

Chapter 7

Understanding Thermal Face Detection: Challenges and Evaluation

Janhavi Agrawal, Aishwarya Pant, Tejas I. Dhamecha, Richa Singh and Mayank Vatsa

Abstract In thermal face detection, researchers have generally assumed manual face detection or have designed algorithms that focus on indoor environment. However, facial properties are dependent on body temperature, surrounding environment, and any accessories or occlusion present on the face. For instance, the presence of scarfs, glasses, or any disguise accessories will alter the emitted heat pattern, thereby making it challenging to detect the face in thermal images. Similarly, daytime outdoor image acquisition has certain effects on the heat pattern compared to nighttime (or indoor controlled) image acquisition settings that affect automatic face detection performance. In this research, we provide a thorough understanding of challenges in thermal face detection along with an experimental evaluation of traditional approaches. Further, we adapt the AdaBoost face detector to yield improved performance on face detection in thermal images in both indoor and outdoor environments. We also propose a region of interest selection approach designed specifically for aiding occluded/disguised thermal face detection. Experiments are performed on the Notre Dame thermal face database as well as the IIITD databases that include variations such as disguise, age, and environmental

Janhavi Agrawal, Aishwarya Pant, Tejas I. Dhamecha: Equal contributions by the student authors.

J. Agrawal · A. Pant · T.I. Dhamecha · R. Singh · M. Vatsa (✉)
IIIT Delhi, New Delhi, India
e-mail: mayank@iiitd.ac.in

J. Agrawal
e-mail: janhavi11056@iiitd.ac.in

A. Pant
e-mail: aishwarya11011@iiitd.ac.in

T.I. Dhamecha
e-mail: tejasd@iiitd.ac.in

R. Singh
e-mail: rsingh@iiitd.ac.in

(day/night) factors. The results suggest that while thermal face detection in semi-controlled environments is relatively easy, occlusion and disguise are challenges that require further attention.

7.1 Introduction

Decades of research in face recognition has seen several research directions, mostly in the visible spectrum and many high-performing algorithms have been developed for this purpose. To instigate further research, several research programs such as Janus¹ have been initiated where the goal is to take face recognition to the next significant level. It is also well understood that in order to have a large-scale application, the technology has to encompass *face recognition both in and beyond visible spectrums*, i.e., developing capabilities to recognize face images/videos in visible, near infrared, and thermal spectrums. Compared to the visible spectrum, research in face recognition beyond visible spectrum is relatively less explored and has primarily focused on near infrared and thermal imagery [2, 3, 5, 6, 15, 19]. As shown in Fig. 7.1, face images in these three spectrums provide non-overlapping information and can be individually or in-combination used for identity management.

For recognizing face images captured in thermal images (spectrum range of 8–12 μm), the first step is the face detection followed by feature extraction and matching against gallery image(s). Similar to visible spectrum, thermal face detection can be modeled as a two-class problem (*face* and *non-face*). Trujillo et al. [25] proposed a thresholding-based approach for detecting faces in thermal images. Since the goal is to recognize expressions, face detection accuracy is not reported in that study. Selinger and Socolinsky [21] and Socolinsky and Selinger [23] applied boosted class-cover catch digraph (CCCD) [24] for face detection. They [23] observed that thermal face recognition performance degrades in outdoor environments. Since the overall goal was to identify the subject, the results of the intermediate face detection stage were not reported. In [23], the authors focused on face recognition and the results of detection were not reported. However, it is possible that in thermal spectrum the outdoor setting affects the detection stage too, particularly in thermal spectrum. Martinez et al. [18] utilized GentleBoost along with Haar-like features [26] and, the results showcase that, to an extent, boosting with Haar-like features can be utilized for face detection. However, evaluation in challenging environments remains an open research problem. Wang et al. [28] observed that Haar-like features with AdaBoost can be useful for detecting eyes, even in the presence of glasses. Zhang et al. [29] proposed a modified boosting approach in which visible images could also be utilized along with the images of other spectrum

¹<http://www.iarpa.gov/index.php/research-programs/janus>.

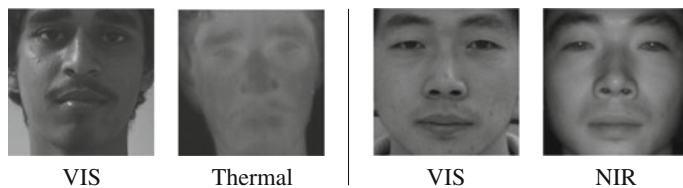


Fig. 7.1 Sample face images captured in visible, thermal, and near-infrared (NIR) spectrum. NIR image has been taken from CASIA NIR-VIS 2.0 face database [14]

to train the cascade model. Table 7.1 briefly summarizes these algorithms on face detection in thermal images.

For designing an efficient face recognition algorithm, it is important that the face detection is accurate. It has been observed that imprecise eye localization and therefore imprecise face localization degrades the performance of the overall thermal face recognition pipeline [6]. Since the majority of researchers have used manually detected face images in the recognition pipeline, thermal face detection has not been well explored in the literature. Moreover, in order to learn an efficient face detector it is imperative to have access to a large amount of face and non-face images. The samples obtained in diverse conditions, such as indoor and outdoor

Table 7.1 Summary of related research for face detection in thermal images

Authors	Objective and technique	Dataset (#Images/#Subjects)
Trujillo et al. [25]	Face detection: thresholding. Facial regions: Harris detector with k -means clustering	IRIS dataset in OCTBVS [12] (4228/30)
Selinger and Socolinsky [21]	Face detection: boosted CCCD classifier [24]. Eye detection: Haar-like features with AdaBoost	Private (3732/207)
Socolinsky and Selinger [23]	Face detection: boosted CCCD classifier [24]. Eye detection: co-registration with the visible images	Private indoor-outdoor (3080/385)
Martinez et al. [18]	Face detection: patch intensities. Eye, nostril and mouth detection: Haar features with GentleBoost and self-similarity descriptor	Private (78/22)
Wang et al. [28]	Eye localization: Haar-like features with AdaBoost for 15 subregions around eyes and majority voting of results of multiple classifiers	NVIE [27] and Equinox dataset
Zhang et al. [29]	Face detection: R-TrBoost to train using visible and other spectrum images together	Private (7000 visible, 1400 spectrum 1, and 705 spectrum 2)
Bourlai and Jafri [1]	Blob analysis-based skin detection, template matching, and integral projection-based eye detection	West Virginia University visible-thermal database (2250/50)

environments, with session variations are necessary to learn a generalizable detector. In our opinion, thermal face detection and recognition research is impaired by the non-availability of a challenging database that includes face and non-face images captured in both indoor and outdoor environments with time lapse variations. Moreover, there exists a very limited literature focusing on detection of occluded thermal faces. Therefore, it is important that the challenge of face detection is addressed to achieve a fully automated and efficient thermal face recognition system. In view of existing limitations, this chapter attempts to bridge the gap in the following ways:

- A database, namely IIITD thermal face database, is prepared that consists of 614 face images pertaining to 65 subjects and 150 non-face images. Face images are captured in two sessions separated by two years time frame. Non-face images are captured in both indoor and outdoor settings. A small set, IIITD-People in Sun and Evening (IIITD-PSE), consisting 22 subjects is also prepared to study the variations due to outdoor day light (sun) and nighttime environments. The database and the ground truth annotations of face regions will be made publicly available for researchers to undertake research on thermal face detection via <https://research.iiitd.edu.in/groups/iab/>.
- Baseline experiments pertaining to face localization are performed on the IIITD thermal face, IIITD-PSE, and Notre Dame (ND) thermal face [6] databases with Haar- and LBP-cascaded AdaBoost to analyze the challenges associated with face detection in thermal images. Challenging scenarios such as cross-sensor thermal face detection and the effects of outdoor conditions (i.e., day or night) are also examined. A baseline evaluation of detecting faces under occlusion (using disguise accessories) is also performed on the IIITD In and Beyond Visible Spectrum Disguise (I²BVSD) face database [8, 9]. A skin detection-based region of interest (ROI) selection is proposed, to improve the face detection performance. We also propose a novel face detection evaluation measure to evaluate the performance of face detection algorithms.

