REAL-TIME FACIAL LANDMARK LOCALIZATION USING ADVANCED CORRELATION FILTERS

Vitomir Štruc, PhD

Faculty of Electrical Engineering, University of Ljubljana, Slovenia

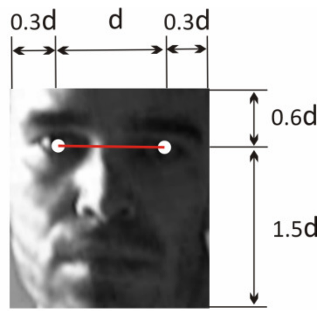
Friday, 21.9.2012, Zagreb, Croatia

OUTLINE

- Motivation
- Review of ASEF filters
- Generalization of ASEF filters
- PSEF filters for localization
- Experiments
- Demonstration
- Conclusion

MOTIVATION

- Face alignment techniques often rely on eye localization



- Accurate eye localization is crucial for recognition performance
- Many existing recognition techniques are highly susceptible to alignment (registration) errors

MOTIVATION

- Several ongoing project require an eye-localization procedure
- Common requirements:
  - Fast training phase,
  - Efficiency,
  - Robustness,
  - Computational simplicity,
  - Real-time capabilities.
- An appropriate solution: **correlation filters**
- Some options:
  - MACE (Mahalanobis et al., 1987)
  - OTF (Refregier, 1991)
  - UMACE (Savvides et al., 2003)
  - ASEF (Bolme et al., 2009)

MOTIVATION

- Several ongoing project require an eye-localization procedure
- Common requirements:
  - Fast training phase,
  - Efficiency,
  - Robustness,
  - Computational simplicity,
  - Real-time capabilities.
- An appropriate solution: **correlation filters**
- Some options:
  - MACE (Mahalanobis et al., 1987)
  - OTF (Refregier, 1991)
  - UMACE (Savvides et al., 2003)
  - **ASEF (Bolme et al., 2009)**

REVIEW OF ASEF FILTERS

- Average of Synthetic Exact Filters (ASEFs)[1]
- Given a training image $f(x,y)$ and a desired correlation output $g(x,y)$ of the following form:

$$g(x,y) = e^{-\frac{(x-x_i)^2+(y-y_i)^2}{\sigma^2}}$$

we define a SEF computed in the frequency domain as:

$$H^*(\omega,\nu) = \frac{G(\omega,\nu)}{F(\omega,\nu)} = \frac{G(\omega,\nu) \odot F^*(\omega,\nu)}{F(\omega,\nu) \odot F^*(\omega,\nu)}$$

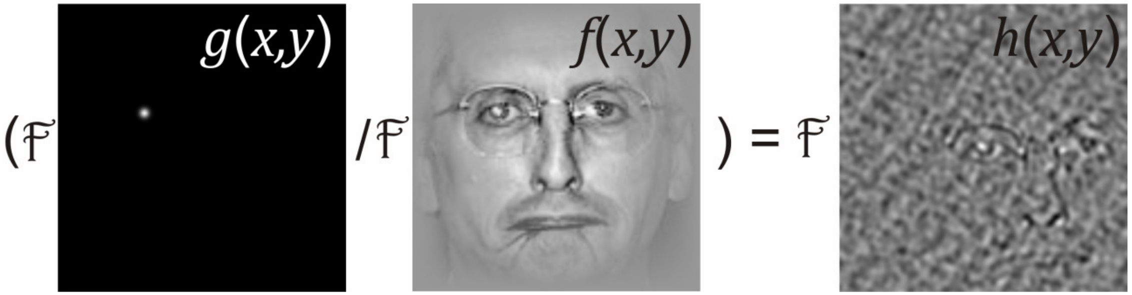
- Justification:

