# Non-intrusive Car Driver's Emotion Recognition Using Thermal Camera

Abhiram Kolli[1], Alireza Fasih[2], Fadi Al Machot[3], Kyandoghere Kyamakya[4]

*Transportation Informatics, University of Klagenfurt*
*Universitätstrasse 65-67, A-9020, Klagenfurt-Austria*

1akolli@edu.uni-klu.ac.at

[2]alireza.fasih@uni-klu.ac.at
[3]fadi.alMachot@uni-klu.ac.at

[4]kyandoghere.kyamakya@uni-klu.ac.at

***Abstract*— Several intrusive and non-intrusive techniques have been proposed in the past to monitor car driver's emotions but very little light was shed on using thermal cameras for such applications. This paper details with one such system that uses a single infrared thermal camera. Such a camera was used to overcome the issues pertaining to usage of single audio/visual sensors. Fusion of the outputs with the audio/visual sensors could provide a better ADAS (Advanced Driver Assistance System). The paper also details face detection techniques in each thermogram prior to emotion recognition.**

***Keywords—Infrared thermal camera; ADAS; thermogram; car driver's emotion recognition; face detection; classification of emotions***

## I. INTRODUCTION

Aspiration to provide better safety and comfort to car drivers has greatly influenced research in the field of ADAS Though there are several systems which alert driver of a possible undesirable event (possibility of collision with another vehicle or persons, a steep turn ahead, exceeding speed limit etc.), fatal accidents are reported daily. This could be due to lack of penetration of ADAS into every vehicle and it also indicates that more factors are necessary to be considered in order to provide safer and driver friendly experience. One such factor is to monitor driver's emotions which influence driving.

Data provided by analysts indicates that road accidents have great socio-economic impact across the world [1-6]. In order to improve traffic safety researchers focused on developing ADAS applications as listed in the references [7-9]. Several driver monitoring systems (fatigue/emotions) have been proposed in the past which include intrusive and non-intrusive techniques [10-21].

Illumination and pose variations are considered as major concerns in facial emotion recognition [22-27]. Several compensation techniques were proposed to overcome these issues however, were successful for face recognition in partly lightened faces and not for facial expression recognition (FER) [27]. Attempts were made to implement FER in [24], [28], [29]. However these were not focused for driver's emotion recognition/monitoring. They lack the region of interest (ROI, in this case face detection) while processing, which is crucial for environment such as in a car (a possibility of another person behind/beside the driver). Apart from face detection, this paper tries to classify more emotions. These facial emotions were considered on the basis of Ekman's basic emotions (anger, disgust, fear, joy, sadness, surprise) classification [30].

## II. RELATED WORK

Study of facial skin temperature measurement techniques and baseline characteristics of facial skin temperature has led to better understanding of the facial skin regions [31-33]. Several attempts were made to implement FER [28], [29], [24]. These techniques were not optimal for a system deployed in a car (another person can be seated behind and/or beside the driver) due to lack face detection technique before feature extraction. Entire scene was considered for feature extraction by these techniques. The process of feature extraction detailed by Khan *et al.* was that each thermogram was divided into grids of squares using CMView Plus software. Square grids of 128 were considered to be optimal; however realignment of grids led to the selection of 75 square grids. To the obtained grids, the highest temperature in each grid was recorded and that recorded value was considered as thermal intensity value (TIV). The TIVs were considered as the extracted features. Linear discriminant analyses (LDA) was applied on the extracted features in order to performance the classification of emotions.

The authors reported better results were achieved after reducing the dimensionality of discriminant space by invoking principal component analyses (PCA) and to the resulting values LDA was invoked. The technique proposed by Khan *et al.* [24] showed 66.28% of successful classification. During cross validation test and 56% of successful classification rate was reported during a person independent test [34]. If this technique is adapted into car/automobile the performance of the classifier could worsen as the input image consists of multiple faces.

