# Complex Bingham Distribution for Facial Feature Detection

Eslam Mostafa[1,2] and Aly Farag[1]

[1] CVIP Lab, University of Louisville, Louisville, KY, USA
[2] Alexandria University, Alexandria, Egypt
{eslam.mostafa,aly.farag}@louisville.edu

**Abstract.** We present a novel method for facial feature point detection on images captured from severe uncontrolled environments based on a combination of regularized boosted classifiers and mixture of complex Bingham distributions. The complex Bingham distribution is a rotation-invariant shape representation that can handle pose, in-plane rotation and occlusion better than existing models. Additionally, we regularized a boosted classifier with a variance normalization factor to reduce false positives. Using the proposed two models, we formulate our facial features detection approach in a Bayesian framework of a maximum a-posteriori estimation. This approach allows for the inclusion of the uncertainty of the regularized boosted classifier and complex Bingham distribution. The proposed detector is tested on different datasets and results show comparable performance to the state-of-the-art with the BioID database and outperform them in uncontrolled datasets.

## 1 Introduction

The face analysis pipeline usually consists of four modules: face detection, face alignment, feature extraction and face matching. Face detection is the first step in this process since it segments the facial region from the background before further processing is performed. The next stage is face alignment, where facial components, such as the eyes, nose, and mouth and facial outline, are located. Accurate detection of these facial components is crucial to the success of the later stages of the pipeline.

Valstar et al. [1] describe the difference between facial component detection, where entire facial features (e.g., mouth region) are detected, and feature point detection, where more detailed points inside the facial features (e.g., mouth corners) are located. Tasks such as face recognition, gaze estimation, facial expression analysis, gender and ethnicity classification often rely on these finer points. In this work, we propose a novel facial feature point detector that will locate 15 feature points, as illustrated in Fig. 1.

There have been a number of recent methods that have shown great accuracy in locating feature points in mostly frontal images and controlled environments. Our immediate goal is facial feature point detection in challenging real-life situations such as face recognition at-a-distance, where there are different illumination

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part II, LNCS 7584, pp. 330–339, 2012.


**Fig. 1.** Sample of results of the proposed facial feature detector on our collected outdoor at-a-distance images (**Top row**) and Labeled Faces in the Wild (LFW) dataset (**Bottom row**)

conditions, variability in pose and expression, existence of occlusion and image acquisition of faces is performed outdoor and at-a-distance (50m to 150m) from the camera. Moreover, we have taken into consideration the speed of this facial feature point detection approach to approach real-time.

Since no public dataset is available with these requirements, we have acquired a database for testing our detector besides the existing databases in the literature. Our collected images, as shown in Fig. 1(**Top row**), are taken in uncontrolled environments at far distances, where instances such as heavy shadowing across the feature points may occur and parts of the face can be occluded (e.g., by sunglasses, hair, or scarf). Pose is also varied from near frontal to severe pose ($\pm 45\%$), where some feature points are occluded.

Previous work on facial features detectors can be classified into two main groups : (a) view-based and (b) 3D-based detectors. View-based approaches train on a set of 2D models; each model can cope with shape or texture variation within a small range of pose. 3D-based approaches [2] can handle multiple views using only a single 3D model but can be sensitive to initialization and computationally expensive. View-based approaches are widely used compared to its 3D counterpart.

The texture and shape prior models are the main components for building a view-based detector. For the texture model, the local texture around given facial feature is modeled,(i.e., the pixels intensity in a small region around the feature point), while for the shape model, the relationship among facial features are modeled. Both models are learned from labeled exemplar images.

Texture-based detectors aim to find the best suitable point in the face that matches the texture model. The texture model can be constructed using different descriptors such as Haar-like [3], local binary pattern (LBP) [4] , Gabor [5], scale-invariant feature transform (SIFT) [6] features. The search problem can be formulated either as a regression or classification. For the classification problem, a sliding window runs through the image to determine if each pixel is a feature or non-feature. For the regression problem, the displacement vector from an initial point to the actual feature point is estimated.

