Thermal Features for Presentation Attack Detection in Hand Biometrics

Ewelina Bartuzi,
Advisor: prof. Andrzej Pacut

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

Seminarium naukowe 3

Contents

- hand recognition method based on a deep convolutional neural network (DCNN) model trained on images of different types in respect to: **quality** (higher and lower), **spectrum** (visible light images, thermal maps, and images combining thermal and visible-light images)
- utilizing thermal features or visible-light images of the hand for the purpose of presentation attack detection (PAD) in two different operational modes: authenticity- and identity-driven
- using data acquired by mobile device

Personal features of the hand

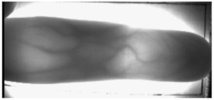
fingerprint

Source: FVC2004

palmprint

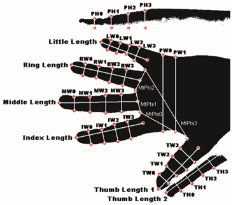
Source: IIT Delhi

finger vein pattern



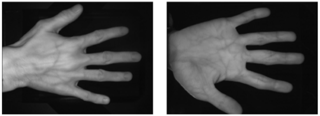
Source: PolyU

geometric features



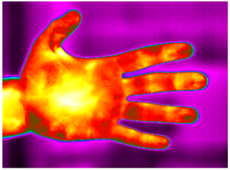
Source: Use of hand in biometrics, A. Czajka

hand vein pattern



Source: A Novel Biometric System Based on Hand Vein, X. Wu et al.

thermal features



Source: BioBase-Hand-Thermal

Presentation Attack Detection for palms: a review

- **features:** palmprints, vein patterns of dorsal side of the hand (taken in near-infrared light)
- **fake samples:** printouts
- **methods:** texture features (LBP, HOG, LoG), set of statistic features
- **fake detection error rate:** 0.16 - 2.73%

No papers employing thermal features for PAD or CNNs.

Motivations

1. Increase of interest in biometric solutions for mobile devices
2. Using built-in component in mobile phones
3. Advantages of hand measurement process:
    - social acceptance (53/53 subjects in MobiBits, German survey: most people accept fingerprint, unaccepted: signature, voice)
    - rarely exposed in whole
    - hygienic, contactless acquisition
    - convenient measurement (also with flash)
    - thermal features are independent of external light
    - difficult to reconstruct heat maps

Experimental data

Dataset of visible light and thermal hand images

- ***MobiBits*** – a subset of a multimodal biometric database including images of palm side of the hand
- **106 classes** $\times$ 3 sessions $\times$ 45 images $\times$ 2 types: **visible light** and **thermal** images
- **three sessions**: with no temperature influence, after warming, and after cooling
- **acquisition**: unconstrained (raised hand) and stabilized by glass stand

◇ **RGB** – images taken with a rear camera of the CAT s60 mobile phone (480$\times$640 px)

◇ **TH** – thermal images, taken simultaneously with RGB images (240$\times$320 px)

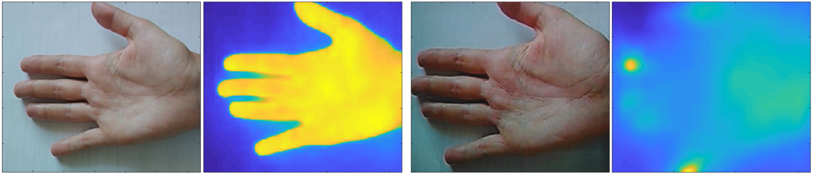
◇ **MSX** – images using the FLIR MSX technology combining thermal images and visible light images at the pixel level (CAT s60, 480 $\times$ 640 pixels)

◇ **HQ** – higher resolution visible-light images taken with rear camera of smartphone (Huawei Mate S, 13 Mpx) with and without flash.

Experimental data

Fake hand representations

- **RGB** – photographed printouts
- **TH** – heat distribution of hand imitated by the hand of a living human placed under the printout, acquired by CAT s60



Quality of data

Tabela: Mean quality indicators for different image types and for *fake* samples (calculated in conformance to ISO/IEC 29794-6:201x(E)).

Quality indicators	HQ with flash	HQ without flash	RGB	RGB fakes	TH	TH fakes
intensity	55	165	159	134	125	130
sharpness	38.46	17.38	36.02	27.12	6.05	0.00
contrast	15.28	15.64	15.78	16.08	9.78	8.59

Experimental scenarios

Palm Recognition:

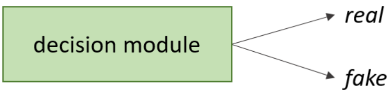
- CNN - based method
- RGB images, TH images, MSX images
- dependence on age and gender

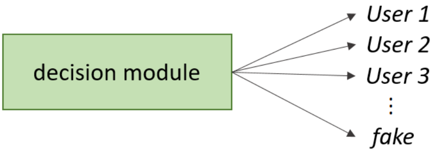
Presentation Attack Detection:

- Authenticity-driven mode:
  - 11 statistical features + PCA + SVM
  - LBP + PCA + SVM
  - BSIF + PCA + SVM
  - CNN - based method
- Identity-driven mode:
  - AlexNet, VGG-19

