Thermal-Visual Facial Feature Extraction Based on Nostril Mask

MOHD NORZALI Haji Mohd*, Masayuki KASHIMA**, Kiminori SATO** and Mutsumi WATANABE**

Department of Information Science and Biomedical Engineering, Graduate School of Science and Engineering, Kagoshima University, Korimoto 1-21-40, Kagoshima, 890-0065, Japan
mnorzali@gmail.com, kashima@ibe.kagoshima-u.ac.jp, kimi@ibe.kagoshima-u.ac.jp, mutty@ibe.kagoshima-u.ac.jp

*Department Of Computer Engineering, Faculty Of Electrical and Electronic Engineering, University Tun Hussein Onn Malaysia (UTHM), Parit Raja, Batu Pahat, Johor, Malaysia.
norzali@uthm.edu.my

Abstract

This paper aims to present facial features extraction by integrating 2 different sensors that will be used in the estimation of internal mental state. Thermal infrared and visible camera are being used in the stimulus experiment by measuring three facial areas of sympathetic importance which is periorbital, supraorbital and maxillary through purely imaging means in thermal infrared spectrums. The development of Automatic Thermal Face, Supraorbital, Periorbital, Maxillary and Nostril Detection to be used for estimation of internal state is also presented. Several faces samples were taken in real time in our experimental setup to measure the effectiveness of our method. Almost 98% of correct measurement of ROI and temperature was detected.

In this paper, a new method for detecting facial feature in both thermal and visual is also presented by applying Nostril Mask, which allows one to find facial feature namely nose area in thermal and visual. Graph Cut algorithm is applied to remove unwanted ROI and correctly detect precise temperature values. Extraction of thermal-visual facial feature images is done by using Scale Invariant Feature Transform (SIFT) Feature detector and extractor to verify the method of using nostril mask. Based on the experiment conducted, it shows 88.6% of correct matching.

KEY WORDS
Face recognition, Thermal face image analysis, Facial feature extraction, SIFT

1. Introduction
Various methods for internal state measurement such as mental stress have been previously proposed which utilizes the changes of physiological quantities such as blood pressure, heart rate, salivary amylase and electromyography (EMG) [1]. These quantities can be measured invasively by the use of expensive instruments. In this paper we introduce an integrated non-invasive measurement through purely imaging means. The proposed booth setup is equipped with Thermal IR and Visual Camera Sensors. A monitor and a speaker are also installed to provide mental and acoustic stimulus to monitor changes in facial sign and brain waves. These experiments will be done at a later stage.

![Fig. 1. Overall System.](image)

Face recognition system based on visual images have reached significant level of maturity with some practical success. However, the performance of visual face recognition may degrade under poor illuminations conditions, for subjects of various skin colour and the changes in facial expression. The use of infrared in face recognition allows the limitations of visible face recognition to be overcome. However, infrared suffers from other limitations like opacity to glasses. Multi modal fusion comes with the promise of combining the best of each modalities and overcoming their limitations [2].

Facial feature in thermal is difficult to locate because of the poorer contrast between the features and the face in IR images. In this paper, one common feature that can be automatically located in both thermal-visible domains is the nostril area. This is due to the changes of temperature in nostril area in thermal domain and the higher contrast value in visual domain. An outline of our approach is shown in Fig. 1. The combined relationship
includes two parts: 1) the relative position between the thermal and visible camera that are calibrated using special calibration board. 2) The relative position and head pose based on nostril mask in thermal and visible.

For the time being, we are focusing on facial feature extraction in thermal IR and visual camera based on nostril mask as a combined relationship. We also analyse the use of SIFT descriptor as feature matching in thermal and visual domain to verify our method using nostril mask.

2. Proposed System

2.1. Facial Feature Extraction

2.1.1. Thermal Domain

Thermal face image analysis has many applications such as sensation evaluation and face recognition. Facial feature extraction in the IR image is an essential step in these applications. Certain facial areas such as periocular, nasal, cheeks and neck region produce different thermal patterns for different activities or emotions. Skin temperature of facial features, such as the nose and forehead, could be an effective indicator in objectively evaluating human sensations such as stress and fatigue. Most existing approaches manually locate the facial feature in IR image or the subject are required to wear marker, as it is hard to automatically locate facial features, even for the obvious features such as the corners of the eyes and mouth. This problem is caused by poorer contrast between the features and the face in IR images.

In our experiment, we have found out that by using the proposed face detection using Cascade structure with Haar-like features. Then 3 ROIs was detected based on detected face ratio. The collaboration of ROI with temperature value was done based on the relationship as in (1). This relationship is concluded through several experiments done to ensure the accuracy. The average brightness value located around these 3 main areas is then converted to the temperature value.

\[ \text{Temperature} = 0.0819 \times \text{GrayLevel} + 23.762 \]  

(1)

Fig.3. Flowchart of face detection and temperature acquisition in thermal domain.

Next, we assume that the centroid of the detected face area as nose area and nostril mask is used to recalculate the detected area and map into the facial thermal image (Fig.4).

We then analyze the detected thermal face with the 3 ROI with sympathetic importance for person wearing spectacles and without it. We found out that there are unwanted ROI such as spectacles and hair which can be excluded so that temperature reading can be done precisely (Fig. 5).
To overcome this problem, unwanted ROI is removed by using Graph Cut method (Fig.6). The average differences before and after cropping is about 0.3-0.6 °C.

Fig.6. The unwanted ROI removed using Graph Cut.