7.2 IIITD Thermal Face Database

As shown in Table 7.1, there are multiple thermal face databases available. However, all of them captured face data with the objective of face recognition in controlled environments and may not be suitable for understanding the state of the art of face detection algorithms in the thermal spectrum. Further, existing face detection algorithms have been optimized for the visible spectrum, and since both visible and thermal spectra have different characteristics, such optimized pre-trained models may not yield the best results. Therefore, we have collected the IIITD thermal face database with a focus on capturing the variations that may affect the

appearance of facial regions in a thermal image, for instance, time lapse and environment. The IIITD database consists of 614 thermal face images pertaining to 65 individuals and 150 non-face images. All the images are captured using a thermal camera having micro-bolometer sensor operating in 8-14 μm spectrum range, also known as long-wave infrared spectrum. Face images are near frontal with neutral expressions and are captured in two sessions:

- Session I is captured in October/November 2011 and it consists of 82 images pertaining to 41 subjects.
- Session II is captured in January 2014 and it consists of 532 images from 65 subjects. There are 41 overlapping subjects in both the sessions.

A set of 150 non-face images is collected out of which equal number of images are captured indoor and outdoor. Since a face can appear very different during day and night in a thermal image, we collected a separate dataset named *IIITD-People in Sun and Evening (IIITD-PSE) database* to capture these variations. It consists of 120 images pertaining to 22 subjects acquired in outdoor settings during both day (around 2 p.m. and $\sim 36^\circ\text{C}$ temperature) and night (around 10 p.m. and $\sim 22^\circ\text{C}$ temperature). Both subsets, the IIITD-PSE-Day and IIITD-PSE-Night, contain 60 images pertaining to 15 subjects, with an overlap of 8 subjects.

All the images are of size 720×576 pixels. The details of both IIITD and IIITD-PSE datasets are summarized in Table 7.2. Figure 7.2 illustrates the variety of images contained in the IIITD and IIITD-PSE databases. For evaluating the performance of face detection algorithms, the ground truth has been manually annotated in terms of two eyes, nose, and mouth coordinates. To encourage the research on the problem of thermal face detection, the database and annotated ground truth will be made publicly available to researchers.

Table 7.2 Dataset details pertaining to sessions, subjects, and classes

Name	Class	Number of subjects and images
IIITD Thermal	Face	Session I: 41, 82
		Session II: 65 (41 + 24), 532
	Non-face	75 indoor and 75 outdoor
Total = 65 subjects, 764 images		
IIITD-PSE	Face (outdoor)	Day: 15, 60
		Night: 15 (8 + 7), 60
	Total = 22 subjects, 120 face images	
ND	Face	Train: 159, 159
		Test: 82, 2292
	Total = 241 subjects, 2451 images	
I ² BVSD	Face	Minor: 75, 307
		Major: 75, 231
	Total = 75 subjects, 538 images	

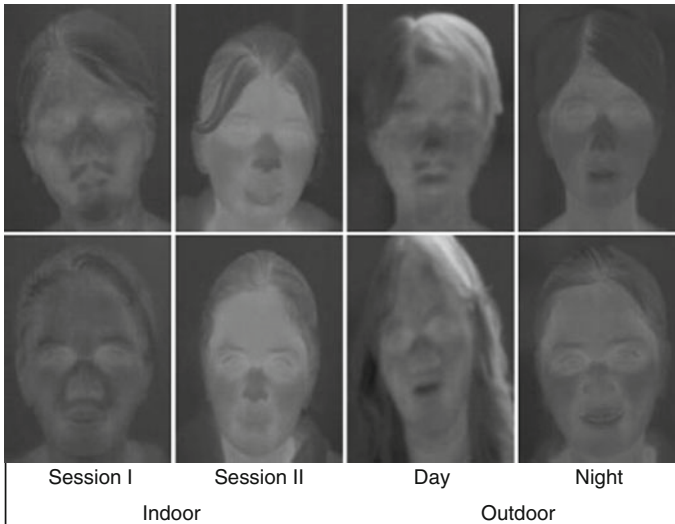


Fig. 7.2 Images illustrating the variations captured in the IIITD and IIITD-PSE thermal face database. Each row contains images of one subject under different environments

7.3 Databases, Algorithms, and Evaluation Measures

For understanding the performance of thermal face detection, Collection XI of the University of Notre Dame (ND) face dataset [6] and I²BVSD database [8, 9] are used along with the IIITD and IIITD-PSE databases. The ND face database consists of 2292 infrared frontal face images of size 312×239 from 82 subjects. It also contains a separate training set of 159 face images. Figure 7.3 shows sample images of subjects from the ND database.

For studying the effect of occlusion using disguise accessories, we utilize the I²BVSD face dataset [8, 9] which is the only publicly available dataset containing images with occlusion. The dataset consists of 681 face images pertaining to 75



Fig. 7.3 Sample images from the ND thermal face database [6]

subjects in both visible and thermal spectra. The utilization of disguise accessories results in varying amount of face occlusion. Depending on the facial regions which are occluded, the dataset is divided into two parts, major disguises (231 images) and minor disguises (307 images). Sample images contained in the dataset are shown in Fig. 7.4. For this research, we utilize only the thermal spectrum images having disguise variations (538 images, 75 subjects).

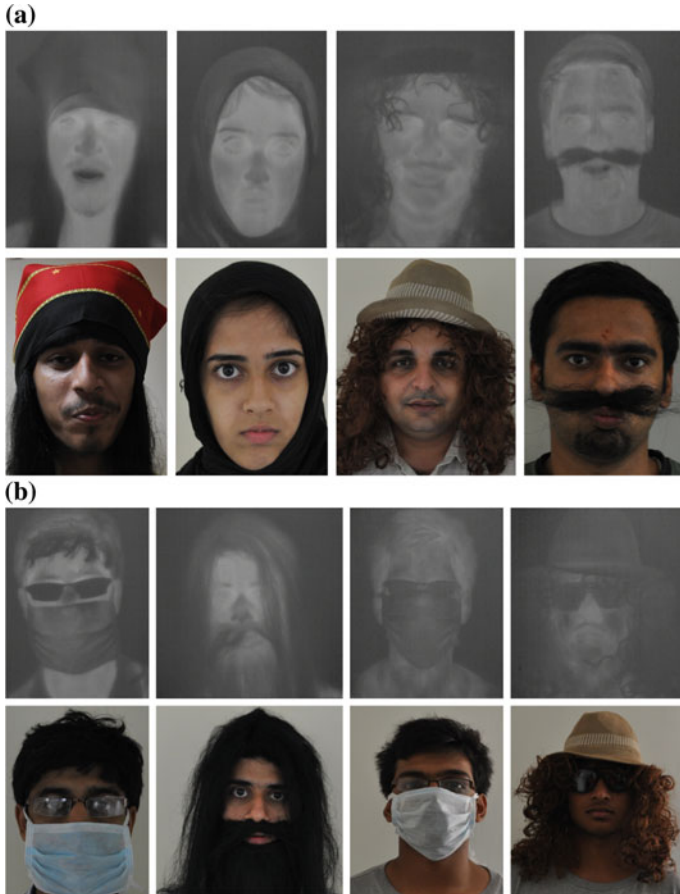


Fig. 7.4 Sample images from the I^2BVSD thermal face database [8, 9]. Face images from subsets pertaining to face occlusion due to **a** minor and **b** major usage of disguise accessories. While I^2BVSD database has images in both visible and thermal spectrum, we have used only thermal images in the experiments

7.3.1 Algorithms

The baseline performance has been established for two face detection algorithms:

1. *Haar-like features* [17, 26] with *cascaded AdaBoost classifier*: Haar-like features are computed from rectangular regions of the image. Every rectangle is divided into multiple non-overlapping sub-rectangles. Pixel intensities of each sub-rectangle are added and the differences of summed intensities of sub-rectangles are used as features. Sub-rectangles are created such that these differences provide coarse information about horizontal, vertical, and diagonal gradients.
2. *Local binary patterns (LBP) features* [16] with *cascaded AdaBoost classifier*: In LBP, the difference between every pixel and its neighbors is computed. The sign of differences is represented using a binary bit. The string of these binary bits for every pixel is converted to decimal. An LBP-coded image representation is obtained by replacing every pixel with its corresponding decimal values. The final feature is represented in terms of histograms obtained from local regions of the LBP-coded image representation.