Experimental scenarios for PAD

Authenticity-driven mode	Identity-driven mode
binary classification **open-set** (subject-disjoint) **2 images per subject** 200 real and 200 fake images	**class-wise prediction** 106 identity classes + 1 class of fake representations **closed-set** (sample-disjoint) **45 images per class**
feature vector + SVM classifier (11 statistical features, LBP, BSIF) **CNN- based method** (AlexNet, VGG-19)	**CNN- based method** (AlexNet, VGG-19)





Employed feature extraction algorithms for PAD

Feature type	Details
Method I: **Statistical features**	**vector of eleven features:** F1: mean, F2: variance, F3: skewness, F4: kurtosis, F5-F7: 10^{th}, 50^{th}, 90^{th} percentile of the image pixel intensities, F8-F9: variance of wavelet coefficients in the first and second level vertically oriented sub-bands, F10: their ratio, F11: kurtosis of the second level vertically oriented sub-band
Method II: **LBP features**	LBP histogram 59-dimensional feature vector 8 neighbours radius $= 1$
Method III: **BSIF features**	BSIF histogram 256-dimensional feature vector filter size: 17×17 px no. filters in set: 8

DCNN architectures

AlexNet	**VGG-19**
'shallow'	'very deep'
5 convolutional layers	16 convolutional layers
3 fully connected layers	3 fully connected layers

- input: $224 \times 224 \times 3$
- pre-trained models with modified bottleneck layers
- fine-tuned with a dataset of Mobibits

Training and evaluation

Overall:

- 10 data splits into **train/validation/test** subset in ratio 60:20:20 (subject-disjoint for authentication and sample-disjoint for identification)
- the networks were trained separately with RGB and TH images (in each mode)
- interpretation of scores obtained from RGB images, TH images separately and by averaging using metrics: classification accuracy, APCER, BPCER

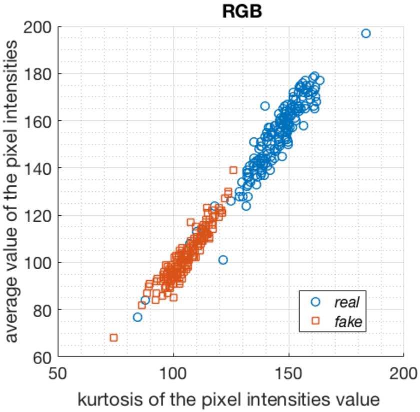
Training:

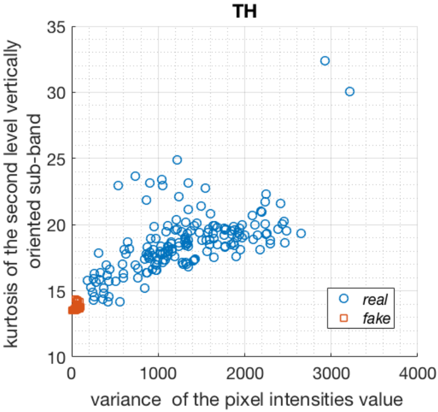
- data shuffling before each training epoch
- optimizer - stochastic gradient descent (momentum = 0.9, learning rate of 0.0001)

Evaluation:

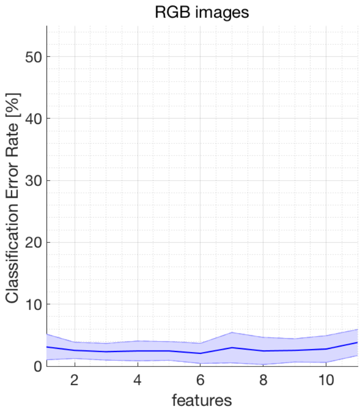
- weights were determined using validation stopping of network training with patience of 10 epoch.

