization, ECCV-2012.

[8] P. Dollar, P. Welinder, and P. Perona. Cascaded pose regression. In CVPR, pp. 1078–1085, 2010.

[9] S. W. Chew, P. Lucey, S. Lucey, J. M. Saragih, J. F. Cohn, and S. Sridharan. Person-independent facial expression detection using constrained local models. In FGR, pp. 915–920, 2011.

[10] H. Gao, H. K. Ekenel, and R. Stiefelhagen. Pose normalization for local appearance-based face recognition. ICB-2009.

[11] N. Wang, D. Tao, X. Gao, X. Li, and J. Li. A comprehensive survey to face hallucination. Journal of Computer Vision, 106(1):9–30, 2014.

[12] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, Robust discriminative response map fitting with constrained local models. CVPR-2013.

[13] Xudong Cao et al., Face Alignment by Explicit Shape Regression, IJCV-2014.

[14] Leo Breiman. Random Forests. ML, 45(1):5–32, 2001.

[15] V. Kazemi and J. Sullivan, One millisecond face alignment with an ensemble of regression trees, CVPR-2014.

[16] X. Zhu and D. Ramanan. Face detection, pose estimation and landmark localization in the wild. CVPR-2012.

[17] J. Saragih, S. Lucey, J. Cohn. Deformable Model Fitting by Regularized Landmark Mean-Shift. IJCV-2011

[18] Baltrusaitis T, Robinson P, Morency L P. Constrained local neural fields for robust facial landmark detection in the wild, ICCV Workshops. 2013.

[19] Liu Liu, Jiani Hu, Shuo Zhang, Weihong Deng, Extended Supervised Descent Method for Robust Face Alignment, ACCV 2014 Workshops, pp.71-84