Hand Imaging in different spectral ranges

Ewelina Bartuzi
Advisor: prof. Andrzej Pacut

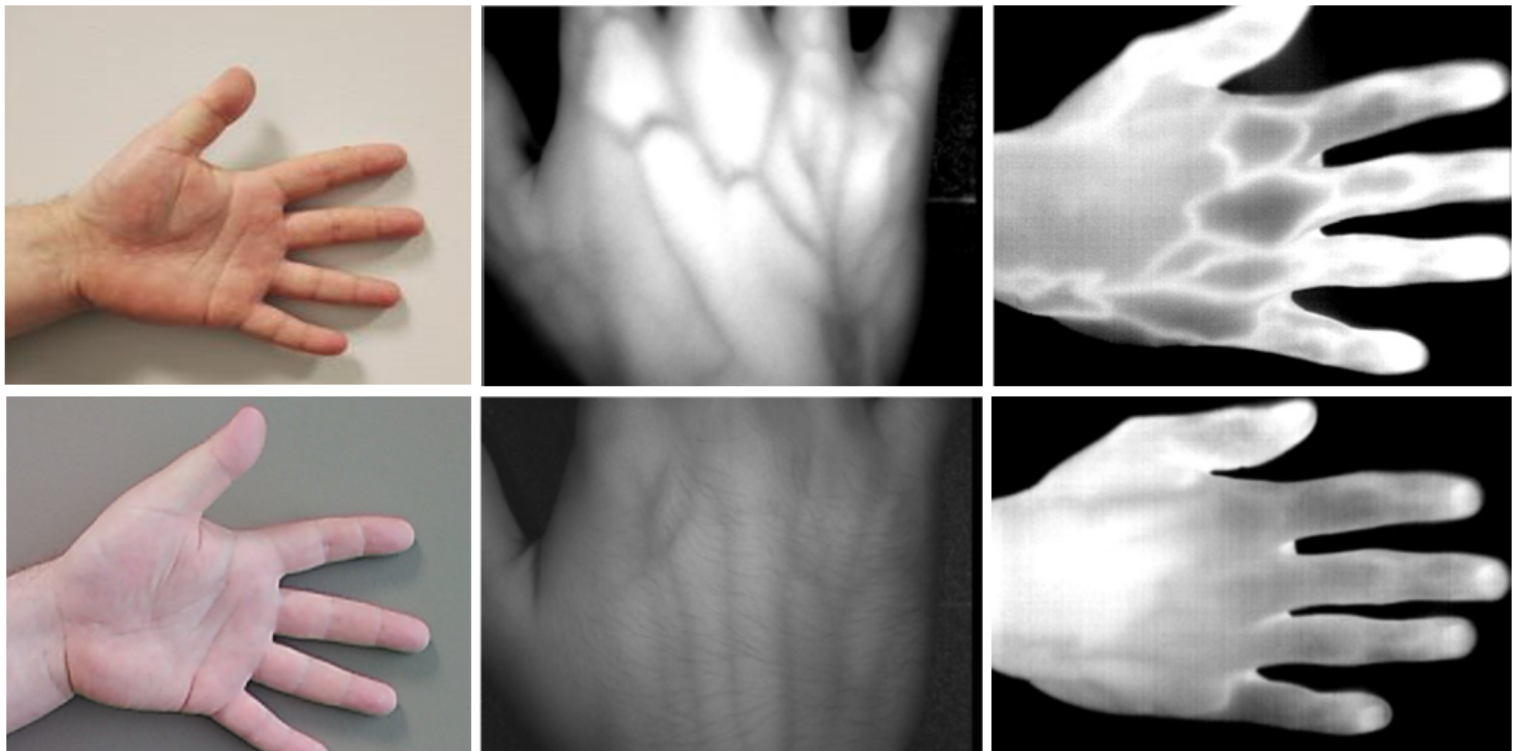
Biometrics and Machine Learning Groups
Institute of Control and Computation Engineering
Faculty of Electronics and Information Technology, WUT

13 czerwca 2018

Introduction

This work proposes a biometric recognition method based on hand images of the outer side of the hand. Hand measurement is hygienic, non-invasive and fast. The analysis includes images collected using three light sensors from various ranges:

visible light - VIS	near infrared - NIR	thermal infrared -TH
$380 - 750\ nm$	$0.7 - 1.4 \mu m$	$3 - 14 \mu m$

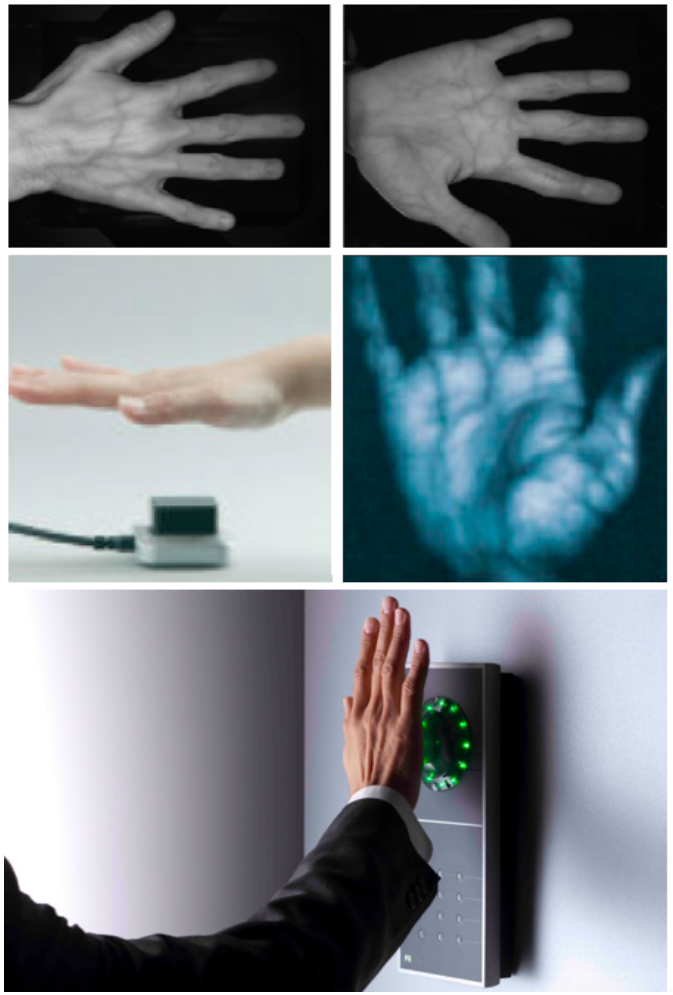


Hand imaging in visible light

- Device: camera, front or rear phone camera
- Hand side: palm side
- Features:
  - main lines: direction, bifurcations, ending, crossing
  - ridges and valley (?)
  - texture
- Accuracy: $65 - 97\%$
- EER: $3.5 - 35\%$

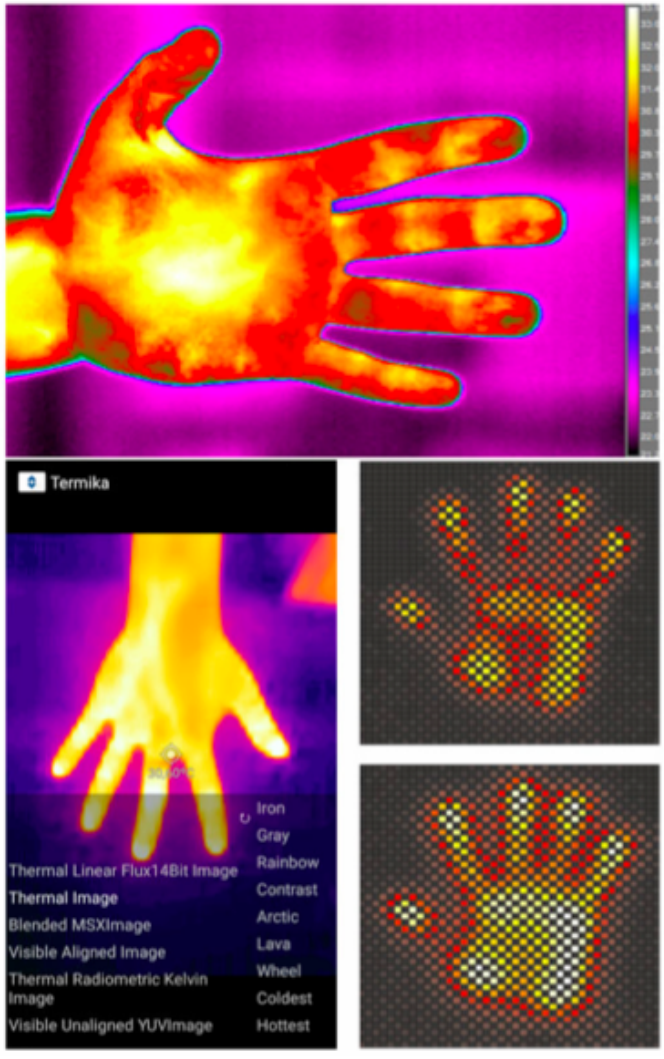
Hand imaging in near infrared

- Device: source and detector of infrared light
- Hand side: both sides
- Features:
  - veins pattern
- Accuracy: $90 - 100\%$
- EER: $0.0 - 15\%$



Thermal hand imaging

- Device: matrix with thermal sensors, thermal camera (professional or compatible with mobile devices)
- Hand side: both sides
- Features:
  - heat distribution
  - vein pattern
  - dynamics of temperature changes
- Accuracy: -
- EER: $5 - 40\%$



Conclusions

- Manifold hand use in biometry: geometry, texture, invisible features to the naked eye
- Incredible high efficiency of methods based on vein patterns