$$g(x,y) = (f \otimes h)(x,y) = \mathcal{F}^{-1}(F(\omega,\nu)H(\omega,\nu))$$

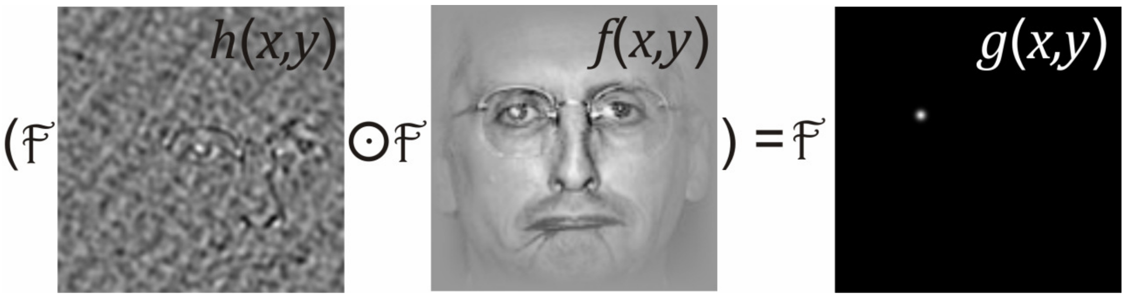
[1]Bolme, D.S, Draper, B.A., Beveridge, J.R.: Average of synthetic exact filters. In: Proc. Of CVPR'09, pp. 2105-2112 (2009).

REVIEW OF ASEF FILTERS

- A visual example
  - SEF construction:



  - SEF filtering:



REVIEW OF ASEF FILTERS

- ASEF construction
  - Compute SEFs for all *n* training images $f_i(x, y), i = 1,2,\ldots, n$
  - Average SEFs to improve generalization:

$$h(x,y) = \frac{1}{n}\sum_{i=1}^{n} h_i(x,y), \qquad h_i(x,y) = \mathcal{F}^{-1}(H_i(\omega, \nu))$$

  - A visual example

LUKS
Laboratory of artificial perception systems and cybernetics

GENERALIZATION OF ASEF FILTERS - PSEF

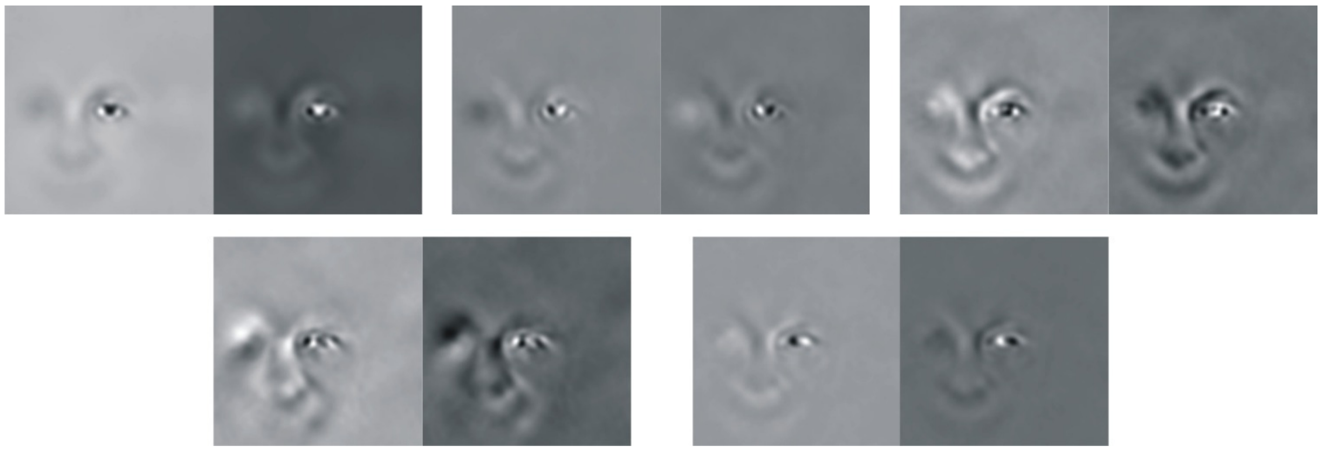
- ASEF filters rely only on the sample mean of the SEFs to characterize the SEFs distribution
- If we assume that the SEFs are drawn from a uni-modal multivariate Gaussian distribution, we can also compute the maximum variance (or principal) directions of the SEFs distribution using:

$$\boldsymbol{\Sigma}\boldsymbol{p}_i = \lambda_i \boldsymbol{p}_i,\ \ i = 1,2,\ldots, \min(d,n)$$

- We define the eigen-vectors $\boldsymbol{p}_i$ of the above equation as PSEF filters
- The procedure resembles PCA (difference in $\boldsymbol{\Sigma}$)

GENERALIZATION OF ASEF FILTERS - PSEF

- Properties of PSEF filters
  - First PSEF filter equals ASEF filter
  - Remaining filters carry additional information about the distribution of the SEFs
  - Filters exhibit sign ambiguity that needs to be resolved
- Visual examples of first five filters - multiplied by 1 and -1