Texture-based detectors are imperfect for many reasons; visual obstructions such as hair, glasses,and hands can greatly affect the results. The detection of each facial feature is also independent from others and it ignores the relation among these facial feature points. To overcome the disadvantages of texture-based detectors, constraints related to the relative location of facial features from each other can be established from the shape model. The relationship among facial feature positions is commonly modeled as a single Gaussian distribution function [7,8], which is the model used by the Active Appearance Model (AAM) and Active Shape Model (ASM) algorithms.

Cristinacce et al. [9] modeled the relative positions of facial features by a pairwise reinforcement of feature responses, instead of a Gaussian distribution while Valstar et.al [1] modeled shape using the Markov Random Field (MRF). These two approaches use a single distribution, which is not suitable for modeling a wide range of poses. Everingham et.al [10] extended the model of a single Gaussian distribution into a mixture of Gaussian trees. Belhumeur et.al [6] used a non-parametric approach, using information from their large collection of diverse, labeled exemplars.

In this work, we propose a novel view-based detector based on a regularized boosted classifier coupled with a mixture of complex Bingham distributions. The following are the contributions of this paper: (a) use of a mixture of complex Bingham distributions to model various viewpoints, (b) regularizing a boosted classifier with a variance normalization factor to reduce false positives, and (c) a new energy function for facial features detection combining two uncertainty terms related to (a) and (b).

The complex Bingham distribution is more robust in modeling the joint probability of the location of facial features than existing models; existing models need a preprocessing step before using the shape prior to filter out scale, translation, and rotation using least-square approaches (e.g., Procrustes analysis), which can introduce errors to the system due to noise and outliers. Since the probability distribution function (PDF) of a complex Bingham has a symmetric property, there is no need to filter out rotation. Scale and translation can be easily removed by a simple mean and normalization step [11].

We propose to regularize the output of the boosted classifier to handle false positives in the classification step of each pixel in a certain neighborhood as feature or non-feature. The output of the classifier should give a high response in the actual facial feature position and decrease smoothly going away from the actual position. If the neighborhood variance is low, it is certain that one pixel

position is the actual feature point; otherwise, if the neighborhood variance is high, i.e., all classifier outputs in the neighborhood have combined high or low scores, the classifier is uncertain if a feature point exists in the area. Regularization is performed by dividing a normalization term to the classifier output related to the standard deviation of the output probability scores in the search neighborhood.

Finally, we formulate the facial feature point detection problem as an energy minimization function that incorporate information from both texture and shape models simultaneously, while most of the state-of-the-art approaches use the shape model to improve the results of texture-based methods. The proposed method is compared with existing algorithms on different datasets. Figure 1 shows a sample results of our facial features detector method on a sample of tested images.

## 2 Facial Feature Extraction

In this section, we describe our texture and shape prior models and how the problem of facial feature point detection is formulated as an energy minimization function that incorporates the uncertainty of the texture model response and the shape prior model.

### 2.1 Texture Model

In this work, Haar-like features are chosen as the descriptor of local appearance. The first real-time face detector used these features for detection in [12]. The main advantage of Haar-like features over most other features is in the calculation speed. Using integral images, a Haar-like feature of any size can be computed in constant time.

For all training samples, we rescale the images such that the face bounding box is $80 \times 80$. The optimal patch size around a given facial feature position has been empirically determined to be $16 \times 16$. Positive samples are taken at manually annotated locations. Negative samples are taken at least 20 pixels away from the annotated locations. For each facial feature, a feature/non-feature classifier was trained using the AdaBoost learning algorithm on the positive and negative samples.

Given the face detection bounding box of an input image, we extract sub-images for each facial feature point; each sub-image is the search space of a given facial feature point. The center of this sub-image is the mean position of the feature point in all training images after filtering translation, scale and rotation. The width and height of the sub-image is based on the variance of the feature position. A sliding window is run over the sub-image and the AdaBoost classifier assigns a score for each pixel to be the correct position of the facial feature. The score at position Z is given by $S(D_{Z_i}) = \sum_{t=1}^{r} \alpha_{t_i} F_{t_i}(Z_i)$ where $\alpha_{t_i}$ is the weight of weak classifier $t$ for the feature $i$ and $F_{t_i}$ is the binary response of weak classifier.