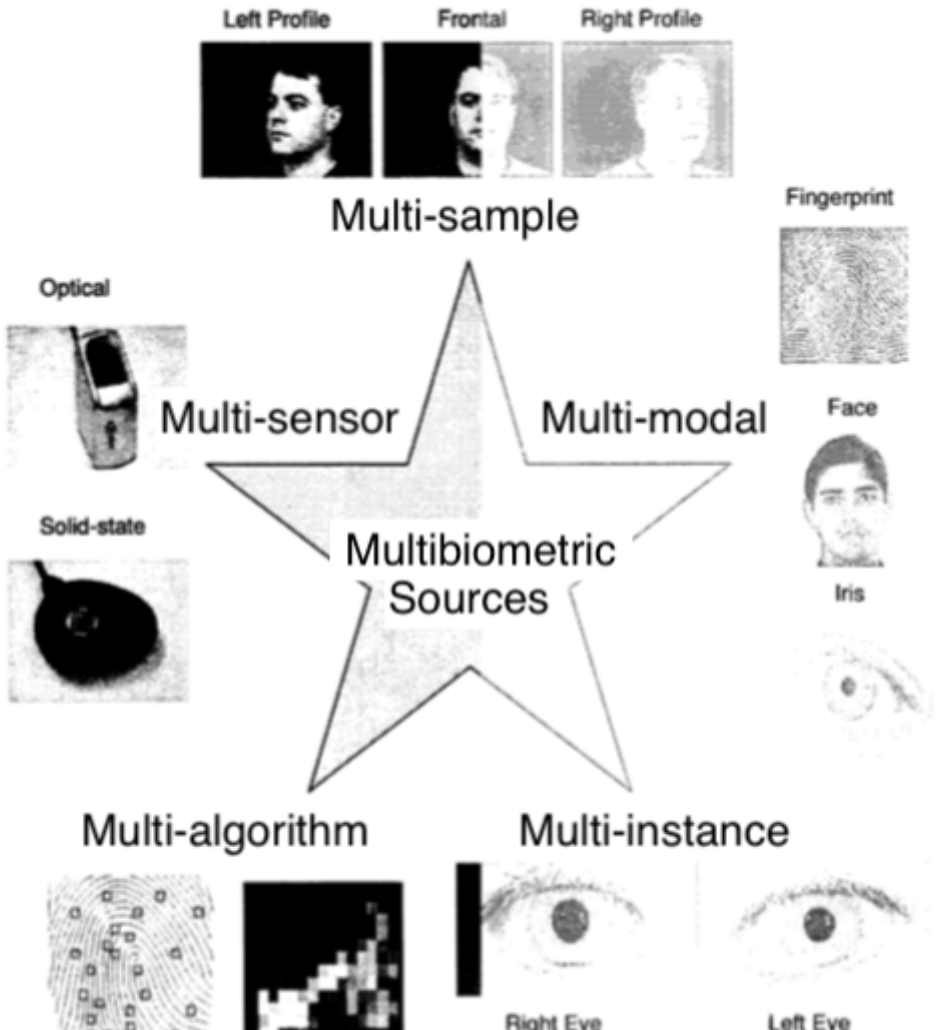
	visible images	*near IR images*	*thermal images*
features	main lines texture	vein pattern	heat distribution vein pattern
disadvantages	external light	hair light source - optical blur	poor stability
cost	cheap	expensive	mid

Multibiometrics - biometrics of many sources

(A. Ross, 2006)

Popular fusions in hand biometry:

- score fusion
- image blending



The purpose of this work

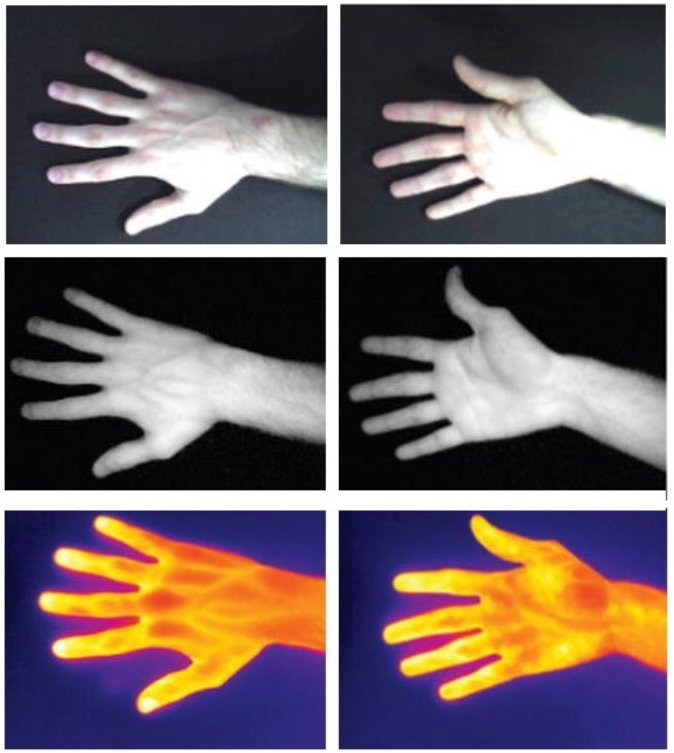
- Correlation between images collected with different sensors
- Estimating the effectiveness of biometric methods based on medium/poor-quality hand images

Questions

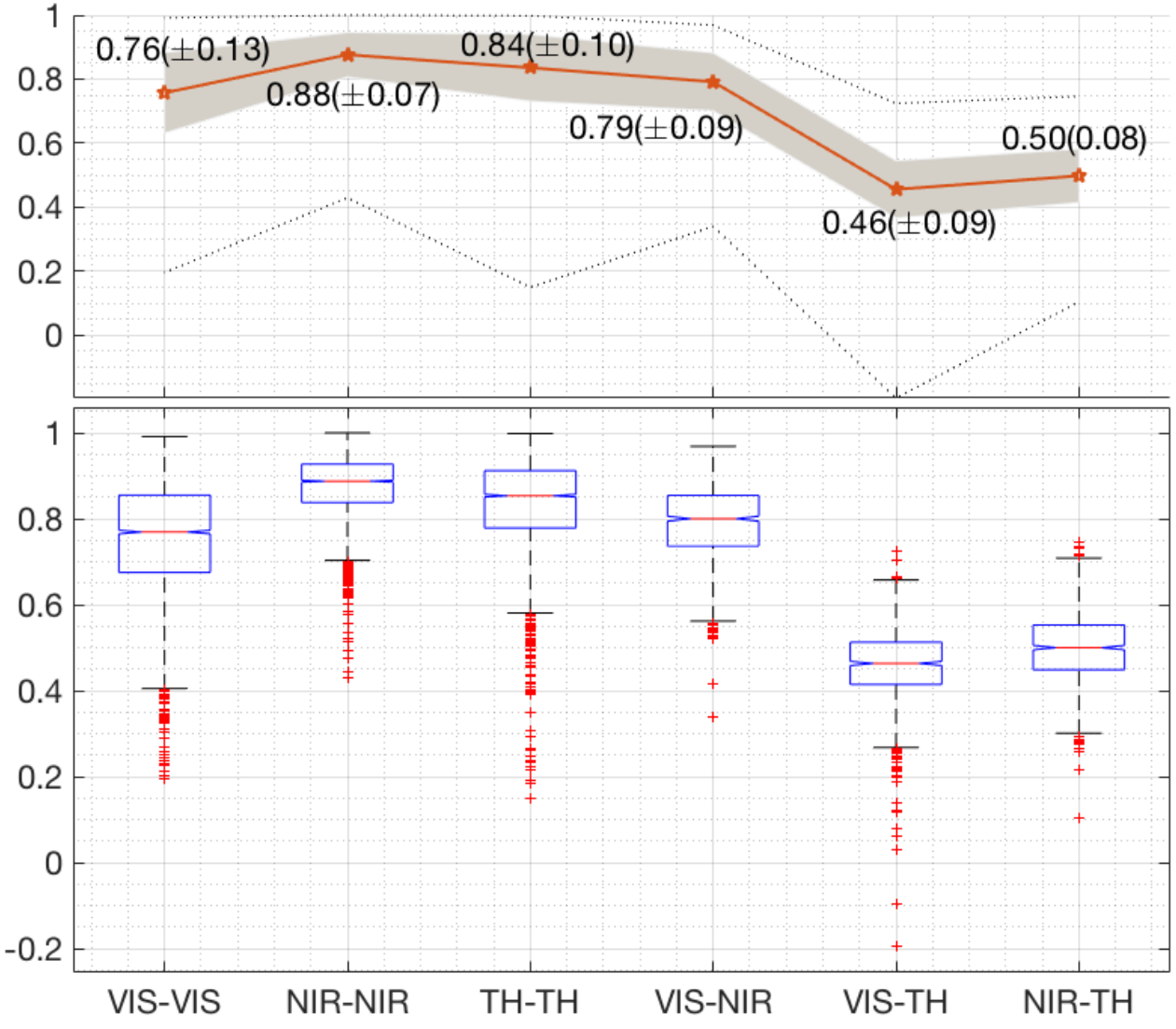
1. Are the hand temperature changes associated with the pattern of blood vessels?
2. What is the similarity between vein patterns obtained from NIR and TH images?
3. What is the effectiveness of systems based on texture descriptors?
4. What is the effectiveness of systems based on CNN?
5. What is the effectiveness of hand-based multibiometric systems?

Multispectral Database of Hand

- *Thermal Hand Image Database* (2012)
- 6.000 images
- 104 subjects $\times$ 5 sessions $\times$ 2 images
- left and right hand, palm and dorsal side
- 3 spectral ranges
  - VIS: visible light - $380 - 750nm$, 640×480 px
  - NIR: near infrared - $0.7 - 1.4\mu m$, 640×480 px
  - TH: thermal infrared - $3 - 14\mu m$, 320 ×240 px



Multispectral Database of Hand - correlation between images



Quality of data

Image Quality is a characteristic of an image that measures the perceived image degradation.