## III. HARDWARE AND DATABASE CREATION

An infrared thermal camera ("PathFindIR") from FLIR systems with spectral band 8-14µ was used in this project. The output of the camera was connected to a digitizer; the digitizer converted analog output of the camera to digital output. Thus obtained output was connected to a personal computer with 1.50 GHz Intel Core2 Duo, 2 GB RAM.

Six evoked emotions (angry, disgust, fear, happy, sad and surprise) of each participant were taken in a controlled environment. Temperatures of the environment were within the range of 19 °C to 22 °C. Each image was placed in corresponding database.

The database consists of 35 volunteers within the age group of 22 - 27 years. The participants were mainly from Europe and Asia.

## IV. Face Detection and Feature Extraction

Face detection in a visual ray camera image has its prominence in driver monitoring systems. Algorithms were developed to provide reasonably detection rates using haar-like features, skin color etc. However this is not the case with the face detection using thermal camera. Research on face detection using a thermal camera is yet to be explored for better face detection rates. Such attempts were made in this paper.

Features were extracted from the detected face. Several feature extraction techniques such as the TIVs, radon transform, wavelet transform, radon + wavelet transform and histogram oriented gradients (HOG) were tested. Issues were encountered during test these features.

### *A. Face Detection*

Three different algorithms were developed after performing careful analyses of hundreds of thermograms to identify the face region. The output of the three algorithms were fused using a pixel by pixel AND operation. Implemented face detection algorithms for an input image (Fig. 1) are listed below:

- Color based detection (Fig. 2)
- Region growing (Fig. 3)
- Morphological operations based (Fig. 4)

As a final face region (Fig. 5), fusion of images (Fig. 1, Fig. 3, Fig. 4) was considered. Ellipse in the fig. 6 indicates the orientation of the face. This orientation concept of face was considered for context aware programming. Algorithm worked for scenarios where there were more than one person as seen in Fig. 7. However, issues of false detection were also identified as shown in Fig. 8.

1) *Color based face detection:* This is an empirical approach in which each thermogram was carefully analyzed in order to obtain a segmented facial region. As a part of the analyses, each channel (RGB channels of the RGB color space) of several pixels related to the same object was analyzed. Such a pixel level analyses of the entire scene has led to design of a classifier that has successfully segmented face region.

Given an input image, Im(x,y), its corresponding R, G and B channel images are $Im_R(x,y)$,$Im_G(x,y)$ and $Im_B(x,y)$ respectively. Intensity of a pixel located at (n,m) in $Im_R(x,y)$,$Im_G(x,y)$ and $Im_B(x,y)$ images are given by $I_R(n,m)$, $I_G(n,m)$ and $I_B(n,m)$ respectively. Th1 and Th2 are the thresholds that have been obtained empirically.

Algorithm of the classifier is given below:

- Consider R, G and B channel images separately
- Check intensity of each pixel corresponding to the individual channel image
- If $Im_R(x,y)$ & $Im_B(x,y) > Th1$ & $Im_G(x,y) > Th2$ then assign the pixel value as "1" else assign the pixel value as "0".

Now a considerable blob of the face region was obtained along with the few other blobs which satisfy the criteria. Post processing of obtained image results in a blob that consists of face region. This face region may be accompanied (depends on the background/environment of the scene) by few unwanted regions. The resulting image, ImC(x,y), was later on fused at pixel level with other images obtained from region growing and morphological operations.

2) *Region growing based face detection:* The basic procedure in this approach was to obtain the segmented region of face and its corresponding boundaries. Thus obtained resulting image, ImR(x,y), was fused with other face detected images.

Region growing is a procedure that group pixels or sub regions into larger regions based on predefined criteria for growth [35]. As an initial step to this approach set of seed points are considered. Given the set of seed points, region is grown until their neighboring pixels satisfy the criteria like gray level, color etc. This is an iterative process where each seed pixel grows iteratively until every pixel is processed and thereby forms different regions whose boundaries are defined by closed polygons [36]. Given the input image Fig.1, Fig.3 illustrates resulting image of region growing.