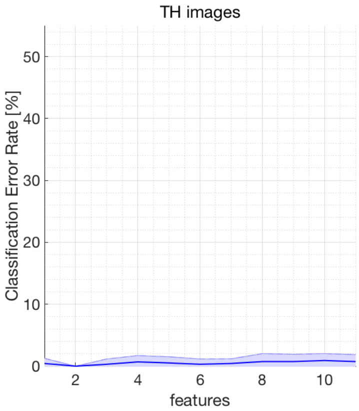
11 statisticall features



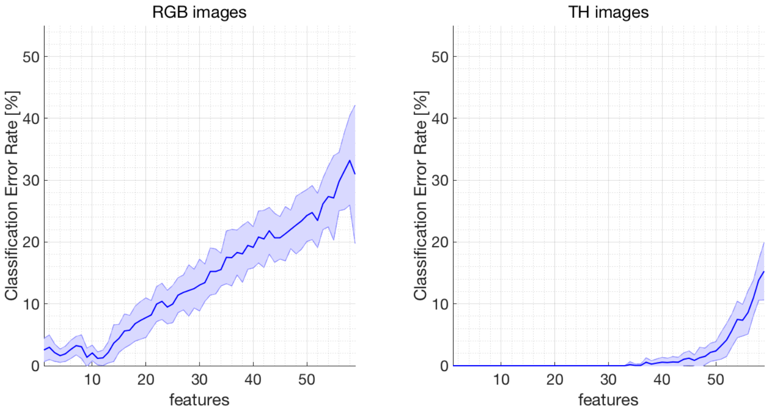


11 statisticall features

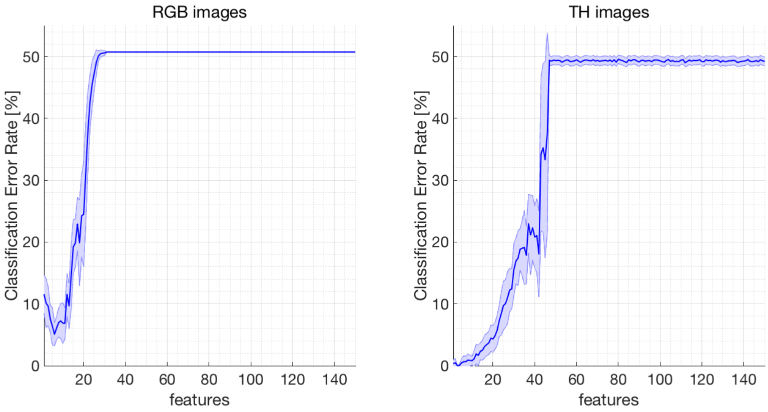




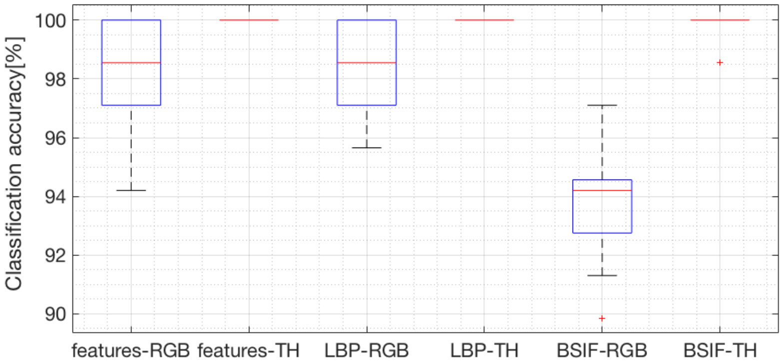
LBP histogram features



BSIF histogram features



Results: feature vectors + SVM - boxplots



Results: feature vectors + SVM - EER, APCER, BFCER

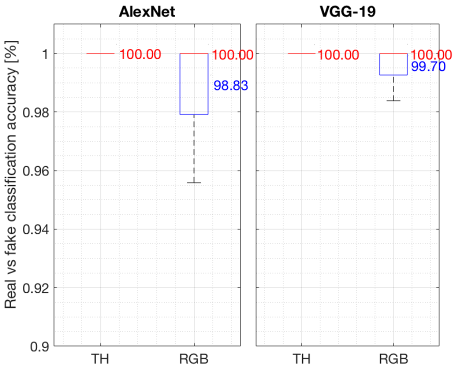
		statistical features	LBP features	BSIF features
RGB	EER[%]	2.06($\pm$0.62)	1.19($\pm$0.95)	5.09($\pm$1.89)
	APCER[%]	0.00($\pm$0.00)	0.61($\pm$0.42)	2.72($\pm$0.72)
	BPCER[%]	4.11($\pm$2.23)	2.38($\pm$1.52)	(7.46($\pm$2.38)
TH	EER[%]	0.00($\pm$0.00)	0.00($\pm$0.00)	0.02($\pm$0.01)
	APCER[%]	0.00($\pm$0.00)	0.00($\pm$0.00)	0.00($\pm$0.00)
	BPCER[%]	0.00($\pm$0.00)	0.00($\pm$0.00)	0.03($\pm$0.01)

Results: authenticity-driven mode

- **Thermal features** allowed to discern *fake* representation from *real* ones with 100% accuracy for both analyzed CNN structures
- APCER = BPCER = 0.00%
- Utilizing **visible-light palm images** allowed to obtain accuracy of 98.83% for AlexNet and 99.70% for VGG-19.
- AlexNet:
  APCER = 0.87%, BPCER = 0.55%
  VGG-19:
  APCER = 0.29%, BPCER = 0.97%

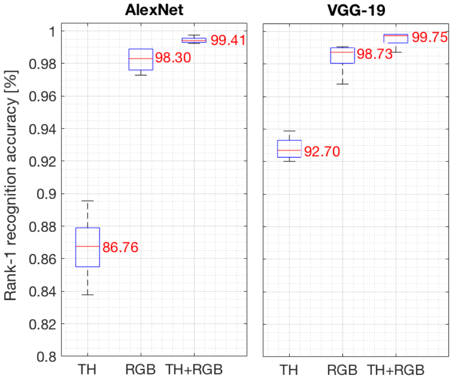
Conclusions:

Thermal hand maps may distinguish *real* and *fake* representations with perfect effectiveness.



Results: identity-driven mode

- Higher accuracy was obtained for VGG-19 structure.
- Accuracy of recognition thermal samples equal 86.76% for AlexNet and 92.70% for VGG-19.
- Utilizing visible light images gives 98.30% and 98.73% accuracy.
- Averaging scores obtained for **visible light** and **thermal** palm images allowed to obtain accuracy of 99.41% and 99.75% for AlexNet and VGG-19, respectively.