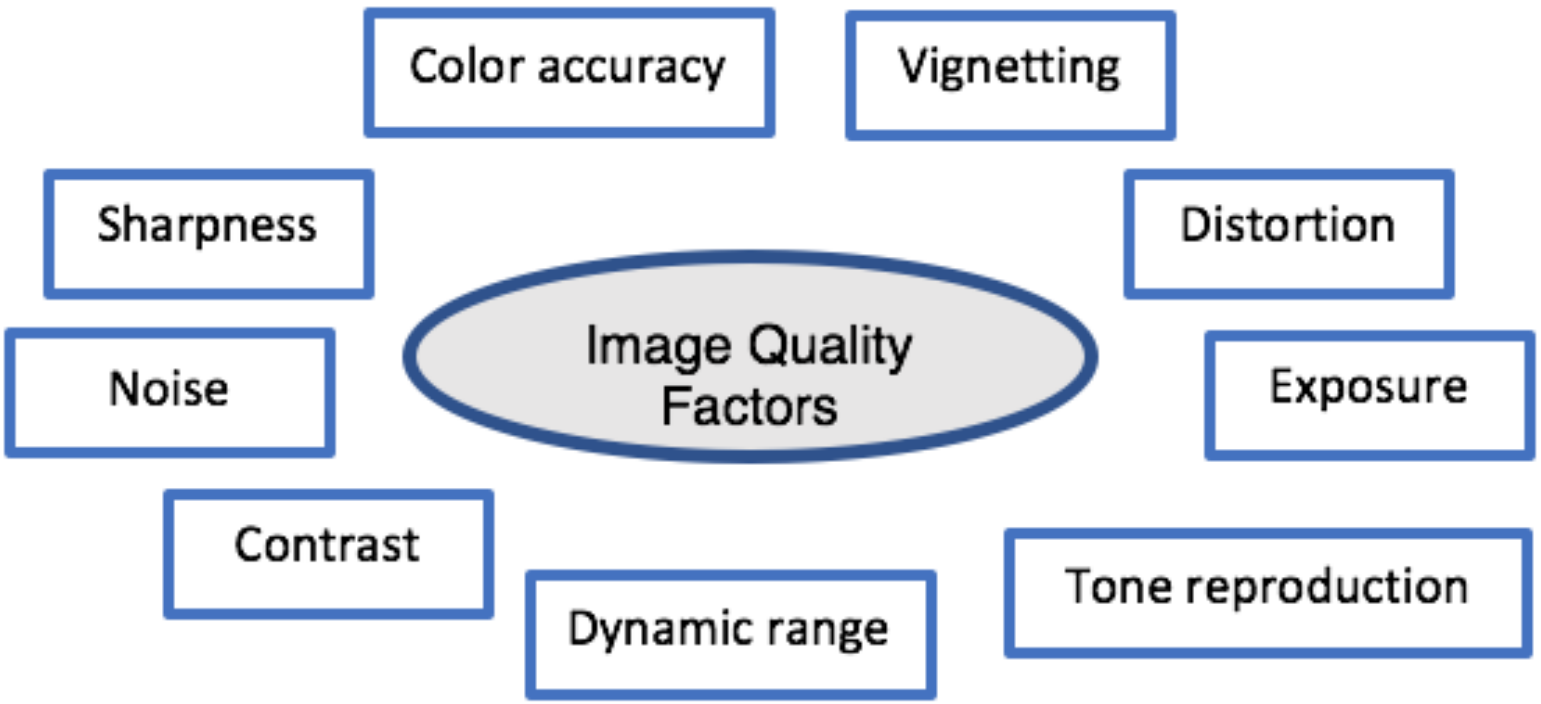


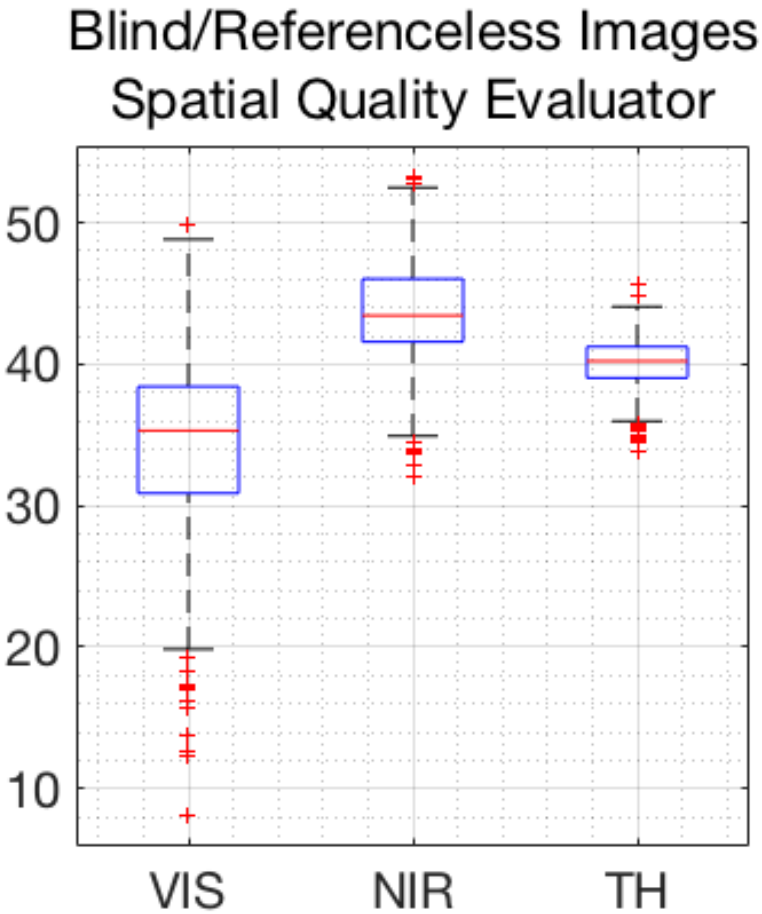
Image quality assessment categories

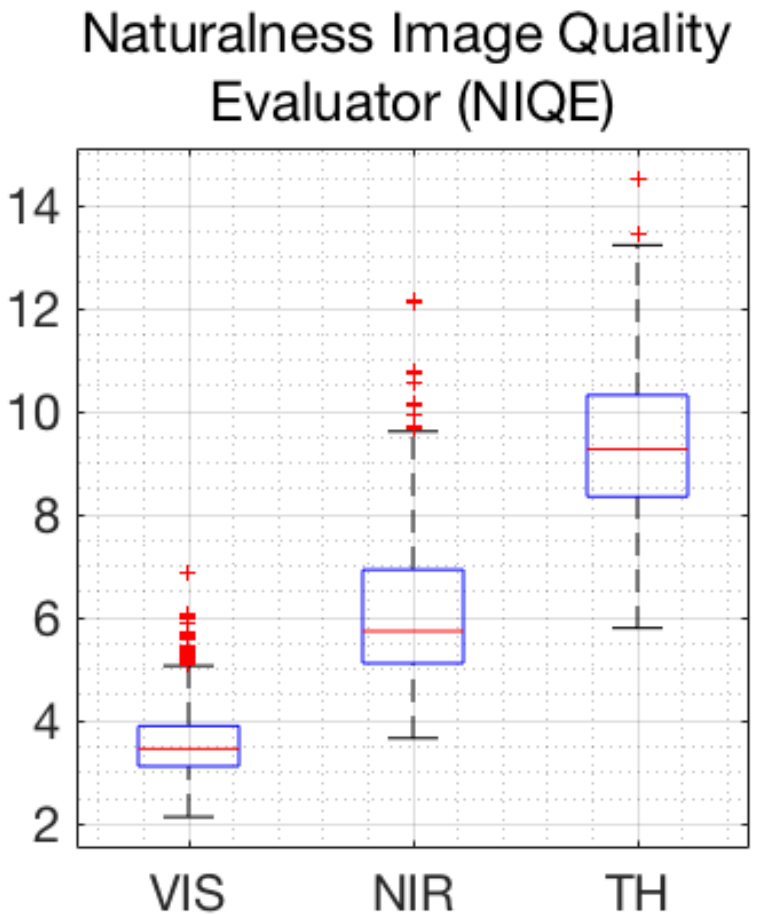
- **Full-reference methods** – FR metrics try to assess the quality of a test image by comparing it with a reference image that is assumed to have perfect quality
- **Reduced-reference methods** – RR metrics assess the quality of a test and reference image based on a comparison of features extracted from both images
- **No-reference methods** – NR metrics try to assess the quality of a test image without any reference to the original one.
  - Blind/Referenceless Image Spatial Quality Evaluator (BRISQE, 2012) - statistics of locally normalized luminance coefficients to quantify losses of 'naturalness' in the image due to the presence of distortions
  - Natural Image Quality Evaluator (NIQE, 2013) - is based on the construction of a 'quality aware' collection of statistical features based on a simple and successful space domain natural scene statistic model.

(*Image quality metrics can also be classified in terms of measuring only one specific type of degradation e.g., blurring, blocking, or ringing, or taking into account all possible signal distortions - multiple kinds of artifacts.*)

Quality of data

- noticeable quality differences (especially between NIR images)





Quality of data

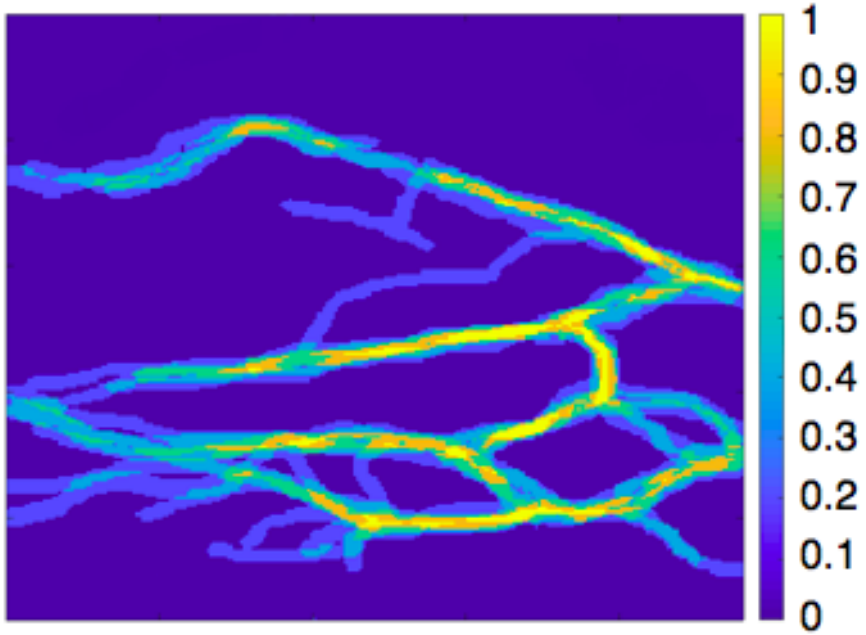
- noticeable quality differences (especially between NIR images)
- subjective division subjects into 4 classes:
  - I - good quality - well visible pattern of veins - 10 classes (9.62%)
  - II - medium quality - clearly visible major part of vein pattern - 27 classes (25.96%)
  - III - poor quality - weakly visible vein structures - 50 classes (48.08%)
  - IV - unusable images - over- or underexposed - 17 classes (16.35%)