Following considerations have been taken care in order to obtain desirable results:

- Selection of initial seeds that represent the regions and properties of face were considered from the resulting image of the color based face detection technique.
- Apart from considering the properties of face region pixels connectivity information was also considered.
- The threshold value which determines the "similarity" was carefully chosen in such a way that the region growing selects face. Similarity refers to the gray level observed between two spatially adjacent pixels or average gray level of a set of pixels, which will yield different regions [36].

3) *Morphological operations based face detection:* Segmentation of face region was performed by background, foreground estimation and using post processing techniques. Thus resulting blob's properties are obtained. The resulting image, ImM(x,y), was further fused with the other resulting images (Fig. 2 and Fig. 3) mentioned in this paper.

Statistical analyses performed on each of the resulting images (around 70 images) facilitated the selection of structuring element's radius. Careful selection of structuring elements has led to better background and foreground detection. The post processing techniques that were performed led to better face region segmentation.

4) *Fusion of Resulting Images:* The resulting images obtained from face detection techniques may consist of unwanted blob regions which have to be eliminated in order to obtain a better face region. To obtain a better face region the following pixel level operation has been performed:

- Consider the resulting images ImC(x,y), ImR(x,y) and ImM(x,y) separately.
- Check the value of each pixel of the resulting images
- If ImC(x,y) & ImR(x,y) & ImM(x,y) = 1 then the resulting fusion image's corresponding pixel was considered as "1" else the corresponding value was assigned "0".

Given the resulting fused image, ImF(x,y), the face region can be represented in the original input image, Im(x,y), by analyzing the regional properties of face blob. The resulting image is shown in Fig. 6. The representation of face region is given by the bounding box while the ellipse represents the orientation of the face (orientated towards left or right). The face region that is represented by bounding box is cropped to obtain the features. This cropped image is referred as face image in this paper.

Figure 2. Output of color based detection.


Figure 1. Input image.

Figure 4. Output of morphology based detection.

Figure 3. Output of region growing based detection.

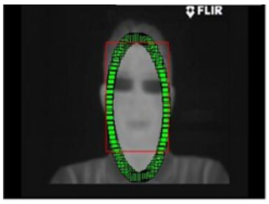


Figure 6. Output of face detection algorithm.

Figure 5. Output of fusion.


Figure 8. False detection of face.


Figure 7. Detection of face.

## B. *Feature Extraction*

Several non-linear features as mentioned in [37] have been extracted apart from the TIVs for the purpose of analyses. Feature vectors that have been extracted from the final detected face image were given below:

- TIVs
- Radon transform
- Wavelet transform
- Radon + wavelet transform
- Histogram oriented gradients (HOG)

Each feature extraction method was tested for its consistency with the training database. However, extraction of such features was computationally intense [37]. Issues related to creation of training databases for several feature vectors have been identified due to the variation of the face size obtained from face detection algorithm. This issue was also observed while testing the system. Such issues were encountered during the creation of TIVs. Another issue encountered with the creation of TIVs was due to the inconsistency of size of face image of different volunteers to that of actual square grids mentioned in [24] and [34]. To overcome such issues each face image was resized. However, such an attempt induced an error at the classification section. This has led to poor performance of classifier.

HOG features were extracted from the face image and tested for their performance. Usage of these features for classification provided better results compared to that of the TIVs. This was due to the consistency in the size of feature vectors (81 X 1. Obtained feature vectors were projected to a higher feature space. Thus obtained feature vectors were considered for classification.

# V. EMOTION CLASSIFICATION

Classification of feature vectors into six different emotions was carried out using modified Hausdorff distance.