2.1.2. Visual Domain

The same approach for locating face image in visual face was also done. At first we detect the face using Viola and Jones’ Boosting algorithm and a set of Cascade structure with Haar-like features. Next, we assume that the centroid of the detected face area as nose area and nostril mask is used to recalculate the detected area and map into the facial image (Fig.7).

Fig. 7. Proposed Automatic Facial and Nostril Detection.

2.2. Image Registration of IR and Visible

2.2.1. Calibration Board

The relative position between the IR and visible cameras is calibrated by using the special calibration board suggested previously [11]. With some small adjustment and preparation where cold fever plasters are attached to the back of the calibration points on a chessboard, the calibration points can be reliably located in thermal IR and visible domain.

One of the common strategies to simplify correspondence problem between IR and visible domain is to exploit epipolar geometry (Fig.8 (b)). The constraint of this method is we consider the stereo camera is modeled as pin-hole camera and lens distortion is not taken into consideration. In this paper, we also tried to propose the use of feature matching as a method for registration of IR and visible because of this constrain. At first, the relative position between thermal-visible stereo cameras is calculated using calibration board (Fig.8 (a)) [14]. The geometry of the visible and thermal IR cameras with respect with world coordinate system is shown in Fig.8 (c). The origin of the world coordinate system is defined to be on the calibration board. Then, the inner parameters of the camera, the rotation and the translation transforms can be obtained by decomposing the transform matrix of the two cameras respectively.

Fig.8. (a) Calibration board (b) Computing fundamental matrix and drawing epipolar lines from Thermal-Visual stereo calibrated images. (c) Initial calibration and fusion of the IR and visible-spectrum sensors.

3. Experiment and Results

In the experiment conducted, Scale Invariant Feature Transform (SIFT) is employed to extract facial features in both thermal and visual and to verify our algorithm (Table 1). SIFT features are features extracted from images to help in reliable matching between different views of the same object [12]. The extracted features are invariant to scale and orientation, and are highly distinctive of the image. They are extracted in four steps. The first step computes the locations of potential interest points in the image by detecting the maxima and minima of a set of Difference of Gaussian (DoG) filters applied at different scales all over the image. Then, these locations are refined by discarding points of low contrast. An orientation is then assigned to each key point based on local image features. Finally, a local feature descriptor is computed at each key point. This descriptor is based on the local image gradient,

<table>
<thead>
<tr>
<th>With Glasses</th>
<th>Without Glasses</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

### Table 1

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
</tbody>
</table>

Fig. 5. Detected thermal face with and without spectacles.
transformed according to the orientation of the key point to provide orientation invariance. Every feature is a vector of dimension 128 distinctively identifying the neighborhood around the key point.

<table>
<thead>
<tr>
<th>Feature Point (Nostril Mask in Thermal)</th>
<th>Feature Point (Facial in Visual)</th>
<th>Sift Feature Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="Image62x580to98x609" alt="Image" /></td>
<td><img src="Image63x537to98x565" alt="Image" /></td>
<td><img src="Image64x628to96x655" alt="Image" /></td>
</tr>
<tr>
<td><img src="Image62x580to98x609" alt="Image" /></td>
<td><img src="Image63x537to98x565" alt="Image" /></td>
<td><img src="Image64x628to96x655" alt="Image" /></td>
</tr>
<tr>
<td><img src="Image62x580to98x609" alt="Image" /></td>
<td><img src="Image63x537to98x565" alt="Image" /></td>
<td><img src="Image64x628to96x655" alt="Image" /></td>
</tr>
</tbody>
</table>

At first, the nostril area in thermal IR is detected using our pair calibrated Thermal –Visible camera. Then, the face part in visible domain is detected and SIFT feature point is calculated in both domain (Table 1). The feature matching is done and recorded as in Table 2.

Table 1. Sift Feature Matching

<table>
<thead>
<tr>
<th>Frame</th>
<th>Processing Time [ms]</th>
<th>Feature Point 1 (TH)</th>
<th>Feature Point 2 (VI)</th>
<th>Correct Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>170.759</td>
<td>2</td>
<td>157</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>168.443</td>
<td>2</td>
<td>93</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>171.866</td>
<td>11</td>
<td>106</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>162.892</td>
<td>2</td>
<td>81</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>170.921</td>
<td>2</td>
<td>168</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>154.105</td>
<td>1</td>
<td>168</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>160.873</td>
<td>1</td>
<td>114</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>153.269</td>
<td>1</td>
<td>112</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>170.35</td>
<td>1</td>
<td>146</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>170.972</td>
<td>1</td>
<td>157</td>
<td>100</td>
</tr>
</tbody>
</table>

At first, the nostril area in thermal IR is detected using our pair calibrated Thermal –Visible camera. Then, the face part in visible domain is detected and SIFT feature point is calculated in both domain (Table 1). The feature matching is done and recorded as in Table 2.

Table 2. Results of Matching Features in Thermal-Visible.

4. Conclusion

We have developed an Automatic Thermal Face, Supraorbital, Periorbital, Maxillary and Nostril Detection to be used for estimation of internal state. Several faces samples were taken in real time in our experimental setup to measure the effectiveness of our method. Almost 98% of correct measurement of ROI and temperature was detected. Feature matching experiment was also done to measure the effectiveness of our method of using nostril mask and shows positive results. We adopt SIFT feature point extraction and matching to verify our method of nostril masking. Experiment showed that this method shows 86% of correct matching. To further improve its accuracy, it requires more accurate samples.

Acknowledgements

This work was partially funded by the Japan Society for the promotion of Science (JSPS), Kagoshima University and UTHM. We would like also to give special thanks to the Laboratory member for their invaluable inputs and assistance.

References