Illumination Robust Facial Feature Detection via Decoupled Illumination and Texture Features

Brendan Chwyl, Alexander Wong, and David A. Clausi

University of Waterloo, Waterloo ON N2L3G1, Canada,
{bchwyl,a28wong,dclausi}@uwaterloo.ca,
WWW home page: http://vip.uwaterloo.ca/ *

Abstract. A method for illumination robust facial feature detection on frontal images of the human face is proposed. Illumination robust features are produced from weighted contributions of the texture and illumination components of an image where the illumination is estimated via Bayesian least-squares minimization with the required posterior probability inferred using an adaptive Monte-Carlo sampling approach. This estimate is used to decouple the illumination and texture components, from which Haar-like features are extracted. A weighted aggregate of each component's features is then compared with a cascade of pre-trained classifiers for the face, eyes, nose, and mouth. Experimental results against the Yale Face Database B suggest higher sensitivity and F_1 score values than current methods while maintaining comparable specificity and accuracy in the presence of non-ideal illumination conditions.

Keywords: illumination robust, object detection, image processing

1 Introduction

The detection of facial features such as the eyes, nose, or mouth on human faces is useful in a wide variety of applications. Examples of such applications include biometric authentication [7], gaze tracking [16], and human-computer interaction [2]. Methods exist to address the problem of facial feature detection in controlled environments, however, their performance suffers when non-ideal illumination conditions are present.

Viola and Jones [22] suggest a method for object detection using a cascade of simple features and apply this method specifically to face and facial feature detection. By cascading a collection of Haar-like features over an image at different scales, strong and weak classifiers can be trained. Weak classifiers determine regions of likely facial features in order to narrow the search area and progressively stronger classifiers are used to positively identify objects of interest. This results in a fast and scale-invariant method for object detection. Various improvements to speed and accuracy [4, 23, 24] have made this method widely used. Naruniec [17] compiled a comprehensive survey of facial feature

* We would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Research Chairs Program, and the Ontario Ministry of Research and Innovation for their sponsorship of this research.

detection methods and while this survey recognizes the accuracy and speed of appearance-based methods, it notes that trained classifiers are unreliable when recognizing features not well represented in the training data, such as facial hair, face orientation, or illumination conditions.

Methods exist to address the problem of inconsistent illumination of human faces. Gourier *et al.* [9] present a method to extract features from a face which are robust to pose and illumination using linear combinations of Gaussian derivatives. Hu *et al.* [12] detect faces under varying illumination conditions by using a YCbCr skin-colour model. Local binary patterns (LBPs) [18] have been used with success for the purpose of face recognition [1, 13, 21] and have been extended to general object detection tasks [26, 27] as well as face detection in particular [10, 11, 20]. One other possible alternative is to first employ Retinex approaches [5, 14, 25] prior to feature extraction.

This paper will be organized as follows: the proposed method is first presented, followed by a description of the experimental setup, a discussion of the experimental results, and conclusions.

2 Proposed Method

The proposed method aims to build upon the framework established by Viola and Jones [22] by extracting illumination robust features for comparison against trained classifiers to compensate for non-ideal illumination conditions. This is achieved by decoupling the illumination and texture components of an image through Bayesian least-squares estimation with Monte Carlo posterior sampling [25]. Weighted contributions from each of these aspects are then used to produce robust features for comparison against trained classifiers. A flow chart illustrating the steps used in this method can be seen in Fig. 1.

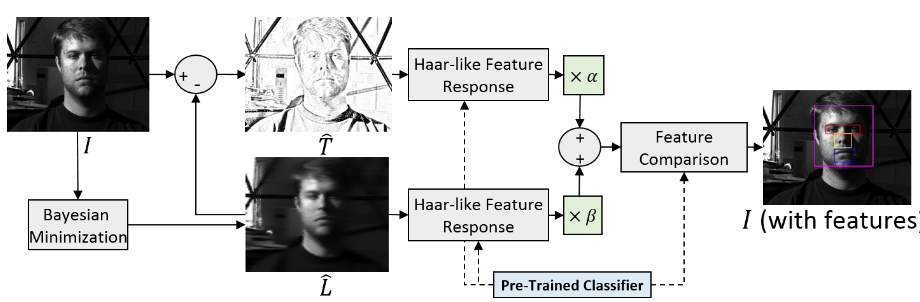


Fig. 1: General illustration of the proposed method. A weighted combination of features extracted from the texture and illumination components is obtained for each classifier in a cascade. This feature is then compared based on methods described by Viola and Jones [22].