Algorithm 1 briefly summarizes the cascade boosting face detection approach utilizing Haar/LBP.

Algorithm 1 Summary of cascaded Adaboost classifier.

- Initialize with equal weights for each training sample
 - For each stage
 1. Normalize the weights such that sum of all the weights is one.
 2. Learn a weak classifier corresponding to every single feature. Compute the weighted error of each weak classifier. Also, assign the decision threshold such that false negatives are minimized.
 3. Choose the weak classifier with the minimum weighted error.
 4. Update the weights of correctly classified training samples by multiplying them with the ratio $\beta = \frac{\text{accuracy}}{\text{error}}$ of the current stage.
 - Final strong classifier: A sub-window is classified as face if all the stages of the cascade classifies it as a face region.
-

7.3.2 Evaluation Measures

In the existing literature, performance and effectiveness of face detection algorithms is measured using metrics such as mean square error (MSE). However, these statistics only present the difference between the ground truth and automatically segmented landmark points. In our understanding, MSE is a more useful metric for

landmark detection and it is not very informative if one wants to compare the regions of interest and analyze falsely detected and falsely rejected regions. The objective of a face detection algorithm is to be able to detect the complete region of interest so that it has maximum intersection with the ground truth. To evaluate the performance with respect to this intersection criterion, we propose the following two measures:

- *Ratio with ground truth (RG)*: RG presents the ratio of intersection of the predicted region and ground truth with the area of ground truth segmentation.

$$RG = \frac{Area_{D \cap G}}{Area_G} \tag{7.1}$$

- *Ratio with detection (RD)*: RD presents the ratio of intersection of the predicted region and ground truth with the area of detected region.

$$RD = \frac{Area_{D \cap G}}{Area_D} \tag{7.2}$$

Here, D and G represent the detected and ground truth face regions (rectangles), respectively. The visual interpretation of RG and RD together is shown in Fig. 7.5. High RD along with low RG indicates that while there is a good overlap between

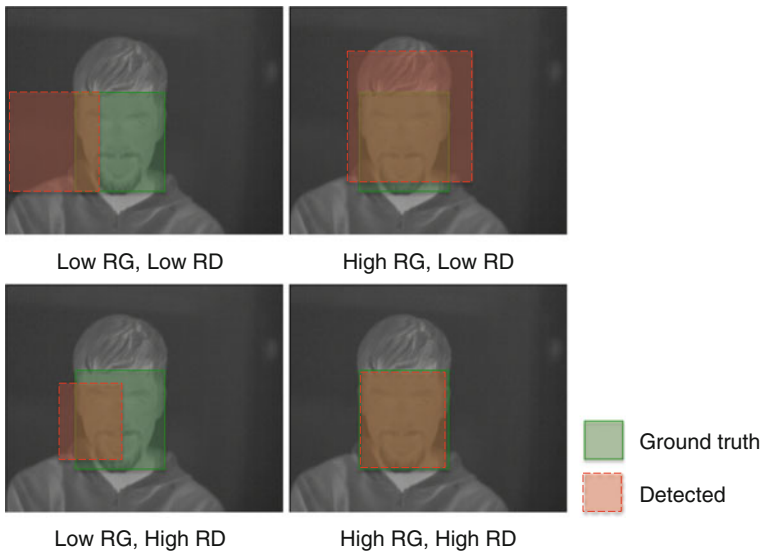


Fig. 7.5 RG and RD measures. High values of RG and RD together ensure a good face detection

the two, a smaller face rectangle is detected compared to the ground truth. Similarly, high RG along with low RD indicates that automatic face detection algorithm has detected a larger face rectangle compared to the ground truth. Their values lie in the range of $[0, 1]$ and for ideal face detection, both should be very close to one. RG and RD can be more effectively used together to analyze the results. In this research, we observe that a threshold of 0.7 for both RG and RD can be used to consider successful face detection.

7.4 Results and Analysis

The performance of the algorithms has been evaluated in four different scenarios:

- *Testing with visible cascade*: The publicly available visible spectrum face detector model is utilized to detect faces in thermal spectrum images. This experiment establishes the baseline for performance evaluation.
- *Learning a model for thermal images*: We learn detectors using thermal face and non-face images. The images are preprocessed using histogram equalization followed by feature extraction using LBP or Haar.
- *Effect of environment and sensor*: The effect of environmental factors such as indoor/outdoor setting and day/nighttime is studied using the IIITD-PSE dataset. We evaluate the effect of sensor interoperability on face detection. This set of experiments is aimed to study the generalizability of face detection models.
- *Effect of occlusion*: In this set of experiments, we study the effectiveness of the learned thermal face detector on occluded faces. Along with LBP and Haar features, a skin detection-based ROI selection approach is also presented.

7.4.1 Testing with Cascade Trained on Visible Images

The first experiment is performed to evaluate the performance of pre-trained Haar cascades (available with OpenCV [4]) on thermal spectrum images. Since Haar cascade is originally trained for visible spectrum, this experiment also provides an understanding about face detection performance with cross-spectral training. For this experiment, the IIITD database (614 images) and the test partition of the ND dataset (2292 images) are utilized as test sets. In all the experiments, when multiple face rectangles are detected in an image, the largest one is considered as the detected rectangle. The graphs of normalized image count verses their RG and RD are shown in Fig. 7.6. The horizontal axis represents RG (or RD), whereas the vertical axis represents the ratio of the number of images having specific RG (or RD) to the total number of images. As it can be seen, very small proportion of face images resulted in high RG or RD values. This shows that pre-trained visible image cascade is not appropriate for face detection in thermal images. No face rectangle is

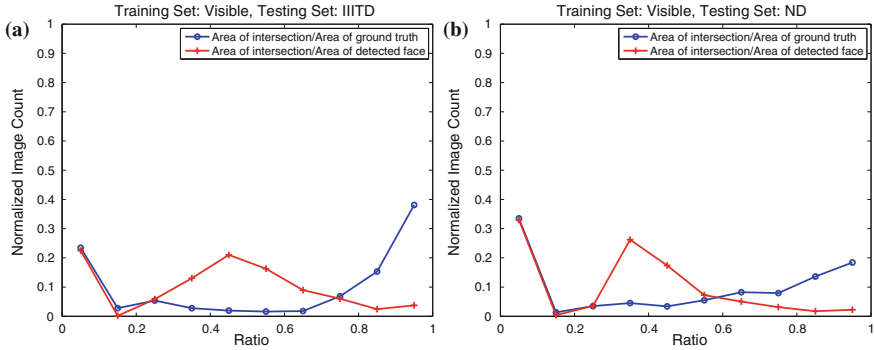


Fig. 7.6 Pre-trained Haar with face detection on the **a** IIITD dataset and **b** ND dataset. *Horizontal axis* represents the value of RG and RD. *Vertical axis* represents the normalized image count with corresponding RG and RD. Normalized image count is computed as $\frac{\# \text{ images with corresponding RG or RD}}{\# \text{ total images}}$

Table 7.3 Summary of face detection results with pre-trained cascade and the cascade trained with thermal images on the IIITD and ND thermal face databases

Experiment	Training set	Test set	Normalized image count		Detection accuracy (RG > 0.7 and RD > 0.7)	Undetected Total
			RG > 0.7	RD > 0.7		
Visible’s Cascade (Haar)	Pre-trained (OpenCV)	IIITD	0.60	0.12	0.05	98/614
		ND	0.40	0.07	0.01	723/2292
Thermal’s Cascade (LBP)	IIITD	IIITD	0.84	0.70	0.62	0/307
		ND	0.78	0.71	0.60	147/2292

detected in 15.9 % images of the IIITD dataset and 31.5 % images of the ND dataset. As shown in Table 7.3, for both IIITD and ND datasets, many images have low RG and RD, which further show poor face detection results.