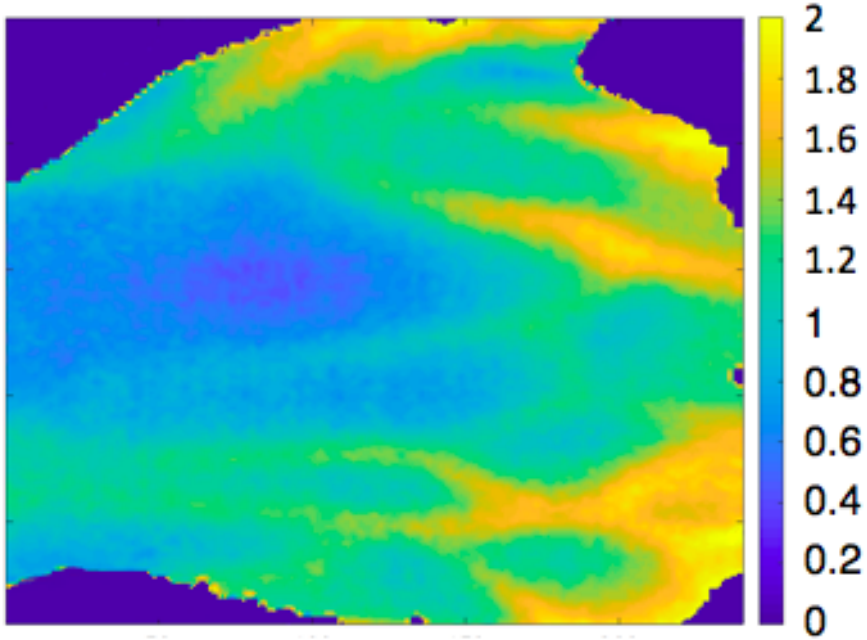
	BRISQE			NIQE		
	VIS	NIR	TH	VIS	NIR	TH
IV	35.36	43.60	39.61	3.69	6.47	8.92
III	35.21	43.92	40.13	3.57	6.04	9.45
II	33.61	44.04	40.14	3.47	5.96	9.11
I	32.68	43.38	40.17	3.52	5.95	9.68
sum	34.54	43.84	40.06	3.52	5.95	9.68

Q1. Are the hand temperature changes associated with the pattern of blood vessels?

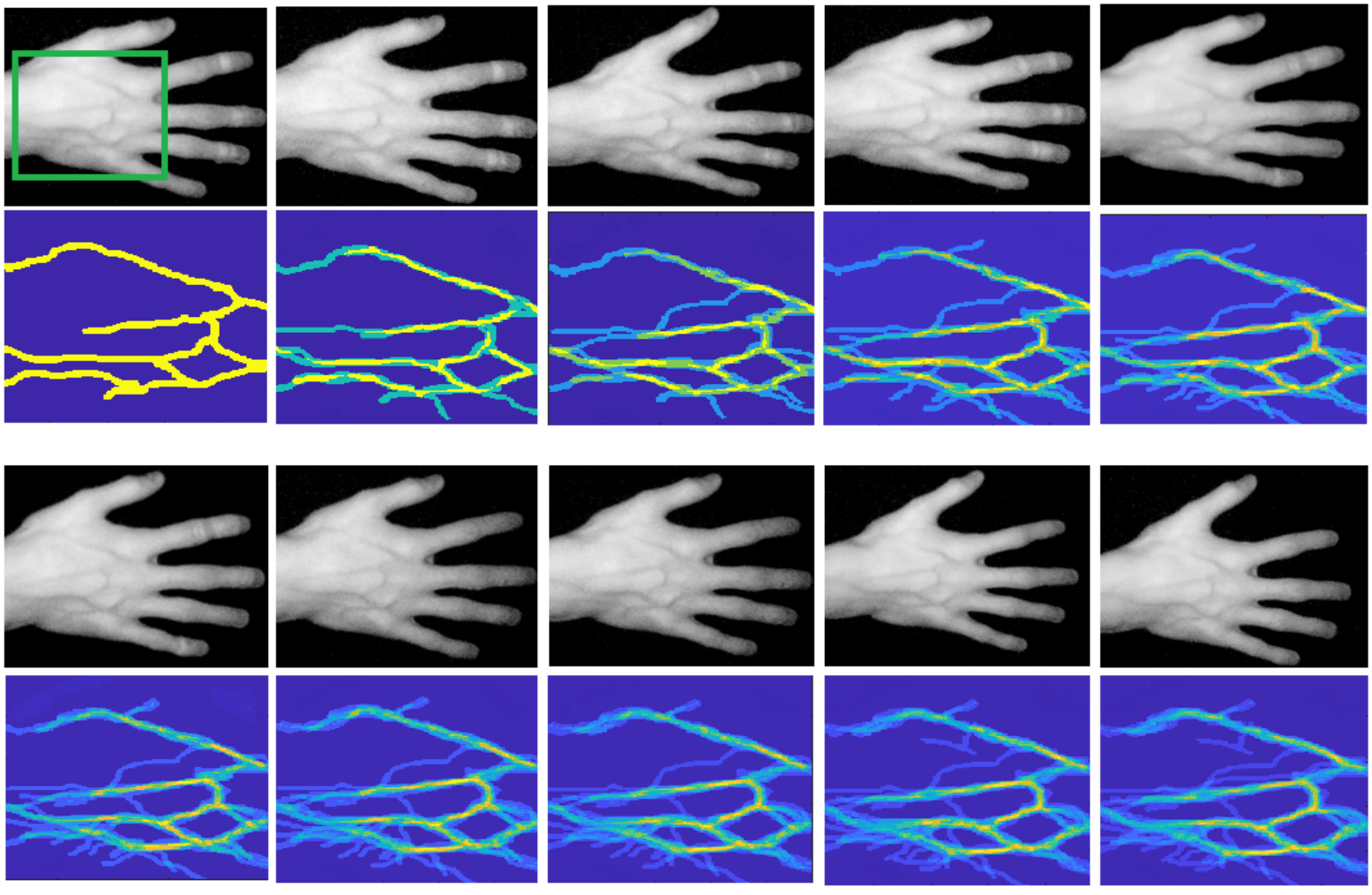
'Thermal stability maps may be related to blood vessels close to the skin surface(...).'
(supposition for palm side, E. Bartuzi, 2017)

manually extracted vein patterns	map of temperature deviations for all sessions $[^oC]$