2.1 Illumination Robust Features

Since the texture component of an image is both useful in quantifying local distinctiveness and relatively insensitive to spatially-varying illumination changes, contributions from texture are helpful in producing illumination robust features. However, texture alone fails to encapsulate sufficient geometric information, leading to the need to also incorporate the information from the illumination aspect of an image. We model the image I as an additive relation of the texture, T, and illumination, L, components:

$$I = T + L. \tag{1}$$

Let f_T and f_L denote sets of Haar-like features [19] extracted from T and L respectively. A weighted combination of these feature sets can be produced as

$$f_{I'} = \alpha f_T + \beta f_L, \tag{2}$$

where α and β are weighting factors and $f_{I'}$ represents the set of illumination compensated features.

2.2 Texture Illumination Decoupling

To produce these features, the texture (T) and illumination (L) aspects of an image are required. We aim to disassociate T and L by first producing an approximation of L (denoted as $\hat{L}$) and calculating an estimate of T (denoted as $\hat{T}$) based off of the model described in Eq. 1 as $\hat{T} = I - \hat{L}$. To obtain $\hat{L}$, a Bayesian least-squares minimization approach is used. This minimization can be formulated as

$$\hat{L} = \arg\min_{\hat{L}} E((L - \hat{L})^2 | I), \tag{3}$$

where $E(.)$ denotes the expectation and $\hat{L}$ represents the estimate of L. Based on work by Lui *et al.* [15], the solution to the minimization can be written as

$$\hat{L} = \int L p(L|I) dL \tag{4}$$

The posterior probability, $p(L|I)$, is necessary for this calculation; however, it is difficult to obtain analytically. For this reason, we adapt the Monte Carlo sampling approach proposed in [25] to infer the required posterior probability distribution. We first establish a set of pixels, $\varOmega$, from a region of interest, $\eta_{\bar{q}}$, surrounding a pixel of interest, $\bar{q}$. From a uniform distribution, $Q(q_k, \bar{q})$, pixels $q_1, q_2, ..., q_M$ are sampled with equal probability. An acceptance probability, $\alpha(q_k|\bar{q})$, is calculated for each sampled pixel, q_k, based on its regional similarity to the center pixel, $\bar{q}$, as follows:

$$\alpha(q_k|\bar{q}) = \exp\left(\sigma - \frac{1}{N}\sum_{i=1}^{N}(\aleph_{q_k}(i) - \aleph_{\bar{q}}(i))^2\right) \tag{5}$$

where $\aleph_{q_k}$ and $\aleph_{\bar{q}}$ are regions of equal size surrounding q_k and $\bar{q}$ respectively, σ is a constant, and N represents the number of pixels in each region. The likelihood that q_k is added to Ω is determined by $\alpha(q_k|\bar{q})$. The sampling process is repeated until M sample pixels are acquired, at which point, the posterior probability can be estimated as

$$\hat{p}(L|I) = \frac{\sum_{k=0}^{M} \alpha(q_k|\bar{q})\delta(L - I(q_k))}{Z}, \tag{6}$$

where Z is a normalization factor such that $\sum \hat{p}(L|I) = 1$ and $\delta(.)$ represents the Dirac function.

2.3 Feature Cascade Object Detection

Feature cascade object detection [22] relies on many classifiers trained over a large number of positive and negative sample images. These classifiers are trained by applying Haar-like features to each training image at various position and scale to produce many classification features. AdaBoost [6] is used to select the features which best classify the objects and combine these features into weak classifiers. By iteratively comparing analogous features from the input image with those in the weak classifiers, areas of unimportance can be quickly discarded. Weak classifiers alone cannot classify an image, however, a strong classifier consisting of the weighted summation of weak classifiers is sufficient to detect objects of interest.