Conclusions:

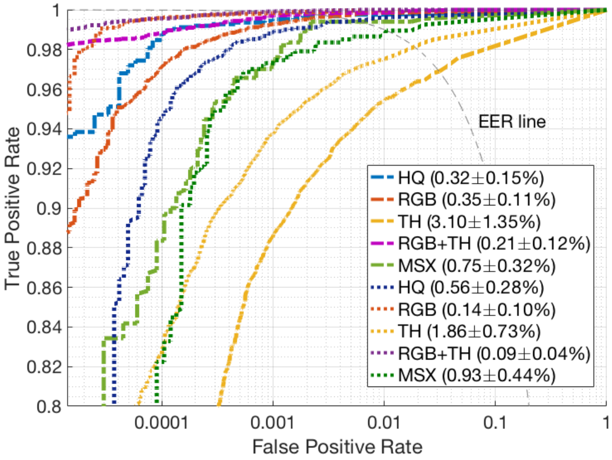
Thermal features improve the overall performance of biometric recognition system.

Conclusions

1. Thermal features are promising as presentation attack detection cues
2. **Authenticity-driven mode** - discerning *fake* representations from *real* ones achieves 100% accuracy using thermal images.
3. **Identity-driven mode** - closed-set classification accuracy reaches 99.75%
4. Thermal features improve the overall performance of biometric recognition system
5. Trained DCNN model weights, example source codes, and a dataset of fake hand representations for a subset of the MobiBits database is made available to interested researchers for non-commercial purposes

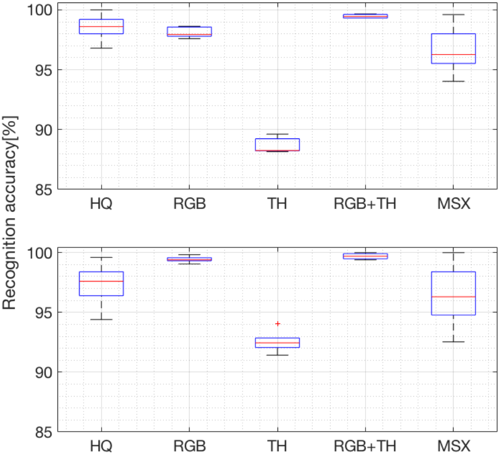
Results: Palm recognition

ROC curves for hand recognition using different image types: HQ, RGB, TH, MSX and for scores obtained by averaging the TH and RGB scores. Results obtained for the AlexNet are plotted with dashed line and for the VGG-19 model with dotted line.



Results: Palm recognition

Boxplots representing differences in accuracy of classification into all hand classes for different hand representations for two DCNN models (AlexNet in the top, VGG-19 in the bottom). .



Results: Palm recognition

Tabela: Mean values of EERs and accuracy in respect to age of subjects.

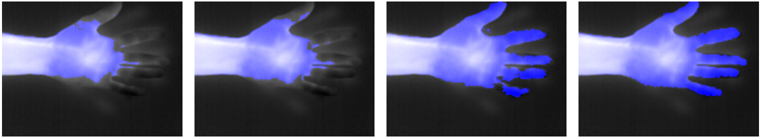
	AlexNet		***VGG-19***	
	EER[%]	Accuracy[%]	EER[%]	Accuracy[%]
RGB				
15-25	0.36	98.46	0.08	99.77
26-35	0.30	98.17	0.11	99.39
36-43	0.30	97.43	0.07	99.34
> 45	0.32	99.01	0.31	99.29
TH				
15-25	2.45	90.31	1.31	93.15
26-35	2.83	88.75	1.67	93.21
36-43	2.87	87.25	2.12	92.75
> 45	3.64	88.21	2.27	91.49

Results: Palm recognition

Tabela: Mean values of EERs and accuracy in respect to gender of subjects.

	AlexNet		***VGG-19***	
	EER[%]	Accuracy[%]	EER[%]	Accuracy[%]
HQ				
female	0.21	99.60	0.33	97.22
male	0.28	98.54	0.98	97.98
RGB				
female	0.37	98.27	0.06	99.66
male	0.28	98.00	0.19	99.28
TH				
female	2.62	90.15	1.52	93.64
male	3.28	87.67	2.07	91.78

Segmentation of thermal spectrum hand images with a pre-trained off-the-shelf DCNN model



Motivation

- one of the most important stages in biometric sample processing is image segmentation
- thermal hand segmentation is a trivial task in controlled environment, but ...
- can be difficult for unconstrained sample acquisition with a thermal camera
- proposition of using a pre-trained DCNN modelfor sementic segmentation
- comparing proposed method with conventional methods such as Otsu's thresholding and method based on Gaussian Mixture Modeling

Hand segmentation methods

Visible light images:

- simple thresholding methods - Otsu's algorithm
- RGB $\mapsto$ HSV $\mapsto$ fuzzy multiscale aggregation
- RGB $\mapsto$ HSV $\mapsto$ mutli-thresholding
- skin color model + NN $\mapsto$ edge detection + voting techniques
- DCNN: SegNet and U-shape models

Thermal images:

- thresholding methods: Otsu's algorithm, GMM-based
- active shape model
- segmentation using masks after geometric transformation