## A. *Modified Hausdorff Distance*

Hausdorff distance measures proximity than exact superposition and the distance can be calculated without explicit pairing of points in their respective data sets [38]. These provide a better advantage over the other metrics. Consider two finite point sets A = $\{a_1, \ldots, a_p\}$, and B = $\{b_1, \ldots, b_q\}$. Then for a given sets A and B the Hausdorff distance is calculated by

$$H(A,B) = \max(h(A,B), h(B,A)), \quad (1)$$

where

$$h(A,B) = \max_{a\in A\ b\in B} \min \| a - b \| \quad (2)$$

In (1) $H(A, B)$ is computed trivially in time $O(pq)$ for point sets of size $p$ and $q$ respectively. A better representation of the value is $O((p + q)\log(p + q))$. In (2) the function $h(A,B)$ is known as directed Hausdourff distance from set A to set B and $\|a - b\|$ is the Euclidean distance between two data points. This function ranks each point of set A that is farthest from any point in set B. After ranking each point in a set, the point which has highest rank is used as the measure of distance i.e. the most mismatched point of set A. This ensures that if $h(A,B) = d$, then each point of set A is within distance $d$ of some point of set B and also there exists a point of set A that is exactly distance $d$ from the nearest point of B [38]. Best partial distance between set A and set B is obtained by ranking each point of A by its distance to the nearest point in B and take the $K^{th}$ ranked value. A better realization of directed ($h(A,B)$, $h(B,A)$) and its corresponding $H(A, B)$ distances between two sets A and B was proposed by Dubuisson and Jain. This is known as the modified Hausdorff distance (MHD).

$$h(A, B) = 1/Na \sum_{a \in A} \min_{b \in B} \|a - b\| \qquad (3)$$

In (3), $Na = p$, the number of points in set A. The modified Hausdorff distance is less susceptible to noise.

### B. Classification of Emotions

A database that contains face region images was taken. The face region images represent six evoked basic emotions namely angry, disgust, fear, happy, sad and surprise. A mean face image was calculated for each of the six emotions. To the obtained six mean facial images HOG features were extracted. To a given input image/video frame the face region image was extracted and its corresponding HOG features were extracted. After post processing, the feature set of the input face region image was compared to that of the feature sets of six individual HOG features. The corresponding ranks of the matching have been listed according to their similarity.

The performance estimation of classifier was conducted using confusion matrix which gave satisfactory results.

## VI. Results

We performed the evaluation of face detection algorithms and the emotion classifier. As a part of the evaluation each face detection algorithm was separately tested for its accuracy and consistency.

TABLE I.PERFORMANCE OF FACE DETECTION ALGORITHM

| Algorithm | Performance (%) | |
|---|---|---|
| | *True Positives* | *False Positives* |
| Color based detection | 55 | 45 |
| Region growing based detection | 80 | 20 |
| Morphological operation based detection | 70 | 30 |
| Fusion of resulting images | 70 | 30 |

### A. Face Detection Results

During the testing phase of face detection algorithms, several inconsistencies were observed with each of the algorithm. It is evident from the Table 1 that the performance of color based detection is relatively poor compared to the other detection techniques. This detection technique was highly sensitive to the major variations in temperature of the environment. This undesirable effect was carefully observed during extreme winter and hot summer. One such inconsistency is show in Fig. 9. Due to such inconsistency in detection, this technique was ruled out to be deployed as a face detection algorithm. Even though the performance of face detection algorithm using region growing is best (detection rate) among the other algorithms, it is tricky to choose a threshold that determines proper “similarity”. Improper selection of threshold that determines the “similarity” can be observed in Fig. 10. Performance of morphological operations based face detection has provided reasonably better consistency in detecting face regions. The major issue with this technique is the right selection of structuring element. Performance of fusion of resulting images was relatively reduced due to the major inconsistency in color based face detection algorithm.

Figure 10. False detection of face region using region growing

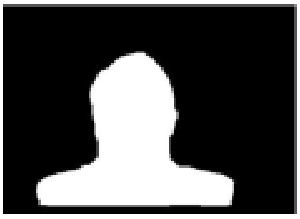

Figure 9. This is a false detection of face region by color based detection.

### B. Emotion Classification Results

“Emotion Face Database” consisting of individual frontal faces that are extracted using the face detection algorithms was considered as a standard database for testing the classifier. It is evident from the Table II that classifier performed better in classifying the emotions happy, sad and disgust. However the performance of the classifier was poor in classifying the emotion fear. This was due to inaccuracy of the evoked emotion and the participants expressed discomfort during expressing this evoked emotion. Accuracy of the classifier can be improved by increasing the number of participants and by reducing the dimensionality of the mean image.