PSEF FILTERS FOR LOCALIZATION

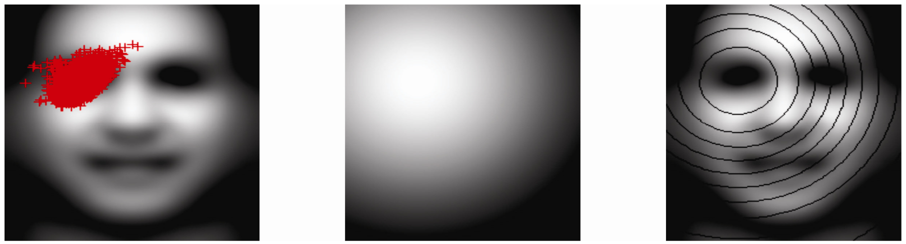
- Using PSEF filters for localization requires a fusion scheme (we have several filters at our disposal)
- Viable option: evaluate weighted sum of correlation outputs
- Fast option: exploit linearity

$$g_c(x,y) = \sum_{i=1}^{k} w_i g_i(x,y) = \sum_{i=1}^{k} w_i\big((p_i \otimes f)(x,y)\big) =$$

$$= \left(\sum_{i=1}^{k} w_i\, p_i(x,y)\right) \otimes f(x,y) = p_c(x,y) \otimes f(x,y)$$

INCORPORATING CONSTRAINTS

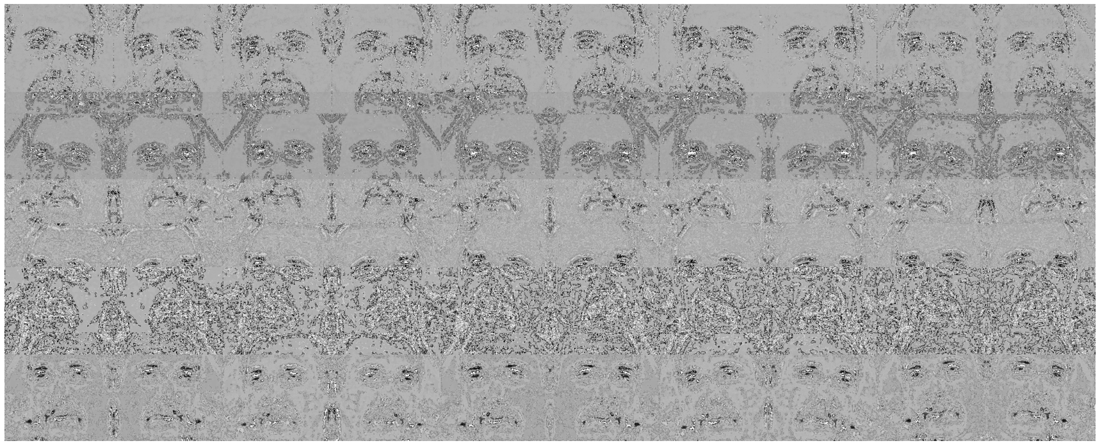
- To boost the localization performance we incorporate two types of constraints into our procedure
- Soft constraint
  - Gaussian shaped weighting function that is multiplied with the correlation output
  - Acts as sort of prior model (estimated on some training data)



- Hard constraint
  - Look for right eye in the upper right corner and for left eye in the upper left corner

EXPERIMENTS

- Training database:
  - Part of the LFW database (640 images)
  - Augmented through affine transformations to 25600 images
    - up to ±5 pixels shift in each direction, rotation up to ±15°, scaling by a factor of at most 1.0 ±0.15, mirroring around the y axes
  - Images size equals 128x128 pixels
  - All facial regions are extracted using the VJ face detector



- 4o modifications per image (robustness, data requirements)
- Retinex preprocessed for illumination invariance

EXPERIMENTS

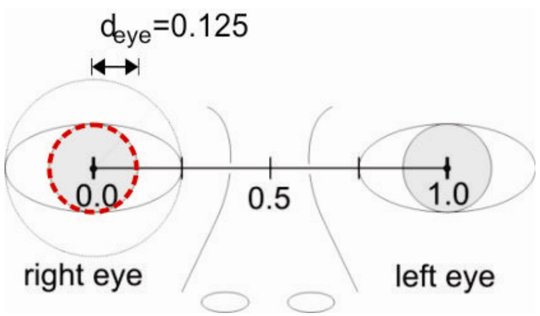
- Test database:
  - 3815 images from the FERET database
  - Augmented through affine transformations to 45780 images
    - up to ±5 pixels shift in each direction, rotation up to ±15°, scaling by a factor of at most 1.0 ±0.15, mirroring around the y axes
    - 12 modifications per image