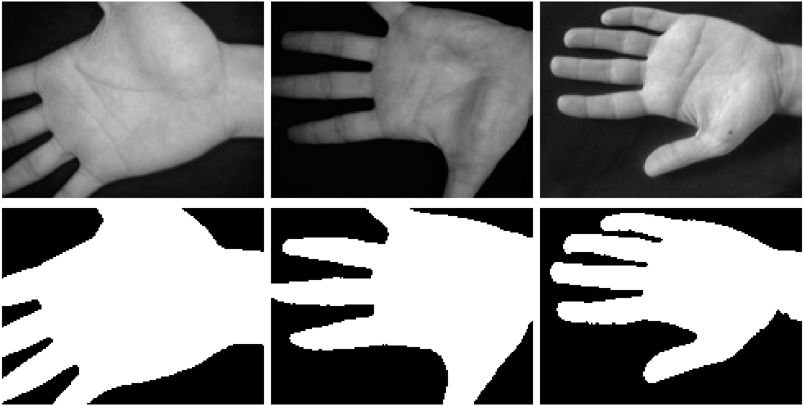
Experimental data

	CASIA - PalmprintV1	Warsaw-BioBase-Hand-Thermal	Tecnocampus Hand Image Database
no subject	5 501	21 000	1 000
no images	312	70	100
image size	640×480	640×480	320×240
image types	**near-infrared**	**thermal images**	**thermal images** near-infrared visible light
acquisition protocol	semi-constrained	unconstrained	semi-constrained
purpose	training backround	segmentation	segmentation
acronym	*CASIA*	*Warsaw*	*THID*

Ground truth binary masks:

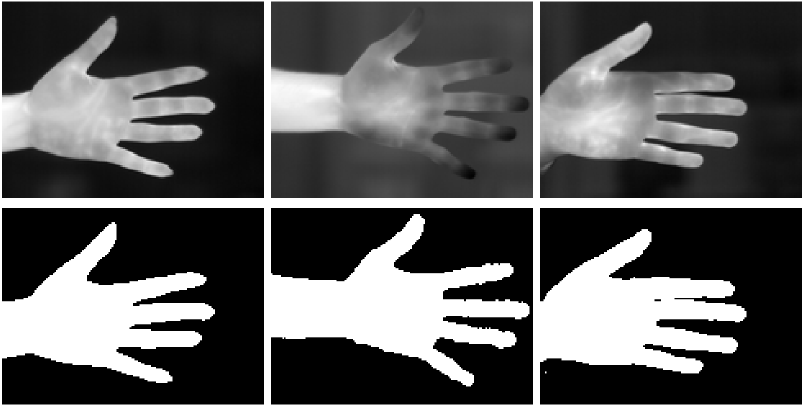
- *CASIA*: 5501 masks for 312 subjects (automatically obtained using thresholding)
- *Warsaw*: 734 masks for 70 subjects (mannually prepared masks)
- THID: 731 masks for 85 subjects (mannually prepared masks)

CASIA - PalmprintV1



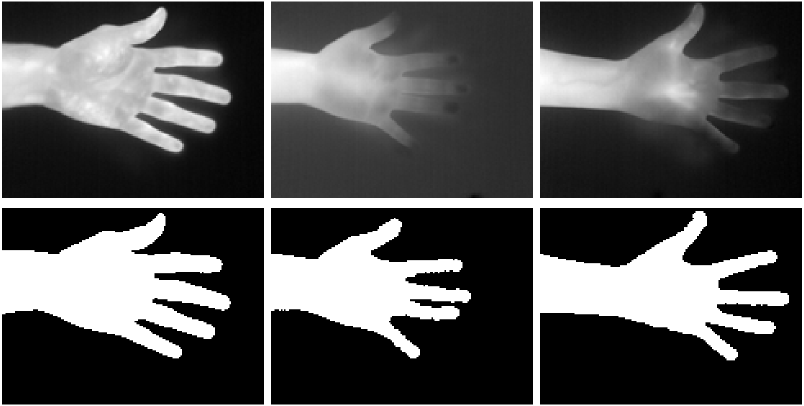
Rysunek: Example data from the CASIA-PalmprintV1 database **(top row)** and their respective ground truth binary masks **(bottom row)**.

Warsaw-BioBase-Hand-Thermal- v1



Rysunek: Example data from the Warsaw-BioBase-Hand-Thermal-v1 database following normalization **(top row)** and their respective ground truth binary masks **(bottom row)**.

Tecnocampus Hand Image Database



Rysunek: Example data from the Tecnocampus Hand Image Database following normalization **(top row)** and their respective ground truth binary masks **(bottom row)**.

Image Qality (1/2)

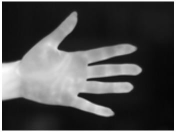
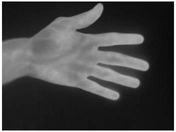
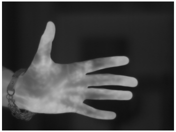
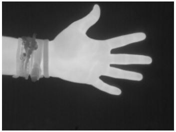
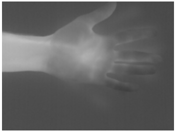
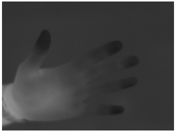
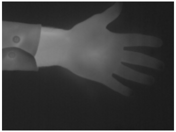
Class	Description	Warsaw	THID
I *warm*	images presenting a palm (and possibly a part of the wrist) against a colder background 33.79% of *Warsaw* 36.78% of *THID*		
II *warm with intrusions*	images similar to those from Class I, but with additional visible clothing and/or jewelry, wristwtches, etc. 33.11% of *Warsaw* 20.16% of *THID*		