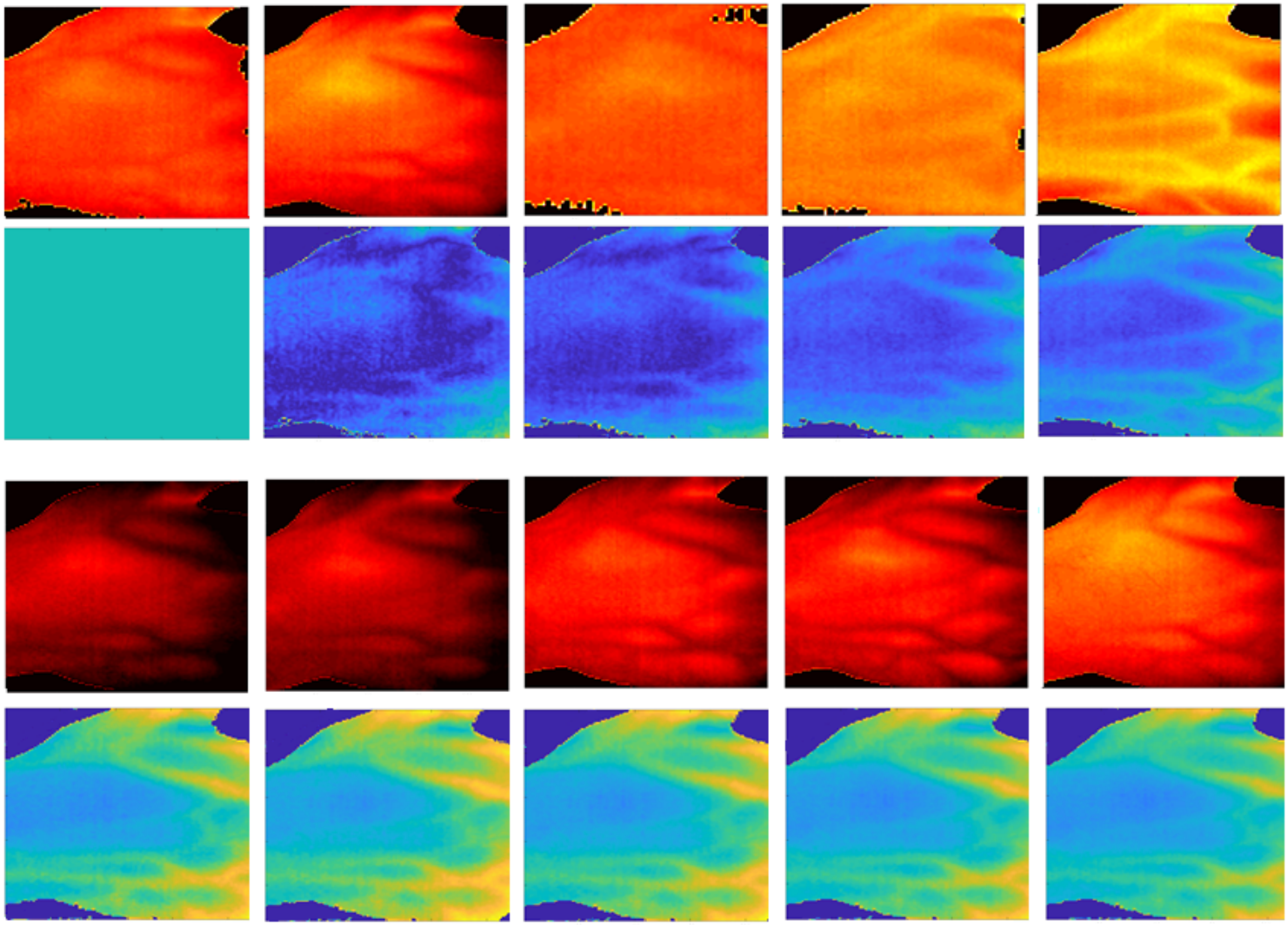




Manual vein extraction - class I

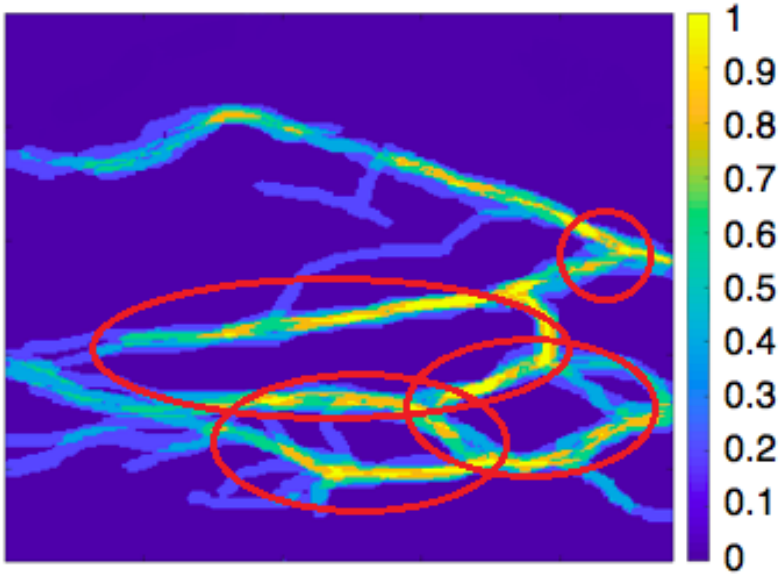


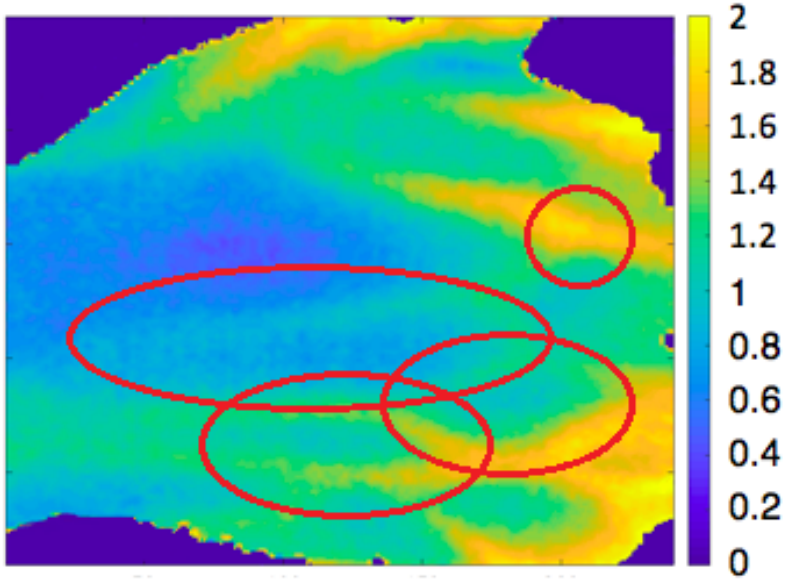
Thermal stability maps - class I



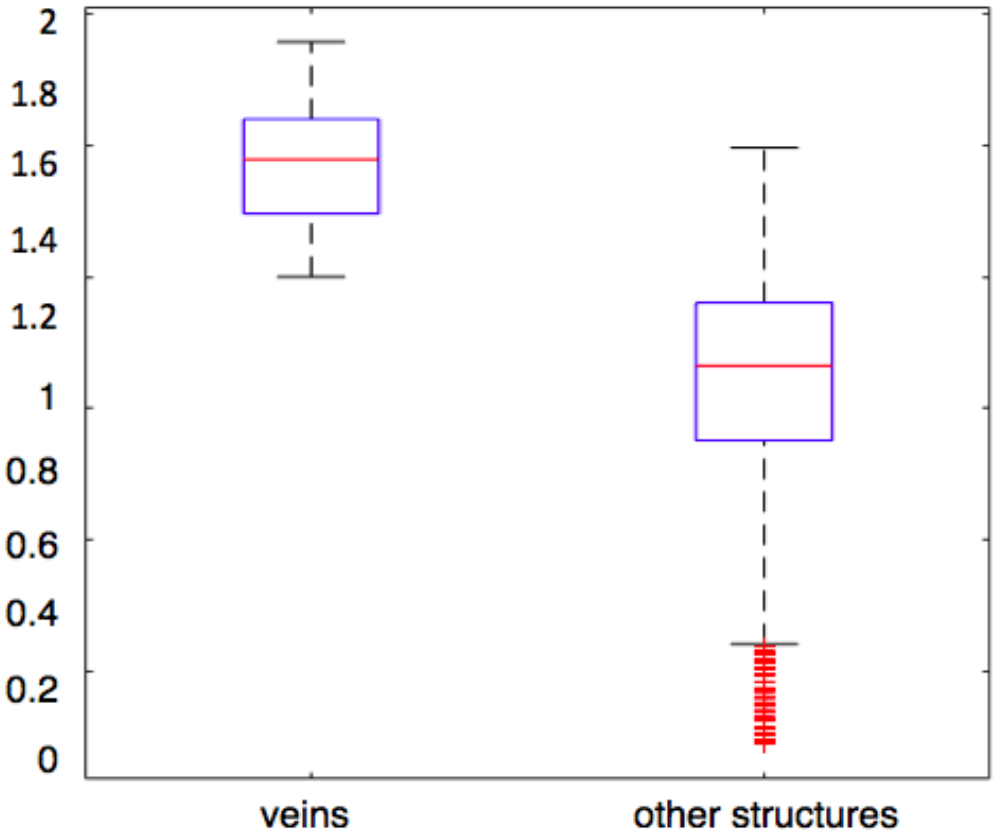
Q1. Are the hand temperature changes associated with the pattern of blood vessels?

- Noticeable similarities between the vascular pattern and the standard deviation map of temperatures
- The largest deviations of the values ($1 - 2^oC$) correspond to the vascular structures
- Maybe a more mathematical measure?



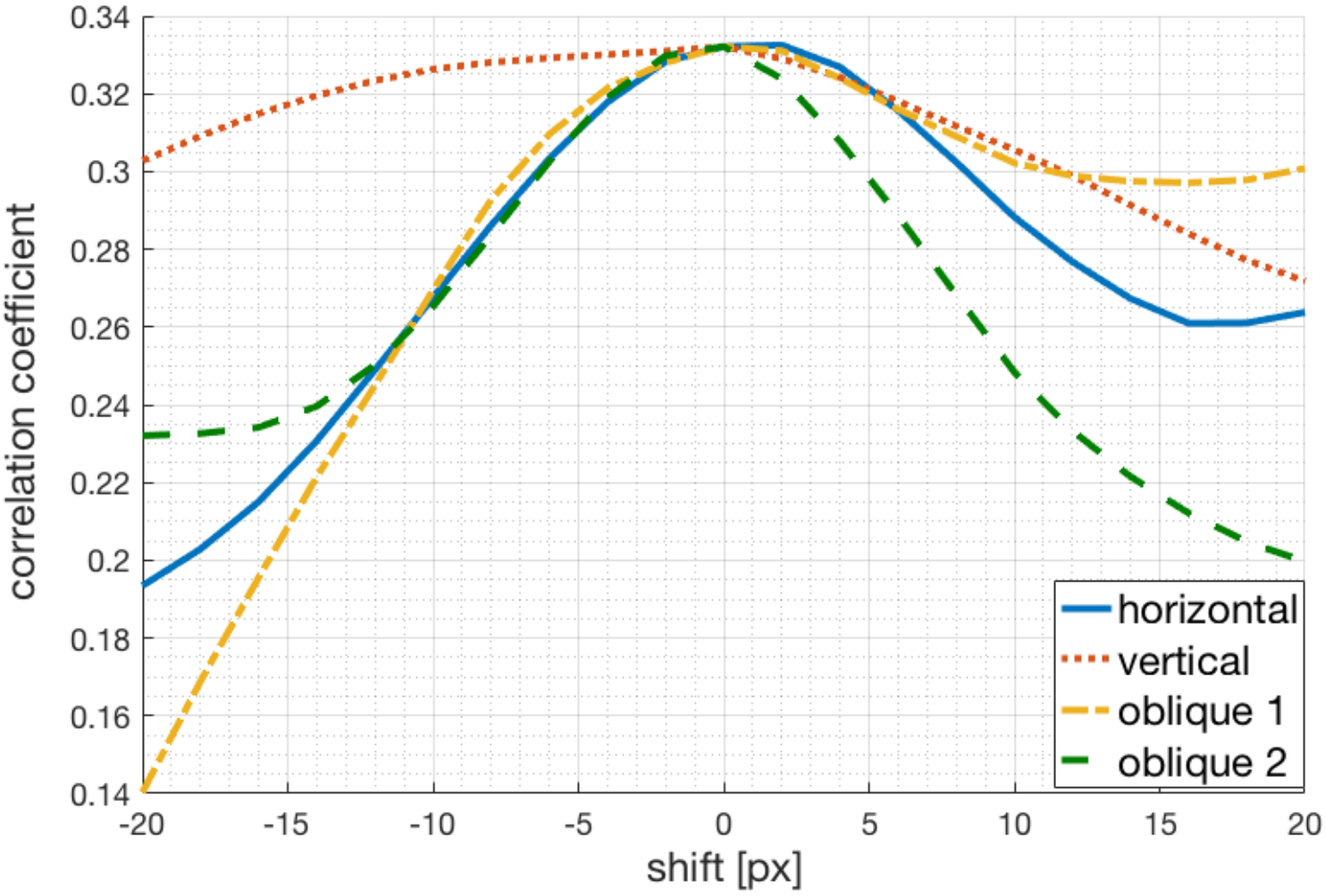


Q1. Are the hand temperature changes associated with the pattern of blood vessels?