7.4.2 Learning a Cascade Model for Thermal Faces

Since pre-trained cascade model does not exhibit effective performance, it is important to train the face detection model using thermal data. We utilize face and non-face images captured in thermal spectrum for this task. From the IIITD dataset, 307 randomly selected face images and all the 150 non-face images are used as the training set and testing is performed on the remaining (unseen) 307 images. For ND dataset, training is performed with a predefined set of 159 train images and testing with 2292 images. The LBP cascade model is trained and the results on the testing

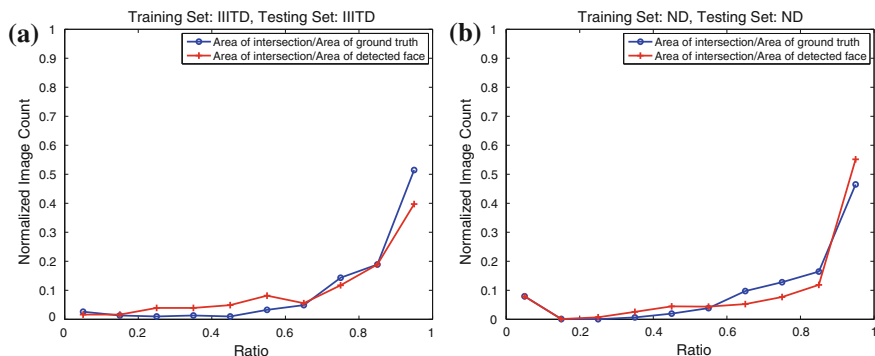


Fig. 7.7 Face detection using LBP cascade learned on the **a** IIITD dataset and **b** ND dataset

database are shown in Fig. 7.7. It can be observed that compared with pre-trained cascade, there is a substantial increase in the number of images with higher RG and RD when cascades trained on thermal images are used. At least one face rectangle is detected in each image of the IIITD dataset, whereas in 6.41 % images of the ND dataset no face rectangle is detected. Further, Table 7.3 also shows that training on thermal images helps in improving face detection results. However, there is a further scope of improvement, as faces are detected reasonably well in only about 60 % images.

7.4.2.1 Learning a Cascade Model from Combined Dataset

One possible way to further improve the face detection performance is to learn the model using data containing large variations. In order to achieve this, we train a model using both the datasets: 307 and 159 images from the IIITD and ND datasets, respectively, comprise the face samples of the training set for this experiment. The cascaded AdaBoost model is trained using LBP features. The results pertaining to this experiment are shown in Table 7.4 and Fig. 7.8a, b. Moreover, Table 7.4 shows that there is a significant improvement in the correct detections ($RG > 0.7$ and $RD > 0.7$ together), with 2 and 3 % improvement for IIITD and ND datasets, respectively. Also the number of undetected faces reduces significantly.

To further reduce the difference between images from the two databases, image histogram equalization is applied. It is our assertion that histogram equalization can help reduce the effect of the sensor- and/or environment-specific variations. Therefore, LBP features are obtained after preprocessing the images using histogram equalization. As shown in Table 7.4 and Fig. 7.8c, d, there is a slight improvement in performance when images are preprocessed using histogram equalization. The detection rate of 0.65 is obtained on both the sets.

On this combined training set, the effectiveness of Haar cascade is also evaluated with histogram equalization preprocessing. For this experiment, the cascaded

Table 7.4 Results of face detection when the model is trained with combined IIITD + ND databases and tested on IIITD and ND thermal face databases

Experiment	Test set	Normalized image count		Detection accuracy (RG > 0.7 and RD > 0.7)	Undetected Total
		RG > 0.7	RD > 0.7		
No preprocessing and LBP features	IIITD	0.89	0.66	0.64	0/307
	ND	0.75	0.76	0.63	15/2292
Histogram equalization with LBP features	IIITD	0.88	0.70	0.65	0/307
	ND	0.77	0.79	0.65	14/2292
Histogram equalization with Haar-like features	IIITD	0.84	0.78	0.70	2/307
	ND	0.83	0.88	0.77	18/2292

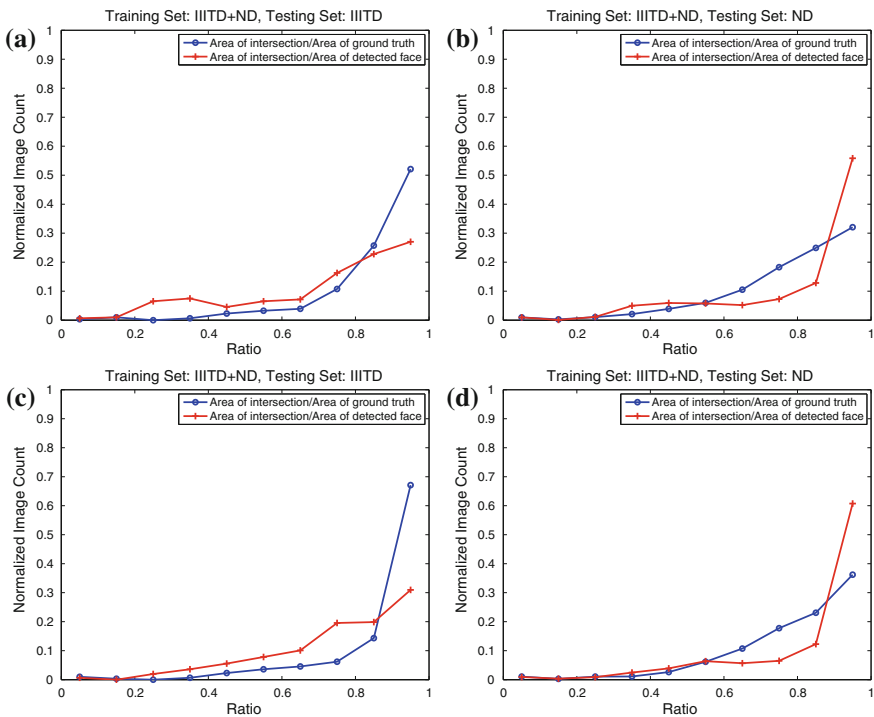


Fig. 7.8 Face detection using LBP cascade with training on combined ND and IIITD dataset and testing on the **a** IIITD dataset and **b** ND dataset. Corresponding result on **c** IIITD dataset and **d** ND dataset when the images are preprocessed using histogram equalization is obtained

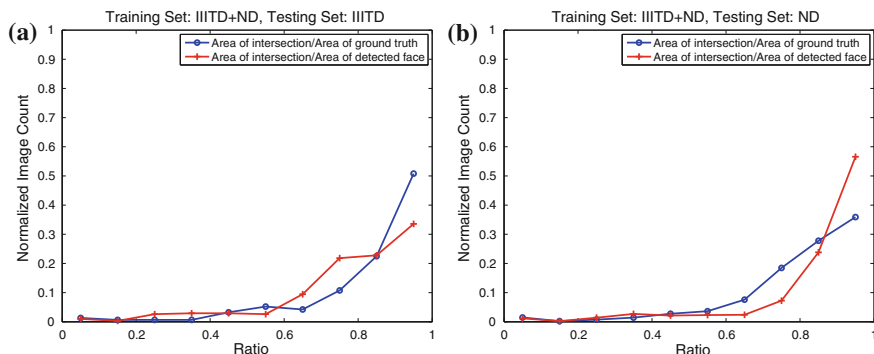


Fig. 7.9 Face detection using Haar cascade with training using histogram equalization preprocessed ND and IIITD dataset and testing on **a** IIITD dataset and **b** ND dataset

