Object detection in thermal images

STUDENT: GUSTAVO SILVA
SUPERVISOR: LUÍS CORTE-REAL (PROF. DR.)
SECOND SUPERVISOR: PEDRO CARVALHO (PROF. DR.)
INSITUTIONAL RESPONSIBLE: ANDRÉ FERREIRA (ENG.)

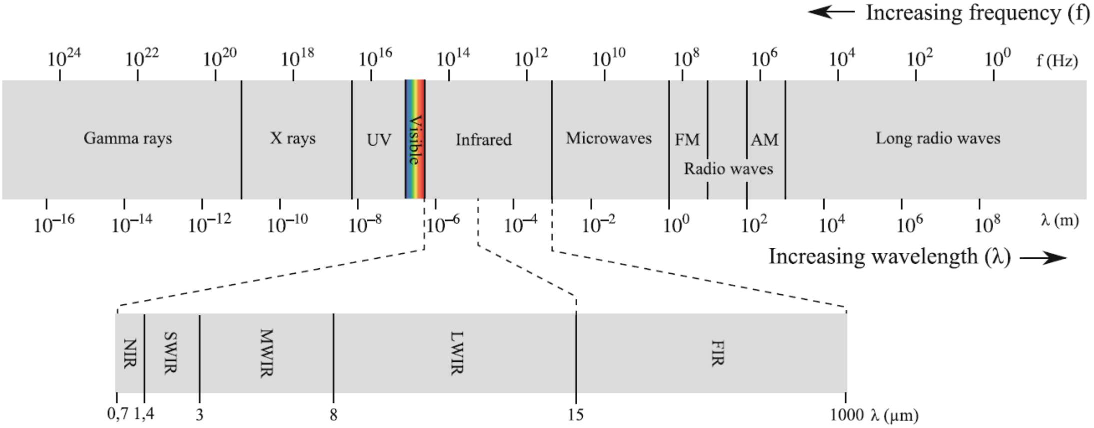
Summary

1. Context and Motivation
2. State of the art
    1. Multimodal sensor fusion
    2. Analyzing the vehicle
    3. Analyzing people inside the vehicle
3. Proposed solution
    1. Description
    2. Camera setup
    3. Datasets
4. Validation
    1. Face/smoking detection
    2. Respiratory/heart rate estimation
    3. Emotion recognition
5. Work plan

Context and Motivation

Context

Objective: develop software to analyze the interior of vehicles using images captured by, but not limited to, cameras that operate in the infrared region of the electromagnetic spectrum.



The electromagnetic spectrum with sub-divided infrared spectrum.

Example of thermal image of a vehicle occupant.

Motivation

Disadvantages of RGB images:

- Require sun or artificial lighting
- Prone to variations in:
  - Light intensity
  - Color balance
  - Light direction
- Cannot detect some events:
  - Overheating objects
  - Human breath

Why monitor the vehicle interior?

- Register rule-breaking activities
- Improve passenger comfort
- Trigger an alert if a child is left alone inside
- Estimate passenger satisfaction

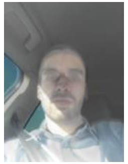
State of the art

Multimodal sensor fusion

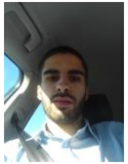
EARLY FUSION

Combines the different types of imagery before passing the data to a classifier. Examples:

- Weighted Fusion

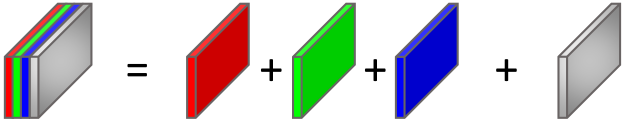


= 0.5 *



+ 0.5 *

- Band combination



LATE FUSION

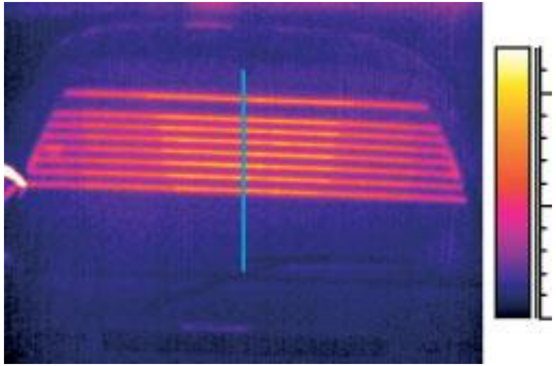
Merging the final results of multiple separate classifiers into a single stronger classifier.

- Ensemble methods
  - Bagging
  - Stacking
  - Voting
  - Boosting

Analyzing the vehicle

Previous attempts found in literature:

- Detect defects in vehicle components
- Measure temperature to improve passenger comfort



Defective defogger captured with a thermal camera.