In the case of a perfect texture-based detector, the classifier response is homogenous as the probability of the pixel being a feature is high at the true position and decreases smoothly going away from this position. Therefore, we regularize the output of the classifier with a variance normalization factor by dividing the output probability of classifier with $\sigma_{\aleph(Z)}$. $\sigma_{\aleph(Z)}$ is the standard deviation of the output probability among the neighborhood $\aleph(Z)$. Thus, the probability of position $Z$ is the position of the feature $i$ based on the texture detector can be written as

$$P(D_{Z_i}) = \frac{K}{\sigma_{\aleph(Z_i)}} \sum_{t=1}^{r} \alpha_{t_i} F_{t_i}(Z_i)$$

where $K$ is the normalization constant.

Since each feature has a corresponding sub-image that has a sliding window classifier running over it, the output of each texture detector can be considered independent from others. Therefore, the overall probability of $\mathbf{Z} = [Z_1, Z_2 ... Z_N]$ is the positions vector of $N$ facial features based on the texture-based detector is given by $P(D_{\mathbf{Z}}) = \prod_{i=1}^{N} P(D_{Z_i})$

### 2.2 Shape Prior Model

Faces come in various shapes due to differences among people, pose, or facial expression of the subject. However, there are strong anatomical and geometric constraints that govern the layout of facial features. The representation of shape, i.e., joint distribution between facial feature points, is described by various models in the literature. The active shape model (ASM) is one example, which is based on a single Gaussian distribution.

Typically, one would like to have a shape representation that is invariant to translation, scale and rotation. A common way to address this problem is to use least-squares (LS) fitting methods, .e.g., Procrustes analysis [11], where misalignments due to noise and outliers may happen [13]. Moreover, an iterative procedure is needed to align multiple shapes.

We propose to use the complex Bingham distribution [14] for our facial feature detection approach. The advantage of using this distribution is that shapes do not need to be aligned with respect to rotation parameters. The probability distribution function of the complex Bingham is

$$P(Y) = c(A)^{-1} exp(Y^* A Y) \tag{1}$$

where $c(A)$ is a normalizing constant.

Since the complex Bingham distribution is invariant to rotation, it is suitable to represent shape in the pre-shape domain, the domain where shapes are zero-offset and unit-scale. In our work, we use the classical way of transforming from the original shape vector to the pre-shape domain by simply multiplying with Helmert sub-matrix(H) to the original shape vector the (matrix) and then performing[11].

Multiplying the original shape vector with the Helmert sub-matrix (H) will project the original facial features position vector $Z \in C^n$ to $C^{n-1}$. Then, the shape representation using complex Binghanm is

$$P(\mathbf{Z}) = c(A)^{-1} exp(\frac{H\mathbf{Z}}{\parallel H\mathbf{Z} \parallel}^{*} A \frac{H\mathbf{Z}}{\parallel H\mathbf{Z} \parallel}) \tag{2}$$

where $A$ is a $(N-1)*(N-1)$ Hermitian parameter matrix, $N$ is number of landmarks or facial feature points. The spectral decomposition can be written as $A = \mathbf{U}\Lambda\mathbf{U}^*$, where $\mathbf{U} = [U_1 U_2 \cdots U_{N-1}]$ is a matrix whose columns $U_i$ correspond to the eigenvectors of A and $\Lambda = diag(\lambda_1, \cdots, \lambda_{N-1})$ is the diagonal matrix of corresponding eigenvalues.

The normalization constant $c(A)$ is given by $c(A) = 2\pi^{N-1} \sum_{i=1}^{N-1} a_i \exp(\lambda_i)$ where $a_i^{-1} = \prod_{m \neq i}(\lambda_i - \lambda_m)$ The log likelihood of parameters is

$$L(\Lambda, \mathbf{U}) = \sum_{i=1}^{N-1} \lambda_i U_i^* S U_i - N \log c(A) \tag{3}$$

where the matrix $S$ is a $N-1 \times N-1$ matrix denoting the auto correlation matrix for manually annotated shapes that have zero mean and unit scale. The maximum likelihood estimators are given by [11]

$$U_i = G_i \quad i = 1, 2, \cdots, N-1 \tag{4}$$

and the solution to

$$\frac{d \log c(\Lambda)}{d\lambda_i} = \frac{l_i}{N} \tag{5}$$

where $\mathbf{G} = [G_1 G_2 \cdots G_{N-1}]$ denotes the corresponding eigenvector of $S$ and $L = diag(l_1, l_2 \cdots l_{N-1})$ is the diagonal matrix of corresponding eigenvalues.