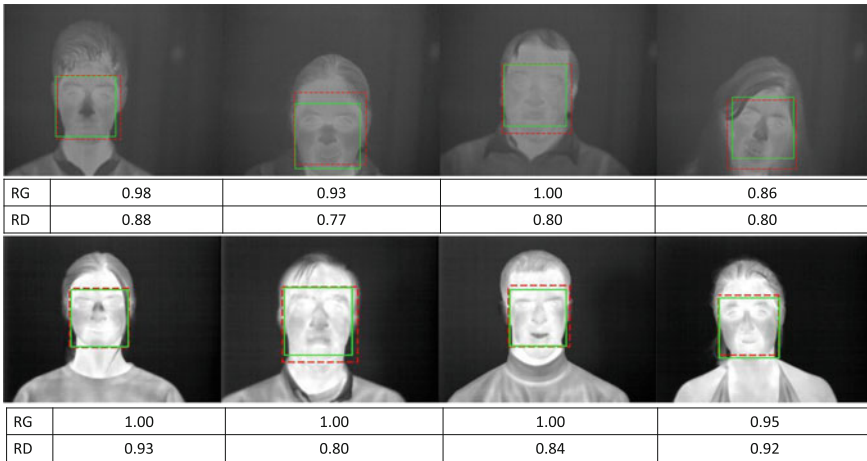
AdaBoost model is learned on the Haar features obtained from histogram-equalized images. The results are shown in Fig. 7.9 and Table 7.4. The results show that Haar cascade with histogram equalization preprocessing performs considerably better in the given scenario by further improving the detection rate to 0.70 and 0.77 on the IIITD and ND datasets, respectively. However, there is a trade-off in terms of training time and accuracy, with Haar cascade requiring more training time and exhibiting better results than that of LBP cascade. Note that there is still scope for improvement as the detection accuracy rate is in the range of 0.70–0.80. Figure 7.10 shows sample detection results of the Haar feature-based cascade learned using combined training set on the IIITD and ND datasets, which yields about 65 % images with successful face detection.

7.4.2.2 Decision Fusion of Haar and LBP Cascades

Since Haar and LBP do not encode the same information, one may expect that both of them should find their applicability in encoding different kinds of variations. Therefore, it is possible that the set of images for which each of the techniques works the best may not be completely overlapping. This is a plausible condition for fusing two techniques and can potentially help further improve the overall accuracy. In order to combine, we follow a simple approach: If face is detected by only one of the techniques, the detected region is taken as the final decision. However, if a face is detected by both the techniques, the following two decision fusion approaches can be applied.

- **Fusion Approach 1:** Out of the two candidate rectangles, select the smaller one. This approach assumes that the detection techniques are prone to overestimating the face rectangle size, thus selecting the smaller candidate rectangle should result in better detection.

(a)



(b)

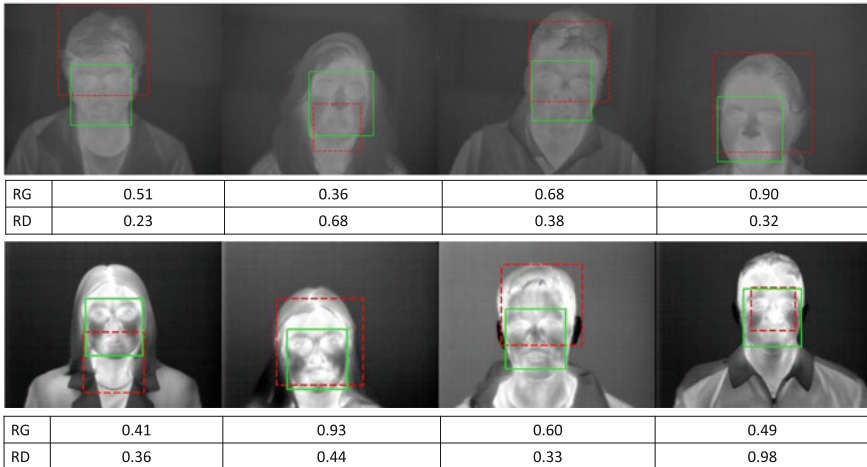


Fig. 7.10 Examples of **a** good ($RG > 0.7$ and $RD > 0.7$) and **b** poor detection results on the IIITD (top row) and ND (bottom row) datasets. Green (solid lines) and red (dashed lines) rectangles represent the ground truth and detected face region (using Haar feature-based cascade learned on combined training set), respectively

- Fusion Approach 2:** Out of the two candidate rectangles, select the larger one. The second approach assumes that the detection techniques are prone to underestimating the size of face rectangle.

Both sets of experiments are performed along with histogram equalization and the results are shown in Table 7.5 and Fig. 7.11. As shown in Table 7.5, in both the

Table 7.5 Summary of face detection results with the proposed fusion approaches on the IIITD and ND thermal face databases

Algorithm	Test set	Normalized image count		Detection accuracy (RG > 0.7 and RD > 0.7)	Undetected Total
		RG > 0.7	RD > 0.7		
Fusion approach 1 (LBP + Haar)	IIITD	0.84	0.85	0.74	0/317
	ND	0.75	0.93	0.73	3/2292
Fusion approach 2 (LBP + Haar)	IIITD	0.88	0.63	0.61	0/317
	ND	0.85	0.75	0.70	3/2292

The training set is created by combining the training sets of both IIITD and ND datasets

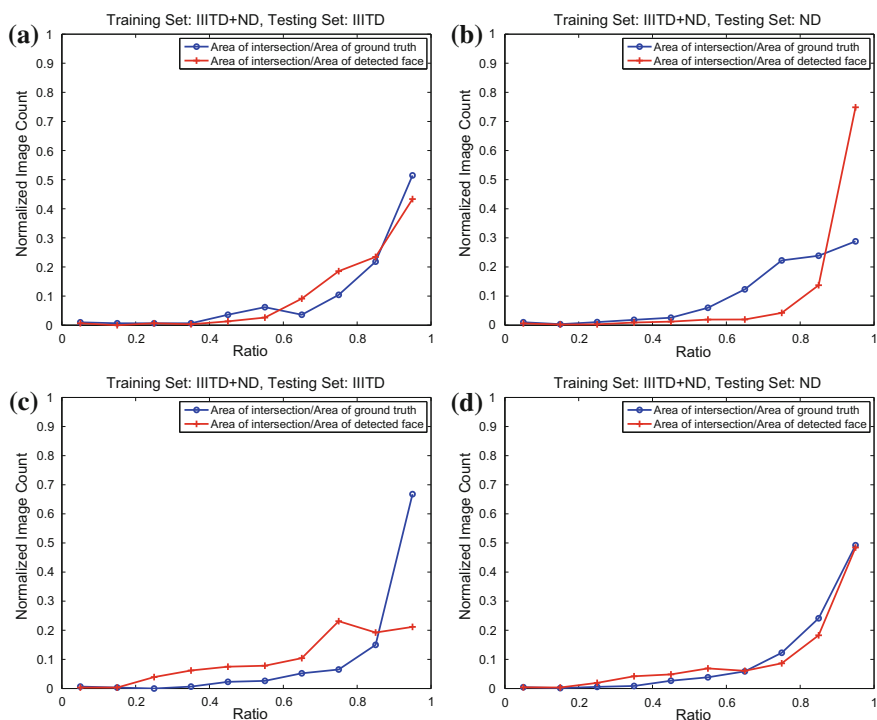


Fig. 7.11 Face detection using fusion approach 1 on the **a** IIITD and **b** ND datasets; and using fusion approach 2 on the **c** IIITD and **d** ND datasets

fusion approaches, at least one face rectangle is detected in all but three images. The results show that the first fusion approach exhibits better results for the IIITD dataset, and fusion has little effect on the ND dataset.

7.4.3 Effect of Sensor and Environment

Since thermal images are dependent on the heat emissivity of the object or surface, they may be affected by environmental aspects such as ambient temperature. These images are also dependent on the type of sensor used, and therefore, the interoperability of sensors can also affect the accuracy of the models learned from one camera. This section studies these two aspects of thermal face detection.