Analyzing people inside the vehicle

Face detection

TRADITIONAL

Learning-based:

- Viola-Jones

Low-level:

- Image segmentation according to pixel temperature
- Horizontal/vertical projection
- Curves with a head-like shape
- …

DEEP LEARNING

- R-CNN
- Fast-R-CNN
- Faster-R-CNN
- SSD
- YOLO
- YOLOv2

Require a large dataset and no pretrained models exist for thermal images

Emotion recognition

Some datasets are available:

- IRIS Thermal/Visible Face Database
- I.Vi.T.E
- Natural Visible and Infrared facial Expression Database (NVIE)
- Kotani Thermal Facial Emotion (KTFE) dataset

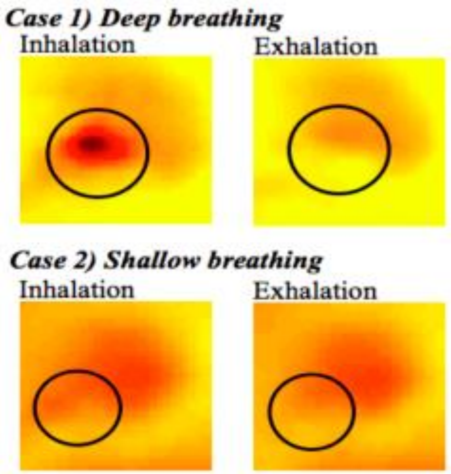
They contain 3 to 6 expressions – happiness, sadness, surprise, fear, anger, disgust

Breath

Thermal cameras are able to detect temperature variations in the nasal area. Two possible ROIs:

- Nostril - requires the camera to be positioned below
- Philtrum - camera can be positioned in front of the face

Challenge: Trepidation and head movement. Some solutions include creating a thermal gradient map and using Median Flow on the KLT tracker.

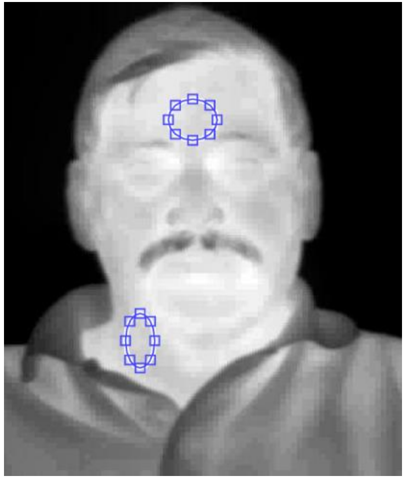


Detecting breath in the nostril.

Heart rate

- Thermal
  1. Measure temperature variation in ROIs
  2. Estimate rate with Fourier Transform
- RGB
  1. Apply Eulerian Video Magnification or other spatio-temporal filter
  2. Measure color variation in ROIs
  3. Estimate rate with Fourier Transform

Challenge: Handle large motions and low-resolution cameras.



Example ROIs for heart rate estimation.

Blush

Why? Identify anxiety, embarassement or deception.

How? Mainly a ROI tracking challenge – the eyes can be a reference point.

S. Yue et al. obtained 77% TPR and 60% TNR with an Argus P7225 (320x240).

Locating ROI for blush detection.

Smoking

- No previous attempts on the detection of the act of smoking.
- The peripheral of the coal burns at around **500ºC**.
- Challenges:
  - False positives (e. g. sun reflection)
  - Occlusions (cigar/cigarette lowered or moved outside the window)

Thermal image of a person smoking.