3 Experimental Setup

For the purpose of testing the performance of the proposed method under drastic lighting conditions, the Yale Face Database B was used [8]. This database provides images of human faces for 39 different subjects facing various angles under 64 different illumination conditions per subject. For this project, only the frontal view of each face was used. Each image is 256 bits, 480 x 640, and grayscale. The various illumination conditions include combinations of lighting angles ranging from -130^o to $+130^o$ horizontally and -40^o to $+90^o$ degrees vertically. In addition, one image taken with ambient lighting is included for each subject. Example images from the database can be viewed in Fig. 2. It should be noted that subject 14 was not available for download and subject 16 had no corresponding frontal view and were therefore omitted from testing.

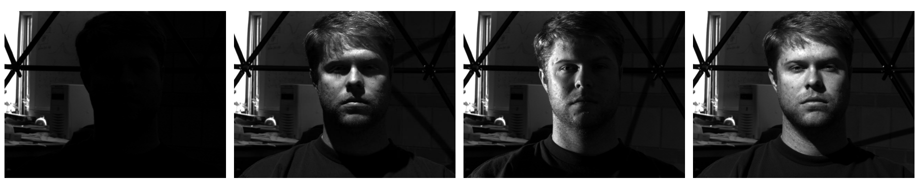
Ground truth was obtained for each subject by performing detection under ideal lighting conditions. These results were visually verified and manually corrected. Because each of the 64 different illumination condition images were taken in rapid succession, it was assumed that the location of facial features for each subject did not vary. For this reason, the ground truth acquired for each subject under ideal lighting conditions was used as the ground truth across all lighting scenarios for the same subject.

For every subject and lighting condition, the method proposed by Viola and Jones [22] as well as our method were applied. For comparison purposes, we also tested two Retinex-based methods for improving illumination robustness in facial feature detection: i) Gaussian Retinex as described by Jobson *et al.* [14], and ii) Bilateral Retinex as proposed by Elad [5]. To increase computation speed, illumination estimation for all decoupling methods was performed on images which were down sampled by a factor of four. In addition, the α and β values used throughout this project are 0.85 and 0.15 respectively.

Pre-trained classifiers available through the OpenCV library [3] were used throughout this paper. Because multiple instances of faces, eye pairs, noses, and mouths may be returned, automatic selection of the most probable true features is performed based on logical assumptions regarding the face. Such assumptions include: each face contains only one eye pair, one nose, and one mouth, and each eye pair, nose, and mouth feature is approximately horizontally centered on the face.

From the detected regions, areas of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) were calculated based on areas of overlap with the ground truth data. The resulting values were then used to calculate four metrics for analysis: sensitivity, specificity, accuracy, and F_1 score. The equations for these metrics are defined as follows:

$$Sensitivity = \frac{TP}{TP+FN}, \qquad Specificity = \frac{TN}{TN+FP},$$
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad F_1\ Score = \frac{2TP}{2TP+FP+FN}, \tag{7}$$



(a) Ambient lighting (b) Light shone at $+45^o$ vertically (c) Light shone at $+90^o$ horizontally (d) Light shone at $+35^o$ horizontally and $+40^o$ vertically

Fig. 2: Selection from Yale Face Database B

4 Experimental Results

The average sensitivity, specificity, accuracy, and F_1 score across all subjects and lighting conditions can be seen in Table 1. Our method achieves sensitivity values and F_1 scores which are generally higher than those achieved by the other presented methods, however, the specificity and accuracy values remain similar. These values are likely skewed due to the difference in the number of detected features; in these cases, regions of false positive are zero while the regions of true negative are the entire image, thus leading to large specificity and increased accuracy values as per Eq. 7. The improved sensitivity values and F_1 scores

Table 1: Average values of sensitivity, specificity, accuracy, and F_1 score calculated across all subjects and lighting conditions. The number of features that went undetected are also shown. The most desirable values for each metric are presented in boldface.

	Viola [22]	Jobson [14]	Elad [5]	Ours	Viola [22]	Jobson [14]	Elad [5]	Ours
	Face				Eyes			
Sensitivity	80.21%	93.31%	90.02%	**93.92%**	45.47%	64.42%	71.62%	**76.34%**
Specificity	98.03%	97.79%	**98.52%**	98.41%	**99.95%**	99.92%	99.90%	99.92%
Accuracy	94.30%	96.85%	96.74%	**97.45%**	98.86%	99.20%	99.33%	**99.44%**
F_1 Score	85.32%	92.38%	91.83%	**93.82%**	61.09%	76.02%	80.65%	**84.25%**
Undetected	73	27	22	**0**	1054	635	478	**358**
	Nose				Mouth			
Sensitivity	52.81%	65.14%	61.27%	**73.01%**	46.38%	62.85%	61.27%	**64.87%**
Specificity	99.66%	99.62%	**99.78%**	99.53%	99.54%	99.58%	**99.70%**	99.64%
Accuracy	99.19%	99.28%	**99.40%**	99.26%	98.82%	99.09%	**99.18%**	99.17%
F_1 Score	55.93%	63.73%	**66.12%**	65.90%	51.13%	64.27%	65.83%	**67.29%**
Undetected	443	313	481	**85**	325	84	71	**31**