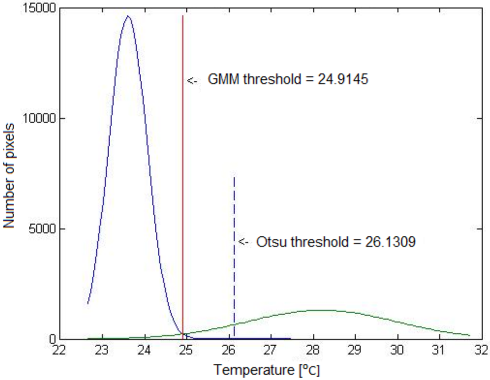
Image Qality (2/2)

Class	Description	Warsaw	THID
III *cold*	images presenting hands with cooler regions, which temperature is similar to this of the background or lower 16.28% of *Warsaw* 23.98% of *THID*		
IV *cold with intrusions*	images similar to those from Class III, but with additional visible clothing and/or jewelry 16.82% of *Warsaw* 6.40% of *THID*		
V *heat shade*	images with heat-shade effect caused by hand movement during image acquisition none in *Warsaw* 12.68% of *THID*	–	

Baseline conventional segmentation methods

Otsu's thresholding and subsequent binarization; Otsu's method selects the threshold by maximizing the inter-class variance (between the background class and the object), and minimalizing the intra-class variance without making any assumptions on the pixel intensity distributions.

Gaussian Mixture Models (GMM), which approximate the distributions of pixels belonging to the hand and those of the background, calculate the intersection point of two Gaussian curves, which approximate the distributions of object's values and of the background's values.



Segmentation accuracy metrics

- Intersection over Union, a metric typically seen in segmentation tasks:
$$IoU = \frac{prediction \cap ground\_truth}{prediction \cup ground\_truth}$$
or
$$IoU = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} P_{ij} \wedge G_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} P_{ij} \vee G_{ij}}$$
- E_1 error metric:
$$E_1 = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} P_{ij} \oplus G_{ij}$$
where P_{ij} and G_{ij} denote the logical values of prediction mask and ground truth mask for the ij-th pixel, respectively, m, n is the image size in pixels, and $\oplus$ denotes the XOR (exclusive or) bitwise logical operator.

DCNN model architecture

- SegNet architecture, which is build around a fully convolutional encoder-decoder architecture
- The encoder stage employs a VGG-16 model graph, whereas the decoder comprises several sets of convolution and upsampling layers, whose target is to retrieve spatial information from the encoder output, to yield a dense, pixel-wise output map of the same size as the input image.
- fine-tune the off-the-shelf weights of the SegNet model pre-trained on the ImageNet database with datasets of thermal hand images and their corresponding ground truth masks,

Training and evaluation of the segmentation method

The devised experiments included tests **within-dataset** (highlighted in blue) and **cross-dataset** performance of the proposed solution:

- training and testing on *Warsaw*
- training and testing on *THID*
- training on *Warsaw*, testing on *THID*
- training on *THID*, testing on *Warsaw*
- training and testing on both datasets

Then, in the second part of the evaluation, the *CASIA* database is included in the training phase. Since it contains a much larger number of images with corresponding ground truth labels than the other two datasets (5501 vs ≈ 730, albeit these are not thermal images, but rather near-infrared ones), the goal is to help the network learn the typical shape of a human hand.

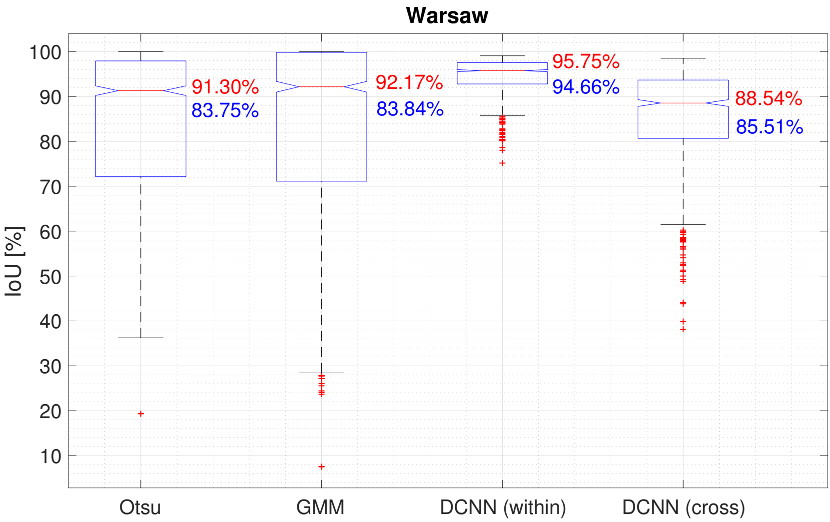
Training and evaluation of the segmentation method

- 10 randomly created subject-disjoint train/test data splits in a ratio of 0.8:0.2.
- the network was trained with stochastic gradient descent as the optimization method.
  Momentum of 0.9, learning rate of 0.001 decreased 10-fold after every 50 epochs, and L2 regularization of 0.0001 were used. Batch size was 4 and the data were shuffled after each epoch.