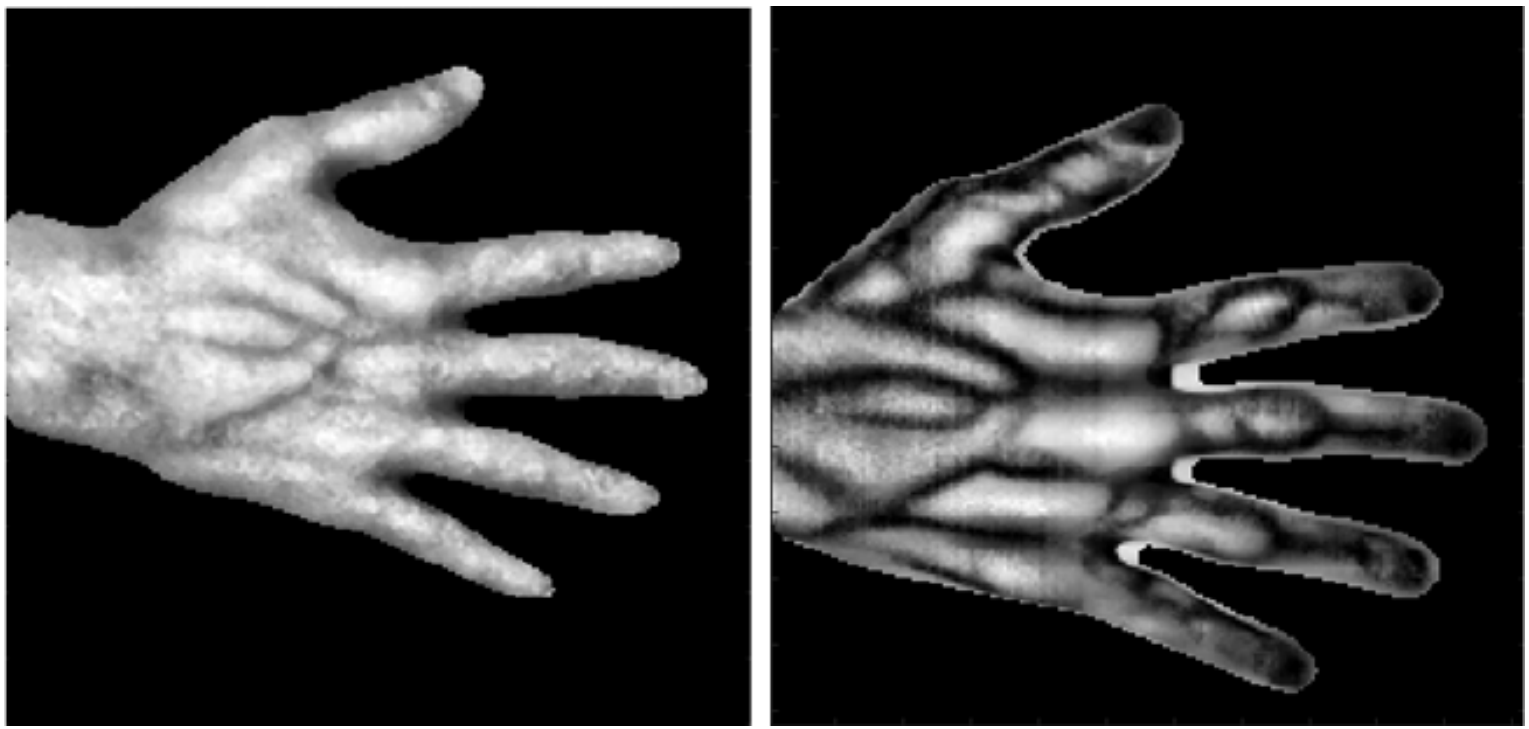


The value of standard deviation for veins pattern and other structure on the hand image

Q1. Are the hand temperature changes associated with the pattern of blood vessels?

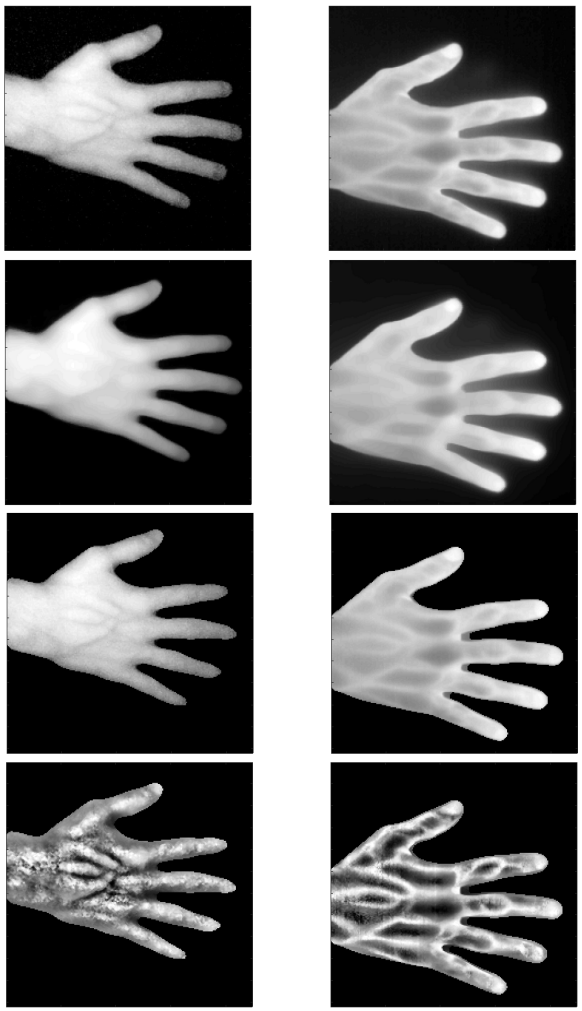


Q2. What is the similarity between vein patterns obtained from NIR and TH images?



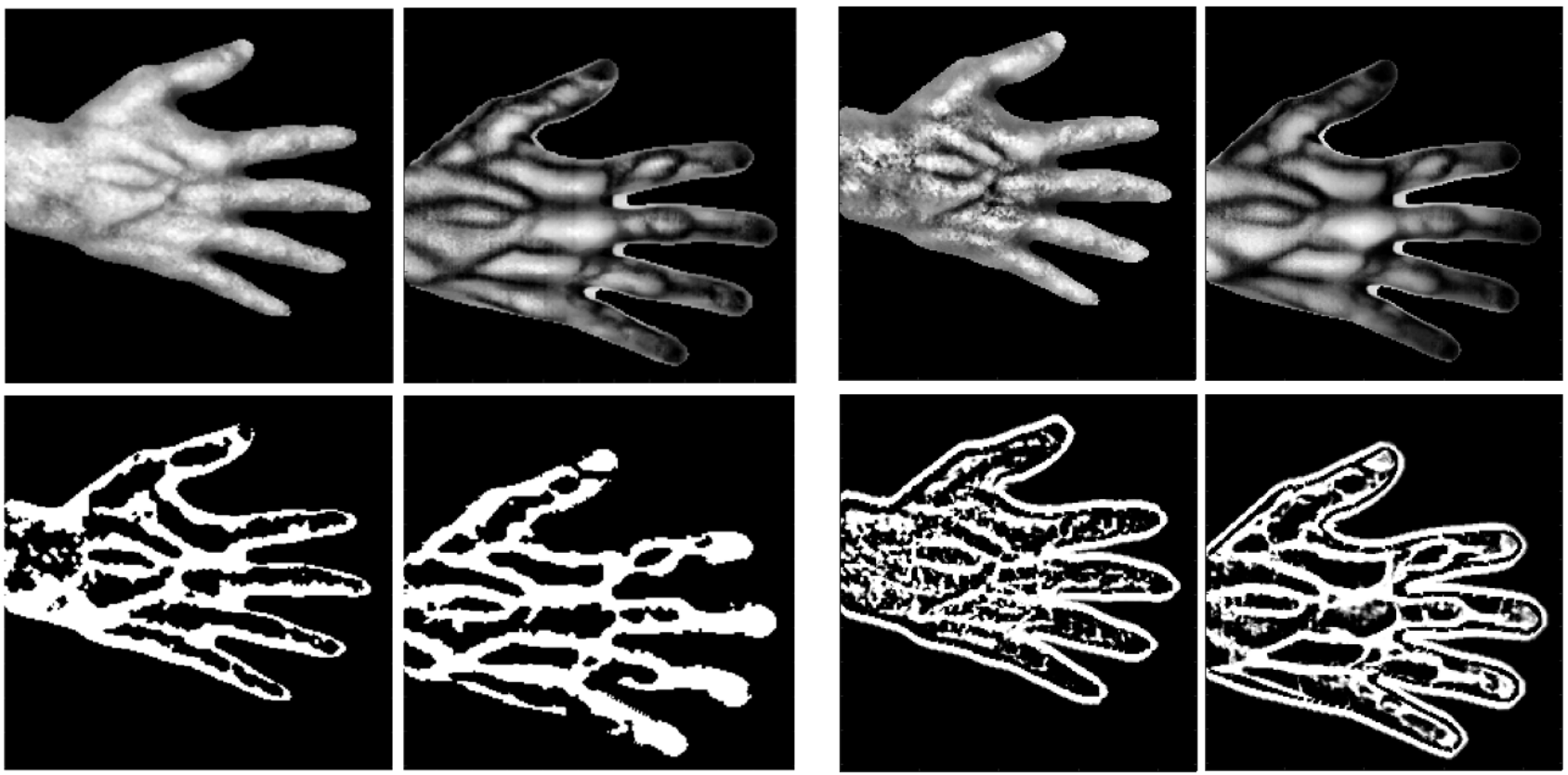
Vein extraction - pre-processing stages

- Cropping and resizing image - output: 224×224 px
- Median filtering - 11×11 px
- Hand segmentation using GMM
  - distributions of object's values and of the background's values on the histogram are approximated by two Gaussian curves
  - threshold - intersection point of these curves
- Image enhancement - CLAHE
  - image division into blocks (vein width $\simeq$ 5px, block size: 11×11)
  - determine the contrast enhancement function for each block
  - calculate new value for the central pixel of the block (with limited contrast)
  - value of other pixels - bilinear interpolation with central pixel of neighboring blocks as nodes