Effect of Day and Night Outdoor Environments We utilize the IIITD-PSE dataset to understand the challenges of thermal face detection in outdoor settings along with the effect of capture during day and night. Figure 7.2 shows sample images of the same person captured in day and night environments. Since fusion approaches yield better results on the IIITD and ND databases, this experiment is also performed with the fusion approaches only. The results of this experiment are reported in Fig. 7.12. It can be observed that the model learned from images captured in indoor settings (IIITD + ND) is not effective on images captured outdoors during daytime (IIITD-PSE-Day). However, as shown in Table 7.6, the results are relatively

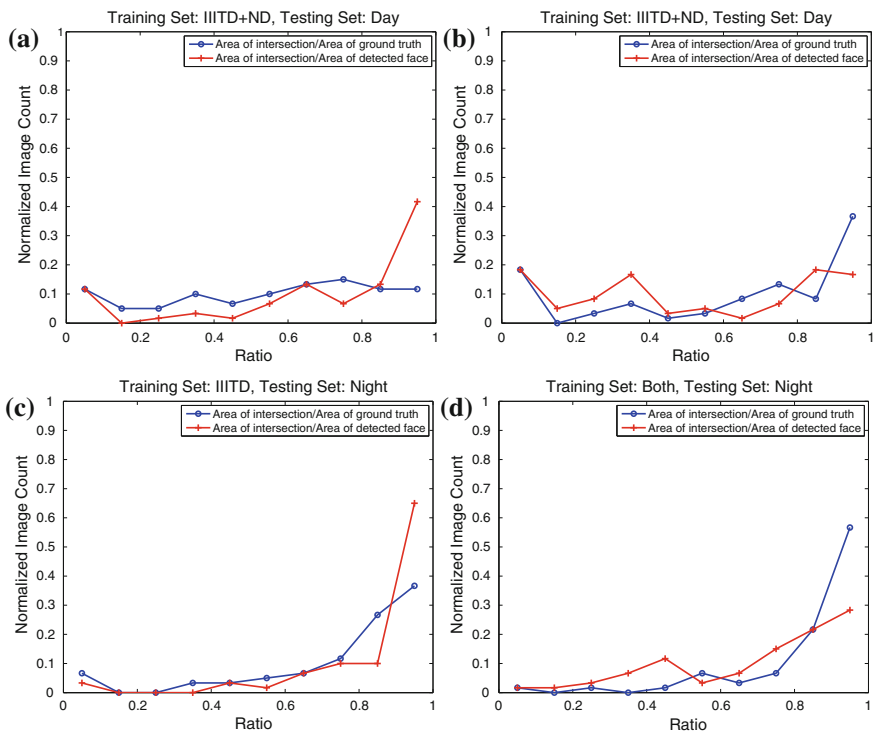


Fig. 7.12 Detection results on IIITD-PSE dataset. Face localization results on IIITD-PSE-Day set using **a** fusion approach-1 and **b** fusion approach-2. Similarly, results on IIITD-PSE-Night set using **c** fusion approach-1 and **d** fusion approach-2

Table 7.6 Summarizing the results of face detection with proposed fusion approaches on thermal images acquired in daytime and nighttime

Experiment	Training set	Test set	Normalized image count		Detection accuracy (RG and RD > 0.7)	$\frac{\text{Undetected}}{\text{Total}}$
			RG > 0.7	RD > 0.7		
Fusion approach 1 (LBP + Haar)	IIITD + ND	IIITD-PSE-Day	0.38	0.61	0.31	0/60
		IIITD-PSE-Night	0.75	0.85	0.70	0/60
Fusion approach 2 (LBP + Haar)	IIITD + ND	IIITD-PSE-Day	0.58	0.41	0.38	0/60
		IIITD-PSE-Night	0.85	0.65	0.65	0/60
Cross-sensor (LBP)	IIITD	ND	0.51	0.65	0.42	211/2292
	ND	IIITD	0.87	0.53	0.47	0/614

The results of cross-sensor face detection experiments are also reported

better for outdoor nighttime images (IIITD-PSE-Night). During daytime, the temperature difference between the skin and the environment is smaller compared to nighttime, which might be affecting the overall contrast of the image. Therefore, daytime outdoor face detection in thermal spectrum needs further research.

Effect of Cross-sensor Training Previous experiments show that if the cascade is trained using the images from same database, it provides good results. However, it is important to have a model which can be utilized across multiple datasets captured using different thermal imaging sensors. As shown in Figs. 7.2 and 7.3, images captured using two different sensors might look quite different. Therefore, a logical step is to examine the challenges due to cross-sensor data (i.e., problem of sensor interoperability).

The cascades trained on individual datasets during the previous experiments are used for detecting faces pertaining to the other dataset. For example, the cascade trained using the ND database is used for detecting faces from the IIITD database and vice versa. The results in Fig. 7.13 show that the model learned from one dataset does not cross-validate well when tested on the other dataset. This may be due to the fact that different datasets include different properties such as capturing environment, set of subjects, imaging resolution, and sensor characteristics.

7.4.4 Effect of Occlusion

In order to study the effect of occlusion, we have performed experiments on the I²BVSD face disguise dataset (thermal spectrum images only). The dataset is divided into two parts, (i) minor disguise and (ii) major disguise. The minor disguise subset consists of images of subjects wearing headgears, hair, and beard extensions which do not cover any of the vital features such as eyes, nose, and mouth. The major disguise subset consists of images of subjects wearing shades,

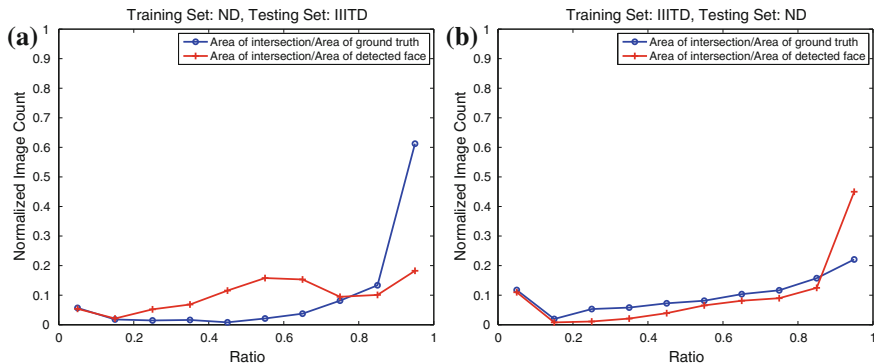


Fig. 7.13 Face detection using LBP cascade with **a** training on ND and testing on IIITD set and **b** training on IIITD and testing on ND dataset

mouth pieces, heavy beards, and/or any accessory which covers one of the vital features. The baseline evaluation is performed with cascaded AdaBoost models learned using combined training set (IIITD + ND). As shown in Table 7.7, detection accuracy for minor disguise subset is about 25 % for LBP and Haar-based cascaded AdaBoost detector. For major disguise subset, the LBP and Haar-based detectors yield 10 and 15 % detection accuracy, respectively. This shows that the cascaded AdaBoost-based approach is not very effective in the presence of occlusions. The corresponding results in terms of RG and RD are reported in Fig. 7.14. If we pose a constrained problem of face localization, i.e., *given that there is a face image, locate it*, we can utilize a skin detection-based approach for approximating facial regions with occlusion variations. Skin color-based region detection has been studied extensively in visible spectrum [11, 13, 20, 22]. However, to the best of our knowledge, skin color detection in thermal spectrum is still unexplored. Skin detection is comparatively easier in thermal spectrum because the heat patterns generated due to body temperature are typically distinct compared to background. Therefore, we present a skin detection-based ROI selection approach.

7.4.4.1 Skin Detection-Based ROI Selection

In order to reduce the number of falsely detected faces, we propose a skin detection-based ROI selection approach as a preprocessing stage to cascaded AdaBoost face detection. The steps involved in the proposed skin detection approach are shown in Fig. 7.15. Further details of skin detection are as follows:

- **Features:** For every pixel, a square neighborhood of $k \times k$ is chosen as the representation. Thus, for every pixel a k^2 dimensional feature vector is obtained; and an image of size $m \times n$ is represented using $mn \times k^2$ feature set. In this work, neighborhood of $k = 3$ is chosen. This feature representation helps encode every skin pixel with respect to its neighborhood.