demonstrated by our method suggest that all facial features were detected with greater regions of true positive and smaller regions of false negatives than in the case without illumination compensation.

While all illumination-robust methods resulted in improved performance when compared to the method proposed by Viola and Jones [22], our method results in better performance with the exception of F_1 score for the nose. This suggests that the proposed method is able to better handle complex non-ideal illumination scenarios to facilitate for improved facial feature detection. While Elad [5] and Jobson *et al.* [14] rely on the assumption that L is piece-wise smooth, our method is able to avoid this assumption, leading to better handling of sharp illumination changes.

5 Conclusions

A method for illumination robust facial feature detection by considering contributions from both the illumination and texture aspects of an image was proposed. Furthermore, it was proposed that the decoupling of illumination and texture be achieved by Bayesian minimization. Results indicate higher sensitivity and F_1 score values while achieving similar specificity and accuracy when compared to the method proposed by Viola and Jones [22], as well as two Retinex-based methods.

Future work will include further validation against a wider variety of state of the art methods. In addition, training of classifiers from illumination compensated features will be explored as a means to achieve improved performance.

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. Pattern Analysis and Machine Intelligence, IEEE

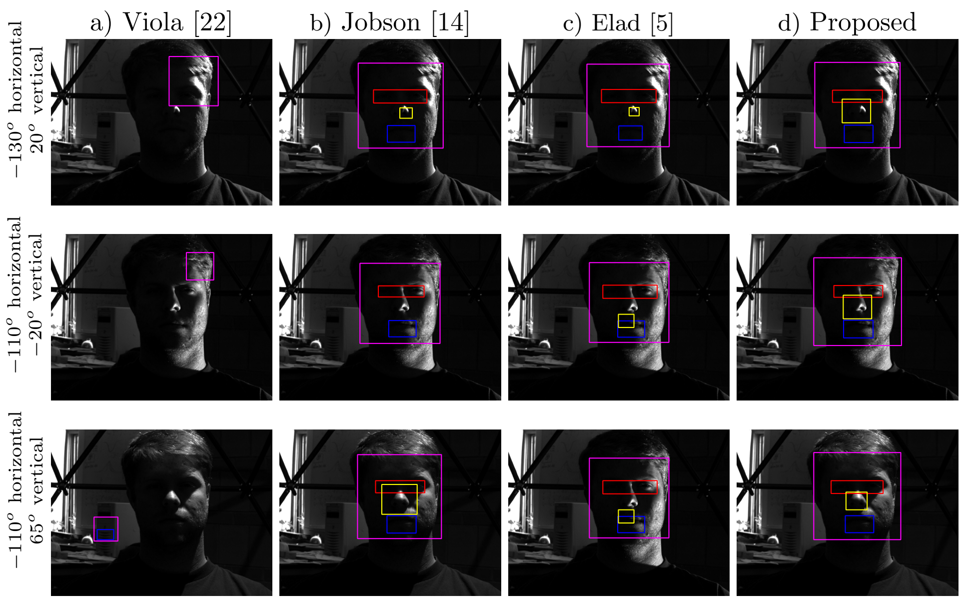


Fig. 3: Examples of facial feature detection performed for the face, eye pair, nose, and mouth. Results from the aforementioned methods are shown in each column, while each row represents a different lighting angle as labelled.