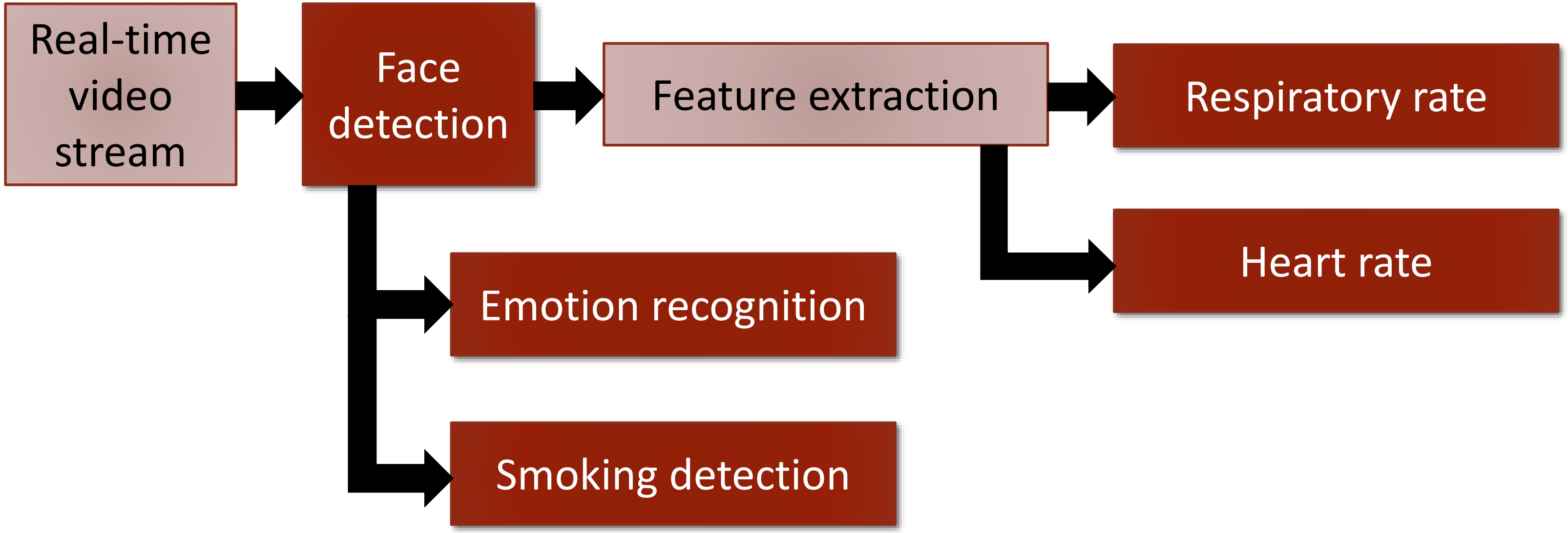
Proposed solution

Proposed solution

Focus on occupants, tackling the following:

- Occupant detection (face detection)
- Emotion recognition
- Respiratory rate
- Smoking
- Heart rate

High-level architecture



Possible cameras

Model	Res.	Waveband	Temp. Range	FPS	FOV	RGB res.
Argus P7225	320x240	8 - 14µm	-40ºC -- 200ºC	60	24ºx18º	-
AXIS Q1922	640x480	8 - 14µm	unspecified	30	57ºx44º	-
FLIR A20M	320x240	7.5 - 13µm	-20ºC -- 250ºC	60	25ºx19º	-
FLIR A40M	320x240	7.5 - 13µm	-40ºC -- 2000ºC	60	24ºx18º	-
FLIR Lepton 25º	80x60	8 - 14µm	0ºC -- 120ºC	8.6	25ºx19º	-
FLIR Lepton 50º	80x60	8 - 14µm	0ºC -- 120ºC	8.6	51ºx38º	-
FLIR Lepton 3	160x120	8 - 14µm	0ºC -- 450ºC	8.8	56ºx42º	-
FLIR ONE Gen II	160x120	8 - 14µm	-20ºC -- 120ºC	8.7	46ºx35º	640x480
FLIR ONE	80x60	8 - 14µm	-20ºC -- 120ºC	8.7	50ºx38º	1440x1080
FLIR ONE Pro	**160x120**	**8 - 14µm**	**-20ºC -- 400ºC**	**8.7**	**55ºx43º**	**1440x1080**
FLIR SC640	640x480	7.5 - 13µm	-40ºC -- 1500ºC	30	45ºx34º	-
NEC R300	640x480	8 - 14µm	-40ºC -- 2000ºC	60	22ºx17º	-
TESTO 880-3	160x120	8 - 14µm	-20ºC -- 350ºC	9	32ºx24º	-

Camera setup

Camera placed in front of the occupant, aiming directly at the face.

- Advantages:
  - Accurate face detection
  - Easier emotion recognition
  - Heart rate measurement
- Disadvantages:
  - Lower robustness in respiratory pattern estimation
  - One camera required per person

Validation

Face/smoking detection

Intersection over Union (IoU)

$$IoU(B_1, B_2) = \frac{B_1 \cap B_2}{B_1 \cup B_2}$$

Average Precision

$$Average\ Precision = \frac{Number\ of\ true\ positives}{Number\ of\ true\ positives + Number\ of\ false\ positives}$$

Mean Average Precision (mAP)

$$mAP(classes) = \frac{\sum_{c \in classes} AveragePrecision(c)}{|classes|}$$

Area Under Curve (AUC)