Table 7.7 Summarizing the results of face detection on the faces occluded using disguised accessories

Experiment	Test set	Normalized image count		Detection accuracy (RG > 0.7 and RD > 0.7)	$\frac{\text{Undetected}}{\text{Total}}$
		RG > 0.7	RD > 0.7		
<i>Without skin detection</i>					
LBP	Disguise minor	0.81	0.33	0.27	0/307
	Disguise major	0.57	0.25	0.10	1/231
Haar	Disguise minor	0.43	0.51	0.26	12/307
	Disguise major	0.31	0.57	0.15	7/231
<i>Only skin detection</i>					
Skin model	Disguise minor	1	0.00	0.00	n/a
	Disguise major	0.99	0.00	0.00	n/a
<i>With skin detection</i>					
LBP	Disguise minor	0.68	0.53	0.42	1/307
	Disguise major	0.20	0.47	0.10	5/231
Haar	Disguise minor	0.39	0.58	0.29	12/307
	Disguise major	0.21	0.65	0.13	0/231
Fusion approach 1 (LBP + Haar)	Disguise minor	0.41	0.64	0.33	0/307
	Disguise major	0.11	0.62	0.08	0/231
Fusion approach 2 (LBP + Haar)	Disguise minor	0.69	0.48	0.38	1/307
	Disguise major	0.31	0.52	0.16	0/231

- **Skin and Non-skin Modeling:** Using the ground truth face and non-face regions, corresponding skin and non-skin pixels are obtained. These distributions of skin and non-skin features are learned from the training data. The distributions essentially capture how the heat patterns of skin and non-skin pixels appear in the local neighborhood. Skin and non-skin distributions are modeled as:

$$f_s(\mathbf{x}) = \mathcal{N}(\mathbf{x}, \mu_s, \Sigma_s),$$

$$f_{ns}(\mathbf{x}) = \mathcal{N}(\mathbf{x}, \mu_{ns}, \Sigma_{ns})$$

where, f_s and f_{ns} denote the probability of \mathbf{x} belonging to skin and non-skin regions, respectively. $\mathcal{N}(\cdot, \mu, \Sigma)$ denotes a normal distribution with mean μ and variance Σ . The training phase includes learning the mean and variance of skin and non-skin distributions.

- **Pixel classification:** A pixel with feature representation \mathbf{x} is classified as skin, if $\log\left(\frac{f_s(\mathbf{x})}{f_{ns}(\mathbf{x}) + \varepsilon}\right) > \varepsilon$ where ε is a very small positive real number.

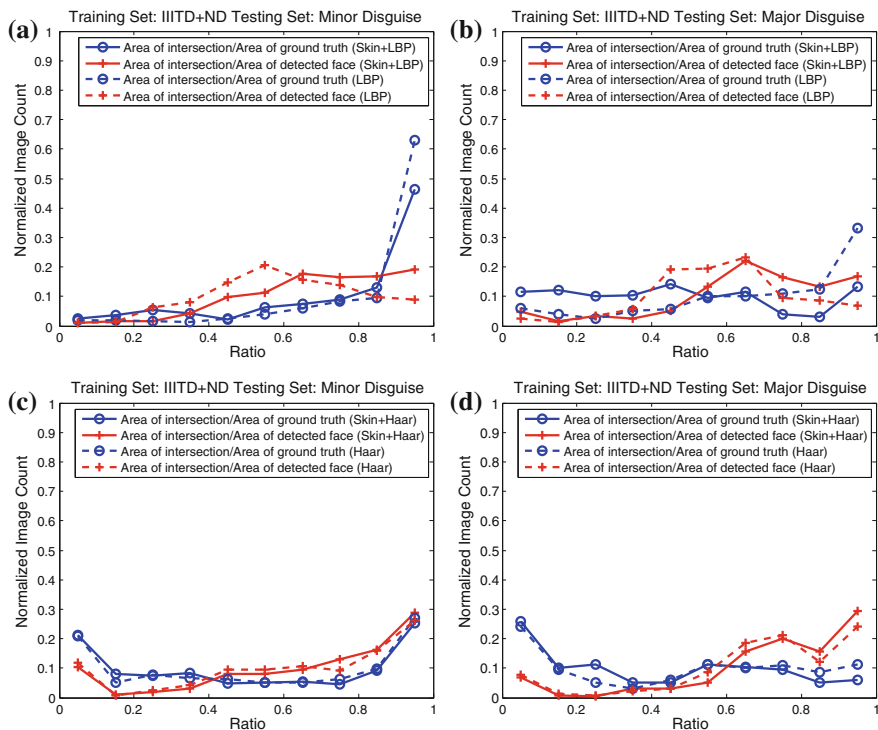


Fig. 7.14 Face detection using LBP cascade on **a** minor and **b** major disguise subsets, and using Haar cascade on **c** minor and **d** major disguise subsets. The results obtained using skin detection-based ROI selection are also reported

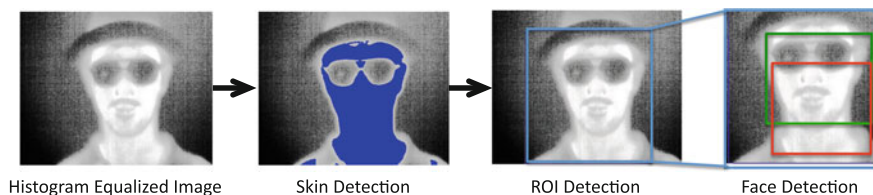


Fig. 7.15 The face detection pipeline with skin detection-based ROI selection. The *blue*, *red*, and *green* rectangles represent the selected ROI, detected face region, and ground truth face region, respectively

- ROI Selection:** At the end of the pixel classification stage, a binary mask is obtained for every image. Although almost all the facial regions are often obtained as skin regions, there may be some holes and/or there may be multiple connected components (due to occlusion). We propose to utilize the largest

connected component of the binary mask, and the corresponding bounding box is utilized as the region of interest. Once the ROI is selected, a cascade classifier is utilized for finding the face location/bounding box.

7.4.4.2 Effectiveness of ROI Selection

Figures 7.16 and 7.17 show some examples of ROI selections and the effectiveness of the ROI selection approach, respectively. It can be observed that the proposed approach effectively rejects non-face regions of the image. As shown in Table 7.7 and Fig. 7.17, the skin modeling-based ROI selection alone, without any further face detector, yields RG of around 100 %. This shows that the ROI is almost always covering the face region. We further applied the AdaBoost cascade, learned in earlier experiments, on the ROI obtained using skin detection. As shown in Table 7.7, Fig. 7.14, and Fig. 7.18 ROI selection helps in improving detection accuracy, especially in case of minor disguise.

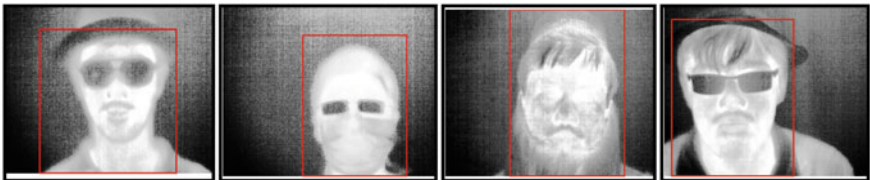


Fig. 7.16 Samples of skin detection-based ROI selection approach. The *red rectangle* represents the ROIs obtained

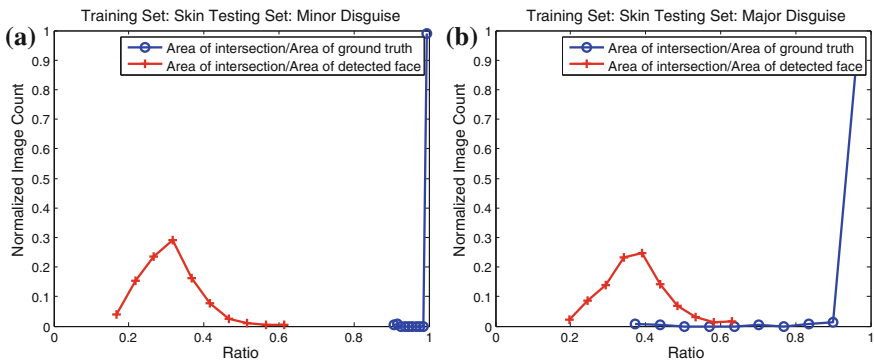


Fig. 7.17 Effectiveness of skin detection-based ROI selection is shown for **a** minor and **b** major disguise subsets. On both the sets, the RG values are comparatively very high for large number of images, suggesting that very little ground truth facial region is discarded in ROI selection