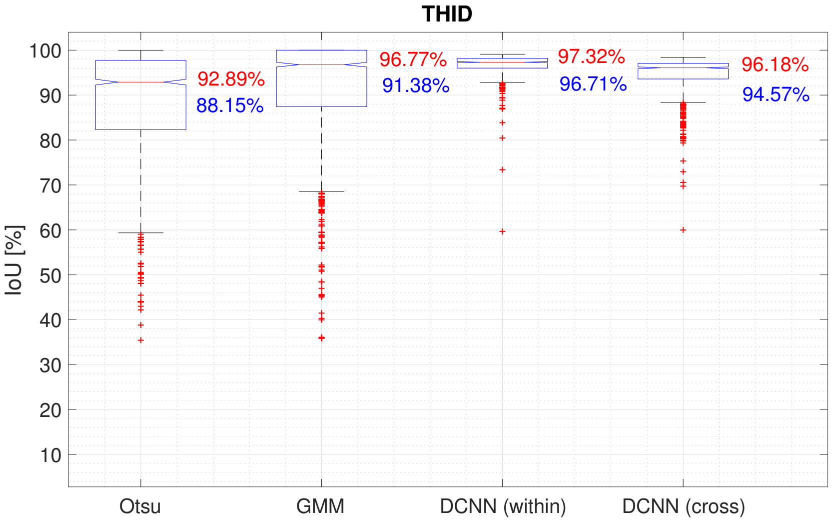
Experimental results

	Mean IoU	Mean E_1		Mean IoU	Mean E_1
Otsu			**GMM**		
Warsaw	83.75%	4.86%	*Warsaw*	83.84%	4.73
THID	88.15%	3.61%	*THID*	91.38%	2.62%
CNN-based method:					
Train: *Warsaw*, test: *Warsaw*	92.75%	2.14%	Train: *Warsaw+CASIA*, test: *Warsaw*	93.98%	1.76%
Train on both, test: *Warsaw*	93.68%	1.83%	Train on all three, test: *Warsaw*	**94.66%**	1.53%
Train: *THID*, test: *Warsaw*	78.51%	6.54%	Train: *THID+CASIA*, Test: *Warsaw*	85.51%	4.37%
Train: *THID*, test: *THID*	94.42%	1.68%	Train: *THID+CASIA*, test: *THID*	95.48%	1.36%
Train on both, test: *THID*	94.46%	1.67%	Train on all three, test: *THID*	**96.71%**	0.98%
Train: *Warsaw*, test: *THID*	94.57%	1.60%	Train: *Warsaw+CASIA*, test: *THID*	94.30%	1.70%

Experimental results

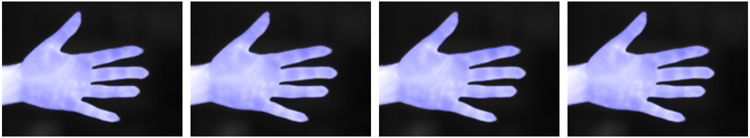


Experimental results - boxplot for *THID*

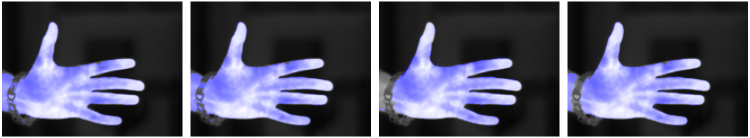


Experimental results - *Warsaw (1/2)*

class I: warm



class II: warm with intrusions



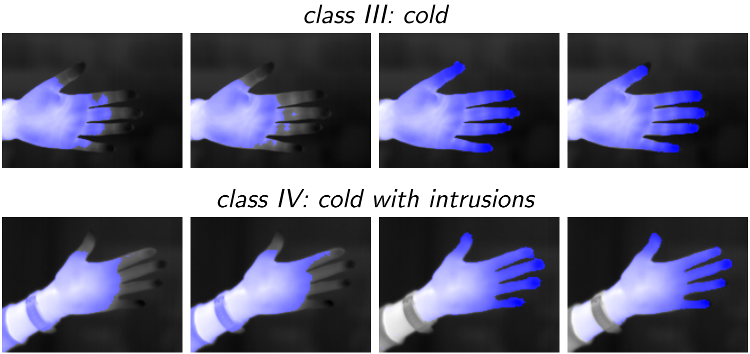
(a) Otsu segmentation

(b) GMM segmentation

(c) DCNN-based segmentation

(d) Ground truth

Experimental results - *Warsaw (2/2)*

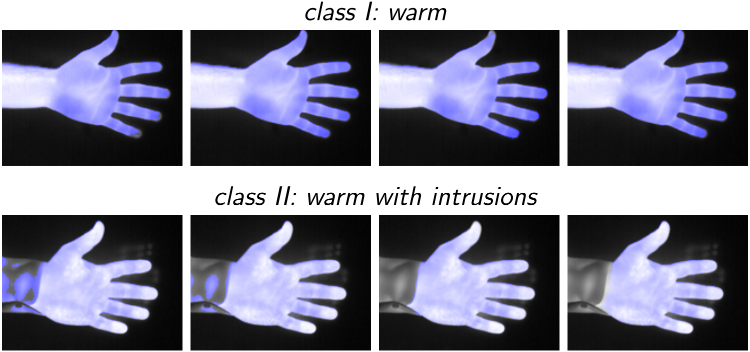


(a) Otsu segmentation (b) GMM segmentation (c) DCNN-based segmentation (d) Ground truth