AUC(r) = $\int_{0\%}^{100\%} Precision(r) * dr$

Respiration/heart rate estimation

Mean Squared Error (MSE)

$$MSE = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}$$

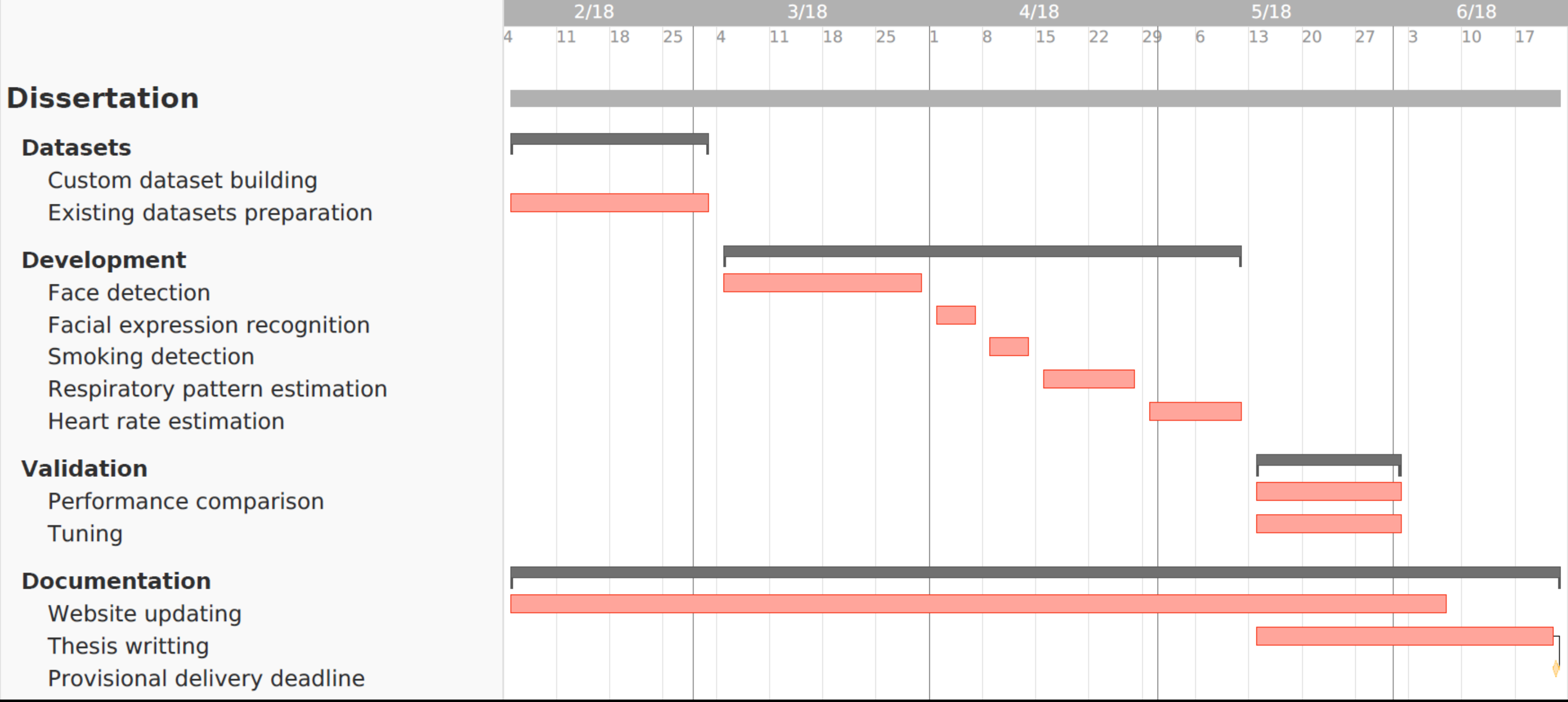
Accuracy

$$Accuracy = \frac{Predicted\ Rate}{Ground\ truth\ rate}$$

Emotion recognition

Categorical cross-entropy (H)

$H_{y'}(y) = -\sum_i y'_i * \log(y_i)$


2/18
3/18
4/18
5/18
6/18
4 11 18 25 4 11 18 25 1 8 15 22 29 6 13 20 27 3 10 17
Dissertation
Datasets
Custom dataset building
Existing datasets preparation
Development
Face detection
Facial expression recognition
Smoking detection
Respiratory pattern estimation
Heart rate estimation
Validation
Performance comparison
Tuning
Documentation
Website updating
Thesis writting
Provisional delivery deadline

Work plan

References

- Gade, R., & Moeslund, T. B. (2014). Thermal cameras and applications: A survey. *Machine Vision and Applications*, *25*(1), 245–262. https://doi.org/10.1007/s00138-013-0570-5
- Corneanu, C. A., Simón, M. O., Cohn, J. F., & Guerrero, S. E. (2016). Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(8), 1548–1568. https://doi.org/10.1109/TPAMI.2016.2515606
- Munz, E. D. (2017). *Psychotherapie in der Psychiatrie*. *Nervenheilkunde* (Vol. 36). https://doi.org/10.1007/s13398-014-0173-7.2
- Bennett, S., Harake, T. N. El, Goubran, R., & Knoefel, F. (2017). Adaptive Eulerian Video Processing of Thermal Video: An Experimental Analysis. *IEEE Transactions on Instrumentation and Measurement*, *66*(10), 2516–2524. https://doi.org/10.1109/TIM.2017.2684518