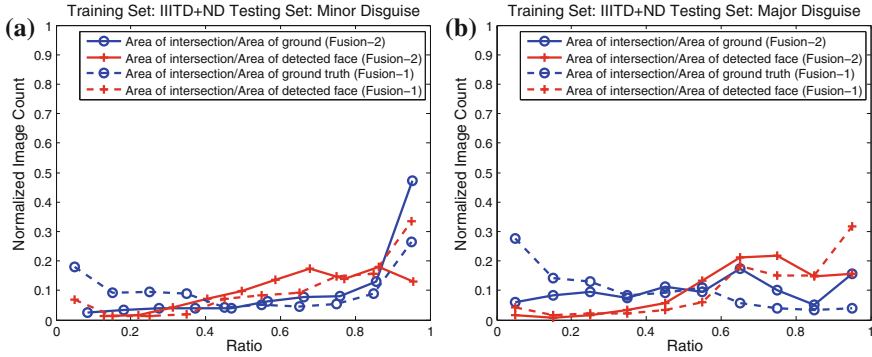


Fig. 7.18 Results of the proposed fusion approaches on **a** minor and **b** major disguise subsets

7.5 Conclusion and Future Research Directions

Thermal face detection has been a relatively unexplored area of research and there are multiple covariates that affect the performance of face detection algorithms. This chapter presents a study to understand the effect of thermal imaging-specific covariates on the performance of face detection in thermal images. There are three contributions of this chapter:

1. We prepared two thermal face databases, namely the IIITD thermal face database that contains 614 face images of 64 subjects and 150 non-face images, and IIITD-PSE database comprising 120 images captured during daytime and nighttime to study the effect of ambient temperature on thermal face detection.
2. We analyzed the performance of three algorithms: AdaBoost-based face detector with LBP features, AdaBoost face detector with Haar-like features, and fusion of the LBP and Haar-like features. The performance is analyzed not only on the IIITD thermal and PSE databases, but also on the Notre Dame and I²BVSD face databases. The use of these two existing databases helps us to understand the impact of interoperability and occlusion on thermal face detection.
3. We propose two new metrics of face localization, RG and RD, which in combination provide the true performance of face detection.

The results show that decision level fusion of Haar-like and LBP features is promising and preprocessing using histogram equalization also helps in improving the detection accuracy. This may point out that preprocessing is one of the key components in addressing environmental covariates in thermal face detection. Further, the results on cross-dataset experiments, indoor–outdoor and day–night variations, and occlusions using facial accessories reveal challenging nature of the problem.

Acknowledgements T.I. Dhamecha is partly supported through TCS Ph.D. research fellowship.

References

1. Bourlai, T., Jafri, Z.: Eye detection in the middle-wave infrared spectrum: towards recognition in the dark. In: IEEE International Workshop on Information Forensics and Security, pp. 1–6 (2011)
2. Bourlai, T., Ross, A., Chen, C., Hornak, L.: A study on using middle-wave infrared images for face recognition. In: SPIE, Biometric Technology for Human Identification IX, Baltimore, USA (2012)
3. Bourlai, T.: Mid-wave ir face recognition systems. SPIE Newsroom Magazine-Defense & Security (2013)
4. Bradski, G.: The OpenCV library. *Dr. Dobb's J. Softw. Tools* **25**(11), 120–126 (2000)
5. Chen, X., Flynn, P.J., Bowyer, K.W.: PCA-based face recognition in infrared imagery: baseline and comparative studies. In: IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 127–134 (2003)
6. Chen, X., Flynn, P.J., Bowyer, K.W.: IR and visible light face recognition. *Comput. Vis. Image Underst.* **99**(3), 332–358 (2005)
7. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
8. Dhamecha, T.I., Nigam, A., Singh, R., Vatsa, M.: Disguise detection and face recognition in visible and thermal spectrums. In: IEEE International Conference on Biometrics, pp. 1–8 (2013)
9. Dhamecha, T.I., Singh, R., Vatsa, M., Kumar, A.: Recognizing disguised faces: human and machine evaluation. *PLoS one* **9**(7), e99212 (2014)
10. Haralick, R.M., Shanmugam, K., Dinstein, I.H.: Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **6**, 610–621 (1973)
11. Hsu, R.L., Abdel-Mottaleb, M., Jain, A.K.: Face detection in color images. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(5), 696–706 (2002)
12. IEEE OTCBVS WS Series Bench; DOE University Research Program in Robotics under grant DOE-DE-FG02-86NE37968; DOD/TACOM/NAC/ARC Program under grant R01-1344-18; FAA/NSSA grant R01-1344-48/49; Office of Naval Research under grant #N000143010022
13. Kovac, J., Peer, P., Solina, F.: Human skin color clustering for face detection, vol. 2. IEEE (2003)
14. Li, S.Z., Yi, D., Lei, Z., Liao, S.: The CASIA NIR-VIS 2.0 face database. In: IEEE Conference on Computer Vision and Pattern Recognition—Workshops, pp. 348–353 (2013)
15. Li, S.Z., Chu, R., Liao, S., Zhang, L.: Illumination invariant face recognition using near-infrared images. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(4), 627–639 (2007)
16. Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.Z.: Learning multi-scale block local binary patterns for face recognition. In: *Advances in Biometrics*, pp. 828–837. Springer, Berlin (2007)
17. Lienhart, R., Maydt, J.: An extended set of Haar-like features for rapid object detection. In: IEEE International Conference on Image Processing, vol. 1, pp. I-900–I-903 (2002)
18. Martinez, B., Binefa, X., Pantic, M.: Facial component detection in thermal imagery. In: IEEE Conference on Computer Vision and Pattern Recognition—Workshops (2010)
19. Osia, N., Bourlai, T.: A spectral independent approach for physiological and geometric based face recognition in the visible, middle-wave and long-wave infrared bands. *Image Vis. Comput.* **32**(11), 847–859 (2014)
20. Phung, S.L., Bouzerdoum, A., Chai, D.: A novel skin color model in YCbCr color space and its application to human face detection. In: IEEE International Conference on Image Processing, vol. 1, pp. I-289 (2002)
21. Selinger, A., Socolinsky, D.A.: Face recognition in the dark. In: IEEE Conference on Computer Vision and Pattern Recognition-Workshops (2004)
22. Singh, S.K., Chauhan, D., Vatsa, M., Singh, R.: A robust skin color based face detection algorithm. *Tamkang J. Sci. Eng.* **6**(4), 227–234 (2003)
23. Socolinsky, D.A., Selinger, A.: Thermal face recognition in an operational scenario. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. II-1012–II-1019 (2004)

24. Socolinsky, D.A., Neuheisel, J.D., Priebe, C.E., De Vinney, J., Marchette, D.: Fast face detection with a boosted CCCD classifier. Technical Report, Johns Hopkins University (2002)
25. Trujillo, L., Olague, G., Hammoud, R., Hernandez, B.: Automatic feature localization in thermal images for facial expression recognition. In: IEEE Conference on Computer Vision and Pattern Recognition—Workshops, pages 14 (2005)
26. Viola, P., Jones, M.J.: Robust real-time face detection. *Int. J. Comput. Vision* **57**(2), 137–154 (2004)
27. Wang, S., Liu, Z., Lv, S., Lv, Y., Wu, G., Peng, P., Chen, F., Wang, X.: A natural visible and infrared facial expression database for expression recognition and emotion inference. *IEEE Trans. Multimedia* **12**(7), 682–691 (2010)
28. Wang, S., Liu, Z., Shen, P., Ji, Q.: Eye localization from thermal infrared images. *Pattern Recogn.* **46**(10), 2613–2621 (2013)
29. Zhang, Z., Yi, D., Lei, Z., Li, S.Z.: Regularized transfer boosting for face detection across spectrum. *IEEE Signal Process. Lett.* **19**(3), 131–134 (2012)