Since no exact solution exists, we estimate $\lambda$ by minimization of function $F$

$$F_i = \frac{d \log c(\Lambda)}{d\lambda_i} - \frac{N}{l_i} \tag{6}$$

This function is linearly approximated and solved iteratively using gradient descent.The update equation of parameter $\lambda$ is given by

$$\lambda_i^{t+1} = \lambda_i^t - \kappa \frac{a_i + \lambda_i^t \sum_{m=1}^{N-2} \prod_{i \neq m \neq k}(\lambda_i - \lambda_k)}{\sum_{i=1}^{N-1} a_i \lambda_i^t} \tag{7}$$

Using the above equation, the parameters of complex Bingham distribution $A$ and $c(A)$ can be estimated off-line from the training shapes examples which are manually annotated from the MUCT dataset. Since the deformation of shape due to different poses is large and cannot be handled by a single distribution [10], [15], we divide the training annotated shapes into $M$ classes. Each class carries a small range of poses and has its own parameters $A_m$ and $c(A_m)$ and a

Bayes classifier rule is used to determine which class the test image belongs to. In this work , we use $M = 5$ which correspondence to poses $-45^o$,$-25^o$, frontal, $25^o$,and $45^o$ The index of class is given by

### 2.3 Combining Texture and Shape Model

In facial feature detection problems, we try to recover numerous of hidden variables (position of facial features) based on observable variables (image gray level). This problem can be formulated as a Bayesian framework of maximum a-posteriori (MAP) estimation. We want to find the vector Z, which maximizes the response probability for the texture model and shape model.

$$\hat{Z} = \arg\max P(I|Z)P(Z). \tag{8}$$

$P(I|Z)$ represents the probability of similarity between the texture of the face to off-line model given the facial feature vector. Since the similarity of the face can be expressed in the similarity of the windows around each facial feature, it can be written as $P(W(Z_i), W(Z_2) \cdots W(Z_N)|Z)$. Where $W(Z_i)$ is the image window around the facial point $Z_i$. The windows around each facial point can be considered independent from each others. Therefore

$$P(I|Z) = \prod_{i=1}^{N} P(W(Z_i)|Z_i), \tag{9}$$

where $P(W(Z_i)|Z_i)$ can be interpreted as the probability of a pixel being feature based on the texture model. Based on boosted classifier and Haar-like feature vector the probability can be written as

$$P(W(Z_i)|Z_i) = P(D_{Z_i}) = \frac{K}{\sigma_{\aleph(Z_i)}} \sum_{t=1}^{r} \alpha_{t_i} F_{t_i}(Z_i) \tag{10}$$

Therefore, the maximum-a-posteriori estimate of facial features can be formulated as an energy minimization of function $E(Z)$

$$E(\mathbf{Z}) = -\frac{H\mathbf{Z}^* AH\mathbf{Z}}{\| H\mathbf{Z} \|^2} - \sum_{i=1}^{N} \log P(D_{Z_i}) \tag{11}$$

This energy function is non-linear and not amenable to gradient descent-type algorithms. It is solved by a classical energy minimization technique, which is simulated annealing where maximum number of iterations is empirically set to 100 iterations.

$$m^* = \arg_m \min \frac{H\mathbf{Z}}{\| H\mathbf{Z} \|}^* A_m \frac{H\mathbf{Z}}{\| H\mathbf{Z} \|} + \log(c_m(A)) \tag{12}$$