TABLE II. PERFORMANCE OF FACE DETECTION ALGORITHM

| Emotion | Angry | Disgust | Fear | Happy | Sad | Surprise |
|---|---|---|---|---|---|---|
| **Angry** | **66.67** | 18.49 | 12.18 | 0.83 | 0.4 | 1.43 |
| **Disgust** | 2.94 | **69.81** | 6.07 | 0.85 | 19.50 | 0.83 |
| **Fear** | 13.66 | 29.90 | **48.37** | 0.71 | 1.91 | 5.75 |
| **Happy** | 3.01 | 1.13 | 1.20 | **82.36** | 0.71 | 11.59 |
| **Sad** | 1.30 | 20.84 | 2.13 | 0.64 | **70.18** | 4.91 |
| **Surprise** | 1.65 | 6.51 | 3.18 | 22.60 | 2.65 | **63.41** |

## VII. ACKNOWLEDGMENT

This work is supported by Lakeside Labs GmbH, Klagenfurt, Austria and funded by the European Regional Development Fund (ERDF) and the Carinthian Economic Promotion Fund (KWF) under grant KWF 20214/18354/27107.v

## VIII. CONCLUSION

This paper presents an approach to solve several issues pertaining to the lack of face detection algorithm, lack of robust feature extraction technique for varying size of face region and lack of optimization for car environment. As a part of the future work, it is intended to fuse thermal camera with several other sensors in order to obtain better classification which in turn provides better safety and comfort to the driver. Thus a framework is being designed to be deployed directly into a car as a part of ADAS.

## REFERENCES

[1] E. Verhoef, "External effects and social costs of road transport," in Transportation Research Part A: Policy and Practice, 1994, pp. 273-287.

[2] A. Trawén, P. Maraste, and U. Persson, "International comparison of costs of a fatal casualty of road accidents in 1990 and 1999," in Accident Analysis & Prevention, 2002, pp. 323-332.

[3] D. Mohan, "Social Cost of Road Traffic Crashes in India," in Proceedings First Safe Community Conference on Cost of Injury, Viborg, Denmark, 2002, pp. 4.

[4] A. Riewpaiboon, and P. Piyauthakit, "Economic burden of road traffic injuries: a micro-costing approach," 2008.

[5] C. Ghee, G. Britain, and T. R. Laboratory, "Socio-economic aspects of road accidents in developing countries," 1997.

[6] R. Pérez-Núñez, M. Híjar-Medina, I. Heredia-Pi *et al.*, "Economic impact of fatal and nonfatal road traffic injuries in Belize in 2007," in Revista Panamericana de Salud Pública, pp. 326-336.

[7] M. Lu, K. Wevers, R. van der Heijden *et al.*, "ADAS applications for improving traffic safety," 2005, pp. 3995-4002.

[8] G. Vivo, P. Dalmasso, and F. Vernacchia, "The European Integrated Project," 2007, pp. 624-629.

[9] T. Michalke, R. Kastner, J. Fritsch *et al.*, "Towards a proactive biologically-inspired advanced driver assistance system," 2009, pp. 621-628.

[10] C. L. Lisetti, and F. Nasoz, "Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals," in Journal on applied Signal Processing, 2004, pp. 1672-1687.

[11] C. T. Lin, R. C. Wu, S. F. Liang *et al.*, "EEG-based drowsiness estimation for safety driving using independent component analysis," in Circuits and Systems I: Regular Papers, IEEE Transactions on, 2005, pp. 2726-2738.

[12] Y. Lin, H. Leng, G. Yang *et al.*, "An Intelligent Non-intrusive Sensor for Driver Pulse Wave Measurement," 2007, pp. 790 - 799.

[13] J. Dai, J. Teng, X. Bai *et al.*, "Mobile phone based drunk driving detection," pp. 1-8.