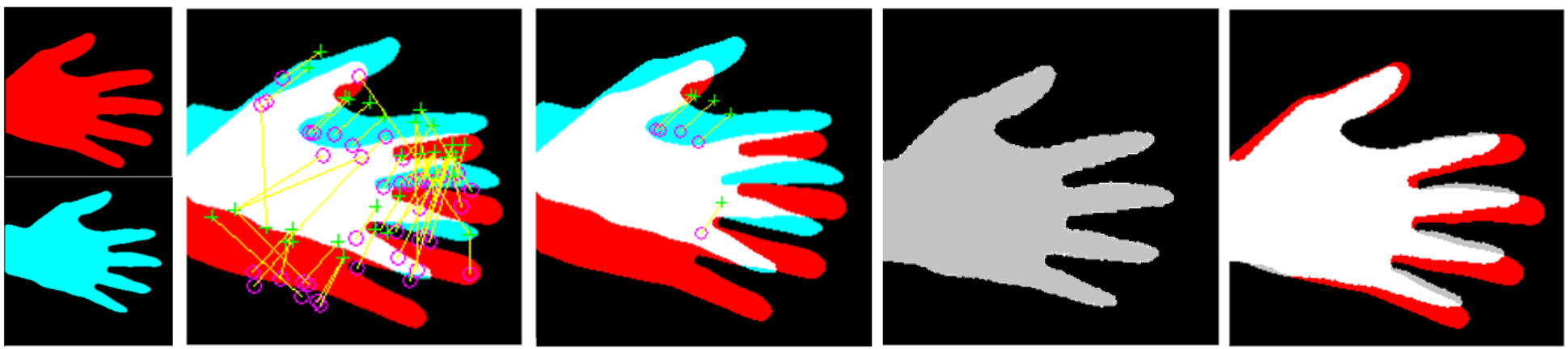
Vein extraction - two approaches

Method I - based on local thresholding	Method II - based on edge filters

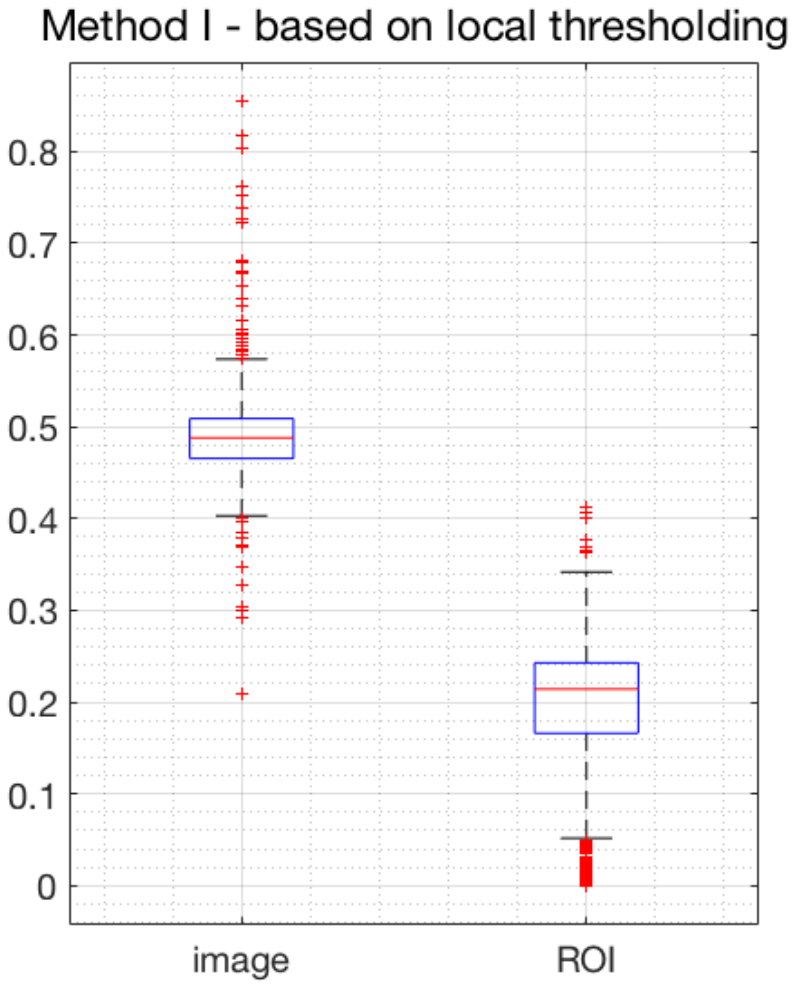


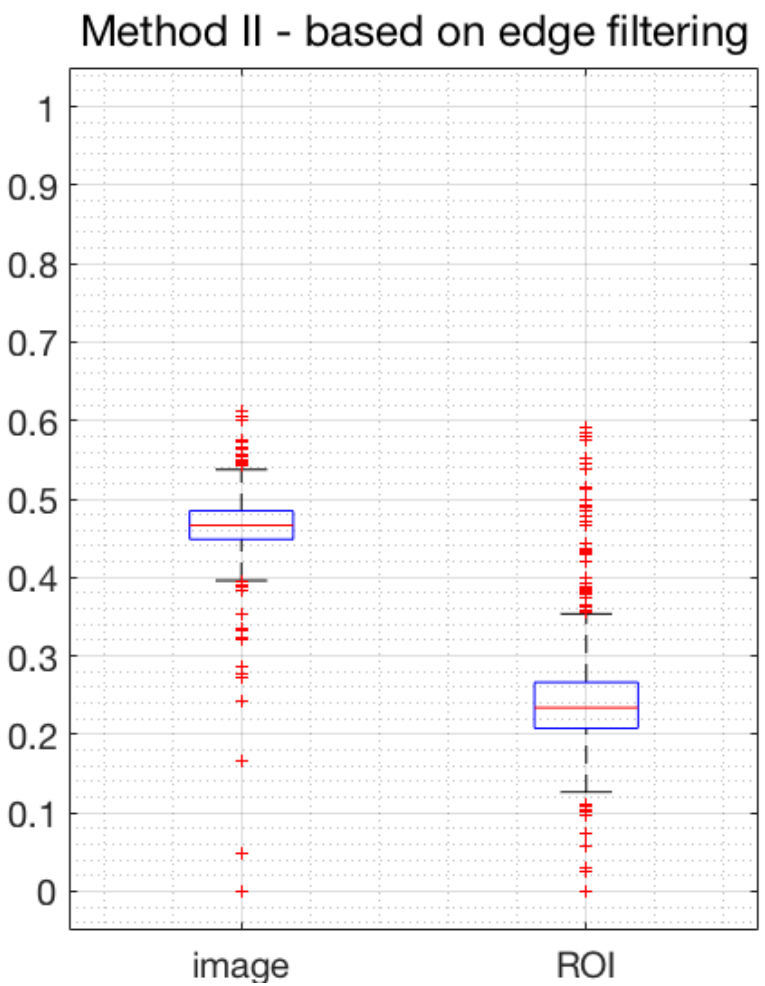
Matching vein patterns

- Hand segmentation from TH and NIR images
- Detect and extract SURF features
- Finding transformation matrix - NIR mask to TH mask
- Apply a transformation matrix to vein pattern



Q2. What is the similarity between vein patterns obtained from NIR and TH images?





Hamming Distance between the vein patterns extracted from NIR and TH

Q3. What is the effectiveness of systems based on texture descriptors?

Experiment 1: Identity recognition using texture descriptors

a. **Changes over time for each image type** - the reference database contained data from the first session. Data from subsequent sessions were test data for comparisons between sessions.

b. **The size of the reference template** - assuming that each session contains data acquired in different conditions (various factors were related to the measurement process and human physiology, and the time between sessions is irrelevant).

c. Multispectral descriptors (TODO) - in this case, the feature vectors obtained from the VIS, NIR and TH images were merged into one larger feature vector for each person.

d. Score fusion (TODO) - the results of the comparisons were combined using score combination methods:
   - objective - determination of a new score by calculating the mean score, product, median value
   - weighted sum
   - logistic linear regression

BSIF: Binarized Statistical Image Features

- generates a series of binary images that can be used as binary codes
- each bit of the output binary code is associated with a specific filter
- the code length is determined by number of filters in the filter bank
- 8 window sizes: from 3×3 to 17×17 pixels
- filters obtained using ICA, trained on natural images *e.g.* grass, stones, fur, trees, landscapes (Huangac, 2008)

	VIS			NIR			TH		
	EER	EER_{ROT}	$diff$	EER	EER_{ROT}	$diff$	EER	EER_{ROT}	$diff$
3×3	38.75%	38.63%	0.12%	33.64%	31.74%	1.90%	42.59%	42.15%	0.44%
5×5	**37.57%**	**37.13%**	0.44%	35.68%	34.25%	1.44%	44.61%	43.80%	0.81%
7×7	51.16%	50.46%	0.70%	42.96%	37.98%	4.98%	46.34%	44.86%	1.48%
9×9	43.63%	44.03%	-0.39%	26.41%	26.25%	0.16%	41.16%	41.13%	0.02%
11×11	44.40%	44.28%	0.12%	24.53%	24.04%	0.49%	**40.70%**	**40.69%**	0.00%
13×13	44.84%	44.31%	0.53%	22.61%	22.81%	-0.21%	41.55%	41.32%	0.23%
15×15	42.80%	42.45%	0.35%	**22.19%**	**22.42%**	-0.23%	41.13%	41.13%	0.00%
17×17	44.14%	43.94%	0.20%	24.86%	24.74%	0.12%	40.74%	40.47%	0.27%

Changes over time for each image type

Value Stability Maps (VSM) - maps of deviations of values from several images (analogue to Thermal Stability Maps (Bartuzi, 2017))

		filter	$S1 \rightarrow S2$	$S1 \rightarrow S3$	$S1 \rightarrow S4$	$S1 \rightarrow S5$	*std*
SEGMENTED	VIS	5×5	38.70%	36.46%	37.99%	35.92%	1.30%
	NIR	15×15	26.87%	24.61%	23.72%	24.15%	1.40%
	TH	17×17	43.83%	30.99%	40.88%	41.68%	5.71%
SEGMENTED + **VSM**	VIS	5×5	20.38%	19.03%	19.22%	21.34%	1.08%
	NIR	15×15	9.56%	10.23%	8.10%	7.44%	1.29%
	TH	17×17	10.39%	17.71%	22.45%	22.71%	5.51%

Mini-conclusions:

- the time difference between sessions is not related to the results for images NIR or VIS, larger deviations of EER for NIR (due to physiological differences and external conditions)
- the use of VSM reduces EER from $\sim$10% to $\sim$ 30%
- the smallest error rates for NIR

The size of the reference template

		$filter$	$S1 \rightarrow S5$	$S1{:}S2 \rightarrow S5$	$S1{:}S3 \rightarrow S5$	$S1{:}S4 \rightarrow S5$
EER	VIS	5×5	21,34%	25,39%	22,31%	16,98%
	NIR	15×15	7,44%	6,47%	6,30%	6,29%
	TH	17×17	22,71%	23,64%	23.53%	21,64%
Accuracy	VIS	5×5	24,82%	27,95%	41,35%	56,74%
	NIR	15×15	87,64%	90,73%	92,78%	92,80%
	TH	17×17	17,22%	19,15%	23,11%	23,13%

Mini-conclusions:

- very poor identification results for NIR and TH
- a small improvement in results with the increase of the reference base for images from subsequent sessions

Q4. What is the effectiveness of systems based on CNN?