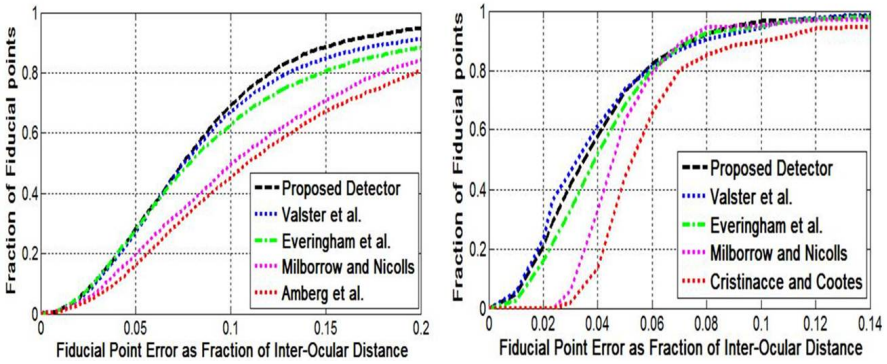


**Fig. 2.** A comparison of the cumulative error distribution measured on our collected images dataset(Left) and BIO-ID(Right)

## 3 Experiments

Our work focuses on facial feature extraction in severe uncontrolled environments, pose, existence of shadow, presence of occlusion objects as sunglasses or subject's hand, existence of in-plane rotation, and blurred images at range distances in real time. Existing available public databases for evaluation facial feature extraction do not consider long distance, above 50 meters, and blurred images. So we collected 1541 faces for 55 subjects. These images are taken at distances of 30, 50, 80, 100, 150 meters with a Canon 7D camera attached to a $800mm$ telephoto lens. Furthermore, most of the researchers about facial features detection in the literature reported results on the BiOID database,therefore we also included it to test the proposed detector. The BioID dataset [16] contains 1521 images, each showing a near frontal view of a face in controlled indoor environments with no illumination and occlusion problems for 23 distinct subjects.

In our experiment, we compare proposed detector with existing algorithms which are the extended Active shape Model (STASM) [17], compositional image alignment (AAM) [18], the detector proposed by Everingham et al. [10], and the detector proposed by Valster et al.[1]. STASM shows good performance in locating facial features on various datasets and most researchers use this detector for comparison, e.g., [18], [1]. The compositional image alignment approach is a modification of the original AAM and has better results than its predecessor. The detector proposed by Everingham et al. [10] shows competitive results for the commercial product COTS on a LFPW dataset [6]. The detector proposed by Valster et al shows the best results on BioID dataset. We excluded from our comparison the detector which is proposed in [6]; however, they reported excellent results on LFPW dataset. Since this task need to be as fast as possible due to being part of real time system, our detector takes on average 0.47 seconds

on an Intel Core $i7$ 2.93 GHz machine for locating all facial feature points. Figure 2 shows the cumulative error distribution of $me_{17}$ defined by [19] for our detector compared to those reported by [19],[1], [5],[17] on collected image at distance and BIO-ID.

For investigation, the effect of the regualized boosted classifier as texture model and the complex bingham as shape model, we conduct an experiment from four tests on our collected images in an uncontrolled environment. First, The boosted classifier without the regulairzed term is used as texture model along with the gaussian distribution as a shape model. Second, the regularized term is added to the boosted classifier while keeping the shape model as gaussian. Third,the regularized term is kept while changing the shape model to complex bingham distribution. Last, the regularized term is discarded while keeping the shape model as complex bingham distribution, as shown in figure 3 (b). It shows that each of regularized term and the complex bingham improves the result especially when the detected facial features are far from the correct one. However,the effect of complex bingham distribution is more significant than the regularized term.