Experimental results

	Class I	Class II	Class III	Class IV	Mean IoU
Otsu: *Warsaw*	97.16%	90.03%	67.87%	62.11%	83.75%
GMM: *Warsaw*	98.11%	90.13%	65.43%	63.11%	83.84%
CNN-based method:					
Train: *Warsaw*, test: *Warsaw*	96.62%	93.08%	91.26%	85.91%	92.75%
Train on both, test: *Warsaw*	97.42%	94.08%	92.48%	88.57%	93.68%
Train on *THID*, test: *Warsaw*	84.46%	80.02%	77.19%	64.74%	78.51%
Train: *Warsaw+CASIA*, test: *Warsaw*	97.35%	94.48%	92.38%	87.98%	93.98%
Train on all three, test: *Warsaw*	97.59%	94.67%	93.58%	90.26%	**94.66%**
Train: *THID+CASIA*, test: *Warsaw*	90.85%	85.57%	82.20%	72.60%	85.51%

Experimental results - *THID* (1/2)



(a) Otsu segmentation

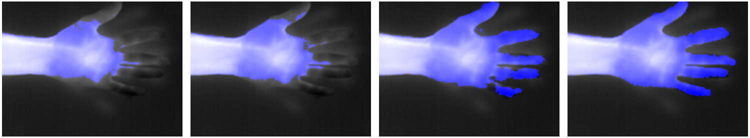
(b) GMM segmentation

(c) DCNN-based segmentation

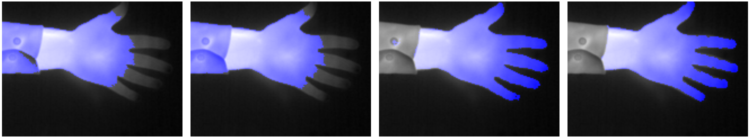
(d) Ground truth

Experimental results - *THID* (2/2)

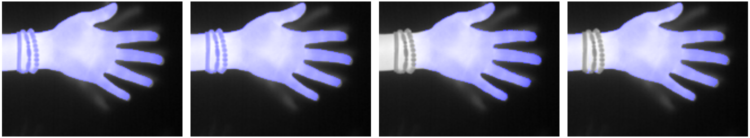
class III: cold



class IV: cold with intrusions



class V: heat shade



(a) Otsu (b) GMM (c) DCNN-based (d) Ground truth

Experimental results

	Class I	Class II	Class III	Class IV	Class V	Mean
Otsu: *THID*	96.13%	86.12%	77.51%	63.84%	93.86%	88.15%
GMM: *THID*	98.76%	88.67%	82.44%	67.68%	96.84%	91.38%
CNN-based method:						
Train: *THID*, test: *THID*	96.55%	93.69%	92.02%	88.51%	95.49%	94.42%
Train on both, test: *THID*	97.07%	95.44%	91.74%	90.83%	95.34%	94.46%
Train: *Warsaw*, test: *THID*	96.86%	93.64%	92.67%	88.40%	96.00%	94.57%
Train: *THID+CASIA*, test: *THID*	97.35%	94.72%	93.37%	90.90%	96.40%	95.48%
Train on all three, test: *THID*	97.83%	95.85%	95.49%	93.66%	97.20%	**96.71%**
Train: *Warsaw+CASIA*, test: *THID*	96.92%	92.42%	92.65%	88.23%	95.85%	94.30%

Conclusion

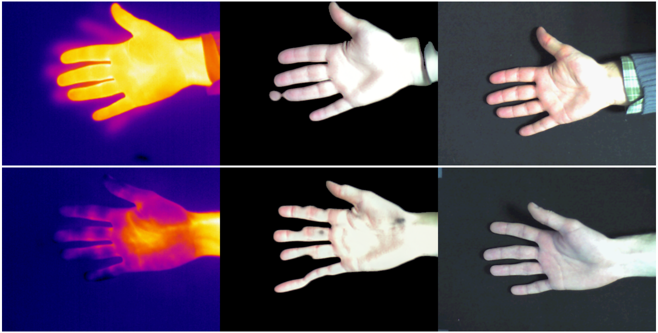
- Proposed model achieves slightly lower performance than the conventional Otsu and GMM methods for *'easy samples'* (hands that are easily discernible from the background)
  ▷ probable reasons: not enough samples, inaccurate *ground truth*, wrong structure of network model, overly aggressive masking

- This method can still be considered as a state-of-the-art solution for segmenting thermal hand images thanks to its good predictions given for difficult samples, such as those with parts of hands or fingers colder than the background, or images with various intrusions, such as wrist-watches or jewelry
- very good results for interbase tests (especially for training on *Warsaw* + CASIA and testing on *THID*: mean(IoU) = 94.30%)

Future work

- impementation of other CNN models for segmentation
- comparison of recognition using the segmentation method
- implementation CNN to verification scenario for mobile solutions (Triplet Network, Deep Siamese Networks)
- extracting cross-spectral features from visible-light images

Generating High-Quality Color Visible Images From Thermal Maps and Vice Versa Using Cascaded Refinement Network



Thermal Features for Presentation Attack Detection in Hand Biometrics

Ewelina Bartuzi,
Advisor: prof. Andrzej Pacut

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

Seminarium naukowe 3