[14] J. H. Yang, Z. H. Mao, L. Tijerina *et al.*, "Detection of driver fatigue caused by sleep deprivation," in Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 2009, pp. 694-705.

[15] G. Yang, Y. Lin, and P. Bhattacharya, "A driver fatigue recognition model based on information fusion and dynamic Bayesian network," in Information Sciences, pp. 1942-1954.

[16] Q. Wang, J. Yang, M. Ren *et al.*, "Driver fatigue detection: a survey," 2006, pp. 8587-8591.

[17] S. Jin, S. Y. Park, and J. J. Lee, "Driver fatigue detection using a genetic algorithm," in Artificial Life and Robotics, 2007, pp. 87-90.

[18] S. Singh, and N. P. Papanikolopoulos, "Monitoring driver fatigue using facial analysis techniques," 2002, pp. 314-318.

[19] L. M. King, H. T. Nguyen, and S. K. L. Lal, "Early Driver Fatigue Detection from Electroencephalography Signals using Artificial Neural Networks," in Proceedings of the 28th IEEE EMBS Annual International Conference, New York City, USA, 2006, pp. 2187-2190.

[20] A. B. Albu, B. Widsten, T. Wang *et al.*, "A computer vision-based system for real-time detection of sleep onset in fatigued drivers," 2008, pp. 25-30.

[21] M. Saradadevi, and P. Bajaj, "Driver fatigue detection using mouth and yawning analysis," in IJCSNS, 2008, pp. 183.

[22] M. Akhloufi, and A. Bendada, "Thermal Faceprint: A new thermal face signature extraction for infrared face recognition," 2008, pp. 269-272.

[23] P. Buddharaju, I. Pavlidis, and P. Tsiamyrtzis, "Pose-invariant physiological face recognition in the thermal infrared spectrum," 2006, pp. 53.

[24] M. M. Khan, M. Ingleby, and R. D. Ward, "Automated Facial Expression Classification and affect interpretation using infrared measurement of facial skin temperature variations," in ACM

[25] J. Wilder, P. Phillips, C. Jiang *et al.*, "Comparison of visible and infra-red imagery for face recognition," 2002, pp. 182-187.

[26] F. J. Prokoski, "History, current status, and future of infrared identification," in Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications, Washington, DC, USA, 2000, pp. 5.

[27] B. Fasel, and J. Luettin, "Automatic facial expression analysis: a survey," in Pattern Recognition, 2003, pp. 259-275.

[28] I. Pavlidis, J. Levine, and P. Baukol, "Thermal imaging for anxiety detection," 2002, pp. 104-109.

[29] I. Pavlidis, and J. Levine, "Thermal facial screening for deception detection," 2002, pp. 1143-1144.

[30] A. Ortony, and T. J. Turner, "What's basic about basic emotions," in Psychological review, 1990, pp. 315-331.

[31] E. Williams, A. Heusch, and P. McCarthy, "Thermal screening of facial skin arterial hot spots using non-contact infrared radiometry," in Physiological measurement, 2008, pp. 341.

[32] B. Nhan, and T. Chau, "Infrared thermal imaging as a physiological access pathway: a study of the baseline characteristics of facial skin temperatures," in Physiological measurement, 2009, pp. 23.

[33] M. P. Cetingul, and C. Herman, "Identification of Subsurface Structures from the Transient Thermal Response and Surface Temperature Measurements," 2008, pp. 1-8.

[34] M. M. Khan, R. D. Ward, and M. Ingleby, "Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature," in ACM Transactions on Applied Perception (TAP), 2009, pp. 1-22.

[35] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, "Digital image processing using MATLAB," 2004, pp.

[36] T. Acharya, and A. K. Ray, "Image processing: principles and applications," 2005.

[37] H. D. Vankayalapati, and K. Kyamakya, "Nonlinear Feature Extraction Approaches for Scalable Face Recognition Applications," in ISAST Transactions on Computers and Intelligent Systems, 2009.

[38] B. Takacs, "Comparing face images using the modified Hausdorff distance," in Pattern Recognition, 1998, pp. 1873-1881.