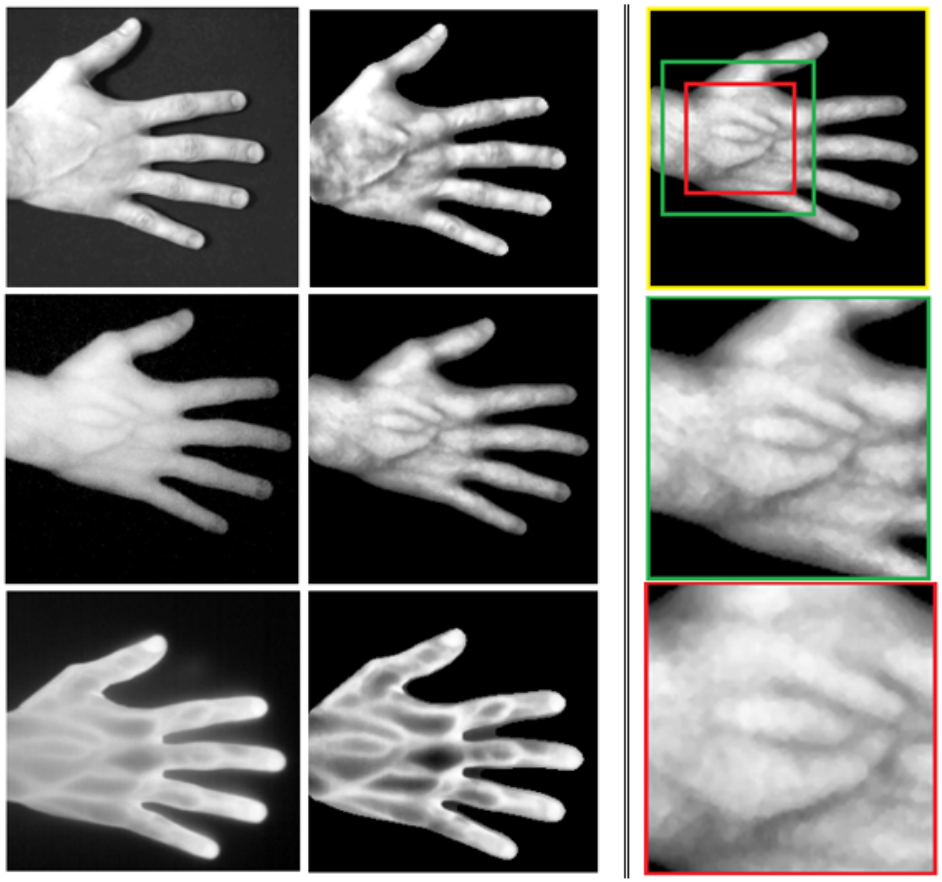
Experiment 2: Identity recognition based on CNN

a. **Impact of the region of interest content** - analysis of input data containing whole hand, central part of the hand with and without margins

b. **Input data comparison** - input data in form of raw images and preprocessed images

c. **Two CNN structures** - simple structure trained on hand images and well known VGG fine-tuned using transfer learning on hand images

d. **Fusion method** - score fusion, multispectral input



CNN-based approach - analyzed structures

	SimpleNet-4	**VGG-16**
layers	4 convolutional layers with ReLU and max pooling 1 fully connected softmax	13 convolutional layers with ReLU and max pooling 3 fully connected softmax
sizes of filters	$17 \times 17, 13 \times 13, 9 \times 9, 7 \times 7$	3×3
no. of filters	N, 2N, 4N, 8N	3, 2:64, 2:128, 3:256, 5:512
optimization	SGD, $m = 0.9$	SGD, $m = 0.9$
learning rate	$0.05 - 0.0005$	0.0001
training	on thermal images	on natural images and fine-tuned using transfer learning on thermal images
other	data augmentation	data augmentation 50% dropout

Results for VGG

Spectrum	Input	Accuracy [%]			EER [%]		
		preprocessed images	*raw images*	*diff*	*preprocessed images*	*raw images*	*diff*
NIR	small ROI	87.11 (±0.27)	87.63 (±0.32)	-0.52	4.90 (±0.13)	4.66 (±0.18)	-0.25
	ROI	90.72 (±0.25)	91.75 (±0.26)	-1.03	2.08 (±0.04)	2.10 (±0.06)	0.02
	image	94.33 (±0.21)	95.36 (±0.20)	-1.03	2.07 (±0.15)	1.70 (±0.02)	-0.37
VIS	small ROI	78.87 (±0.33)	89.18 (±0.23)	-10.31	4.17 (±0.01)	3.67 (±0.01)	-0.49
	ROI	89.69 (±0.22)	95.36 (±0.15)	-5.67	3.61 (±0.15)	1.14 (±0.02)	-2.47
	image	98.45 (±0.09)	100.00 (±0.00)	-1.55	0.57 (±0.01)	0.86 (±0.12)	0.29
TH	small ROI	87.63 (±0.23)	91.75 (±0.19)	-4.12	4.17 (±0.01)	1.56 (±0.02)	-2.61
	ROI	95.88 (±0.14)	99.48 (±0.05)	-3.61	1.03 (±0.06)	0.52 (±0.01)	-0.51
	image	97.94 (±0.10)	100.00 (±0.00)	-2.06	1.04 (±0.01)	0.03 (±0.00)	-1.01

Mini-conclusions for VGG:

- the highest accuracy and the smallest EER for thermal images
- the smaller the content of the input image, the worse the results
- VGG prefers raw images

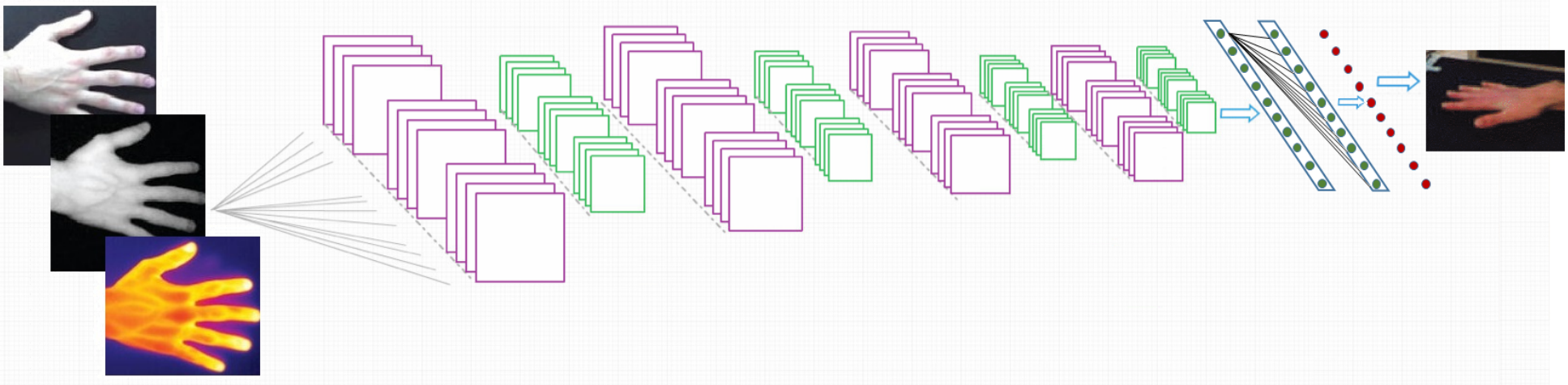
Results for VGG - score fusion

	Accuracy [%]				EER [%]			
	image		ROI		image		ROI	
NIR	95.36	(±0.20)	91.75	(±0.26)	1.7000	(±0.02)	2.0986	(±0.06)
VIS	100.00	(±0.00)	95.36	(±0.15)	0.8595	(±0.12)	1.1430	(±0.02)
TH	100.00	(±0.00)	99.48	(±0.05)	0.0348	(±0.00)	0.5201	(±0.01)
NIR+VIS	100.00	(±0.00)	98.45	(±0.09)	0.0348	(±0.00)	0.6128	(±0.0142)
NIR+TH	100.00	(±0.00)	98.97	(±0.07)	0.0348	(±0.00)	0.5208	(±0.0120)
VIS+TH	100.00	(±0.00)	99.48	(±0.05)	0.0348	(±0.00)	0.5066	(±0.0088)
NIR+VIS+TH	100.00	(±0.00)	100.00	(±0.00)	0.0348	(±0.00)	0.0107	(±0.0009)

Mini-conclusions for VGG:

- identification is satisfactory for single modality
- verification: the fusion of modality gives a smaller EER (EER $= 0.01\%$ for NIR+VIS+TH fusion)

SimpleNet-4 structure



$17 \times 17 \times N$	$13 \times 13 \times 2 \cdot N$	$9 \times 9 \times 4 \cdot N$	$7 \times 7 \times 8 \cdot N$	fully connected
Relu	Relu	Relu	Relu	layer
pooling	pooling	pooling	-	*softmax*

Results for SimpleNet-4

	Input	Accuracy [%]	Accuracy (top-5) [%]
single modality	NIR	95.95 (±0.87)	98.94 (±0.01)
	VIS	93.24 (± 1.64)	98.65 (± 0.78)
	TH	91.84 (± 3.17)	98.02 (± 0.86)
two modalities	NIR & VIS	97.30 (±0.87)	98.95 (±0.00)
	NIR & TH	95.95 (± 1.64)	98.65 (± 0.78)
	VIS & TH	94.60 (± 3.17)	99.47 (±0.00)
three modalities	NIR & VIS & TH	98.65 (±0.57)	100.00 (±0.00)

Mini-conclusions for SimpleNet-4:

- the highest accuracy for NIR images
- the fusion of two modalities increases accuracy by ∼2%, while the fusion of all three modalities allows to achieve Accuracy = 98.65%

Conclusions

- THID - database characteristics
  - correlation between image types: high degree of similarity between NIR and VIS images
  - image quality: a lot of over- or underexposed NIR images, TH - low quality in terms of blurring, artifacts *etc.*
- Similarity between vein patterns obtained from NIR and TH images
  - *Hamming distance* between vein patterns extracted from NIR and TH: $\sim$ 0.5 for whole hand and $\sim$0.2 for ROI. *HDs* are similar for both methods.
- Texture descriptors
  - the smallest errors are obtained for NIR images (EER ¡ 8%), the largest deviation is observed for TH images (EER$\simeq$20%, std¿5%)
- CNN approach
  - VGG: better results for TH images
  - SimpleNet-4: best results for NIR, and the fusion of modalities increases accuracy
  - pre-processing of data or determining too small regions of interest reduces effectiveness

Further work

- assessment of verification and identification based on vein patterns extracted from THID images
- selection of a more accurate features (vein) extractor
- analysis of various data fusion models and multibiometry techniques
- use of other CNN models
- results for vein patterns obtained from NIR and TH
- solution implementation using the capabilities of mobile devices

Hand Imaging in different spectral ranges

Ewelina Bartuzi
Advisor: prof. Andrzej Pacut

Biometrics and Machine Learning Groups
Institute of Control and Computation Engineering
Faculty of Electronics and Information Technology, WUT

13 czerwca 2018