Transactions on 28(12), 2037–2041 (2006)

2. Bartlett, M.S., Littlewort, G., Fasel, I., Movellan, J.R.: Real time face detection and facial expression recognition: Development and applications to human computer interaction. In: Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on. vol. 5, pp. 53–53. IEEE (2003)
3. Bradski, G.: Dr. Dobb's Journal of Software Tools (2000)
4. Cristinacce, D., Cootes, T.F.: Facial feature detection using adaboost with shape constraints. In: BMVC. pp. 1–10. British Machine Vision Association (2003)
5. Elad, M.: Retinex by two bilateral filters. In: Scale Space and PDE Methods in Computer Vision, pp. 217–229. Springer (2005)
6. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences 55(1), 119–139 (1997)
7. Frischholz, R.W., Dieckmann, U.: Biold: a multimodal biometric identification system. Computer 33(2), 64–68 (2000)
8. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intelligence 23(6), 643–660 (2001)
9. Gourier, N., Hall, D., Crowley, J.L.: Facial features detection robust to pose, illumination and identity. In: Systems, Man and Cybernetics, 2004 IEEE International Conference on. vol. 1, pp. 617–622. IEEE (2004)
10. Hadid, A., Heikkila, J., Silvén, O., Pietikainen, M.: Face and eye detection for person authentication in mobile phones. In: Distributed Smart Cameras, 2007. ICDSC'07. First ACM/IEEE International Conference on. pp. 101–108. IEEE (2007)

11. Hadid, A., Pietikainen, M., Ahonen, T.: A discriminative feature space for detecting and recognizing faces. In: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. vol. 2, pp. II–797. IEEE (2004)
12. Hu, W.C., Yang, C.Y., Huang, D.Y., Huang, C.H.: Feature-based face detection against skin-color like backgrounds with varying illumination. Journal of Information Hiding and Multimedia Signal Processing 2(2), 123–132 (2011)
13. Huang, D., Ardabilian, M., Wang, Y., Chen, L.: 3-d face recognition using elbp-based facial description and local feature hybrid matching. Information Forensics and Security, IEEE Transactions on 7(5), 1551–1565 (2012)
14. Jobson, D.J., Rahman, Z.U., Woodell, G.A.: Properties and performance of a center/surround retinex. Image Processing, IEEE Transactions on 6(3), 451–462 (1997)
15. Lui, D., Modhafar, A., Glaister, J., Wong, A., Haider, M.a.: Monte Carlo bias field correction in endorectal diffusion imaging. IEEE transactions on bio-medical engineering 61(2), 368–80 (Feb 2014)
16. Morimoto, C.H., Mimica, M.R.: Eye gaze tracking techniques for interactive applications. Computer Vision and Image Understanding 98(1), 4–24 (2005)
17. Naruniec, J.: A survey on facial features detection. International Journal of Electronics and Telecommunications 56(3), 267–272 (2010)
18. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern recognition 29(1), 51–59 (1996)
19. Papageorgiou, C.P., Oren, M., Poggio, T.: A general framework for object detection. In: Computer vision, 1998. sixth international conference on. pp. 555–562. IEEE (1998)
20. Roy, A., Marcel, S.: Haar local binary pattern feature for fast illumination invariant face detection. In: British Machine Vision Conference 2009. No. LIDIAP-CONF-2009-048 (2009)
21. Shan, C., Gong, S., McOwan, P.W.: Robust facial expression recognition using local binary patterns. In: Image Processing, 2005. ICIP 2005. IEEE International Conference on. vol. 2, pp. II–370. IEEE (2005)
22. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. vol. 1, pp. I–511. IEEE (2001)
23. Vukadinovic, D., Pantic, M.: Fully automatic facial feature point detection using gabor feature based boosted classifiers. In: Systems, Man and Cybernetics, 2005 IEEE International Conference on. vol. 2, pp. 1692–1698. IEEE (2005)
24. Wilson, P.I., Fernandez, J.: Facial feature detection using haar classifiers. Journal of Computing Sciences in Colleges 21(4), 127–133 (2006)
25. Wong, A., Clausi, D.A., Fieguth, P.: Adaptive monte carlo retinex method for illumination and reflectance separation and color image enhancement. In: Computer and Robot Vision, 2009. CRV'09. Canadian Conference on. pp. 108–115. IEEE (2009)
26. Zhang, H., Gao, W., Chen, X., Zhao, D.: Object detection using spatial histogram features. Image and Vision Computing 24(4), 327–341 (2006)
27. Zhang, L., Li, S.Z., Yuan, X., Xiang, S.: Real-time object classification in video surveillance based on appearance learning. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. pp. 1–8. IEEE (2007)