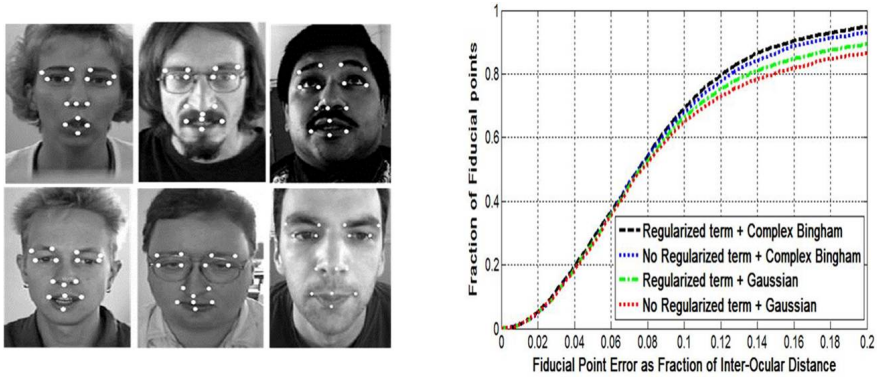


**Fig. 3.** (a) Some results of our proposed detector on the BioID dataset. (b) The Evaluation of the effect of regularized term in the texture model and complex bingham as shape model on our collected images dataset.

## 4 Conclusion

We have described a new approach for facial features detection based on complex Bingham distribution and regularized boosted classifier. We combine the uncertainty of the response of complex Bingham and boosted classifier is an energy minimization function. Our detector is robust under a variation of pose, in-plane rotation, expression, occlusion, and illumination. It shows that we outperform existing facial detector in uncontrolled environment images and achieve a comparable results in the less challenging dataset BIO-ID.

## References


1. Valstar, M., Martinez, B., Binefa, X., Pantic, M.: Facial point detection using boosted regression and graph models. In: Proceedings of IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR 2010), San Francisco, USA, pp. 2729–2736 (2010)
2. Zhang, Z., Liu, Z., Adler, D., Cohen, M.F., Hanson, E., Shan, Y.: Robust and rapid generation of animated faces from video images: A model-based modeling approach. Int. J. Comput. Vision (2004)
3. Eckhardt, M., Fasel, I.R., Movellan, J.R.: Towards practical facial feature detection. IJPRAI
4. Shan, C., Gong, S., Mcowan, P.: Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing, 803–816 (2009)
5. Vukadinovic, D., Pantic, M.: Fully automatic facial feature point detection using Gabor feature based boosted classifiers. In: IEEE Int'l Conf. on Systems, Man and Cybernetics 2005, pp. 1692–1698 (2005)
6. Belhumeur, P.N., Jacobs, D.W., Kriegman, D.J., Kumar, N.: Localizing parts of faces using a consensus of exemplars. In: The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2011)
7. Cristinacce, D., Cootes, T.: Facial feature detection using adaboost with shape constraints. In: 14th British Machine Vision Conference, Norwich, England, pp. 231–240 (2003)
8. Cristinacce, D., Cootes, T.: Boosted regression active shape models. In: 18th British Machine Vision Conference, Warwick, UK, pp. 880–889 (2007)
9. Cristinacce, D., Cootes, T., Scott, I.: A multi-stage approach to facial feature detection. In: 15th British Machine Vision Conference, London, England, pp. 277–286 (2004)
10. Everingham, M., Sivic, J., Zisserman, A.: "Hello! My name is... Buffy" – automatic naming of characters in TV video. In: Proceedings of the British Machine Vision Conference (2006)
11. Dryden, I., Mardia, K.V.: The statistical analysis of shape. Wiley, London (1998)
12. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vision, 137–154 (2004)
13. Hamsici, O., Martinez, A.: Rotation invariant kernels and their application to shape analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1985–1999 (2009)
14. Bingham, C.: An antipodally symmetric distribution on the sphere. The Annals of Statistics, 1201–1225 (1974)
15. Saragih, J.M., Lucey, S., Cohn, J.: Face alignment through subspace constrained mean-shifts. In: International Conference of Computer Vision, ICCV (2009)
16. Frischholz, R.W., Dieckmann, U.: Bioid: A multimodal biometric identification system. Computer, 64–68 (2000)
17. Milborrow, S., Nicolls, F.: Locating Facial Features with an Extended Active Shape Model. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 504–513. Springer, Heidelberg (2008)
18. Brian Amberg, A.B., Vetter, T.: On compositional image alignment with an application to active appearance models. In: Proceedings of IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR 2009) (2009)
19. Cristinacce, D., Cootes, T.: Feature detection and tracking with constrained local models. In: 17th British Machine Vision Conference, Edinburgh, UK, pp. 929–938 (2006)