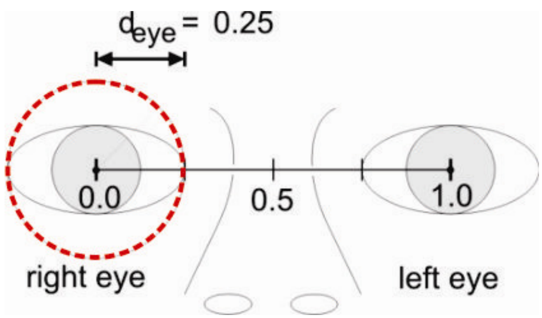
EXPERIMENTS

- Localization criterion (Interoccular distance criterion)

$$d_{eye} = \frac{\max(||\, l_{le} - r_{le} ||, || l_{re} - r_{re} ||)}{|| r_{le} - r_{re} ||}$$

- Examples



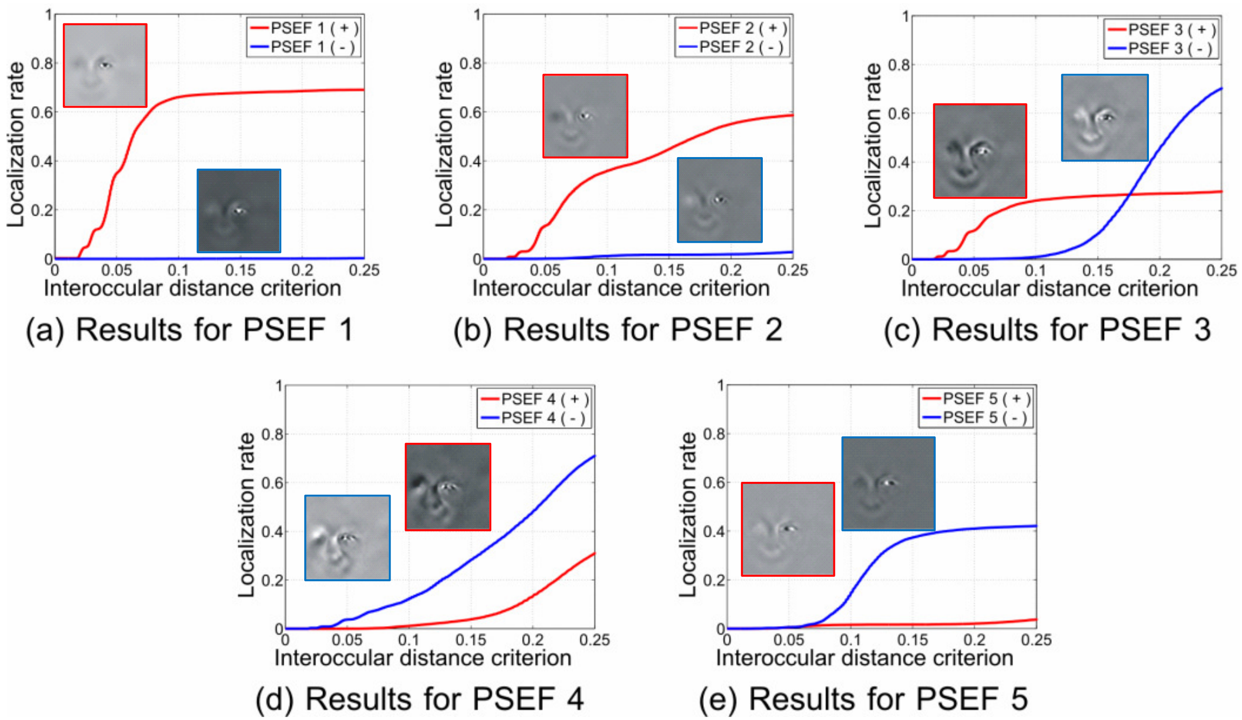


- We plot the plot the proportion of images, on which we achieved d_{eye}, against the value of d_{eye}

LUKS
Laboratory of artificial perception systems and cybernetics

EXPERIMENTS

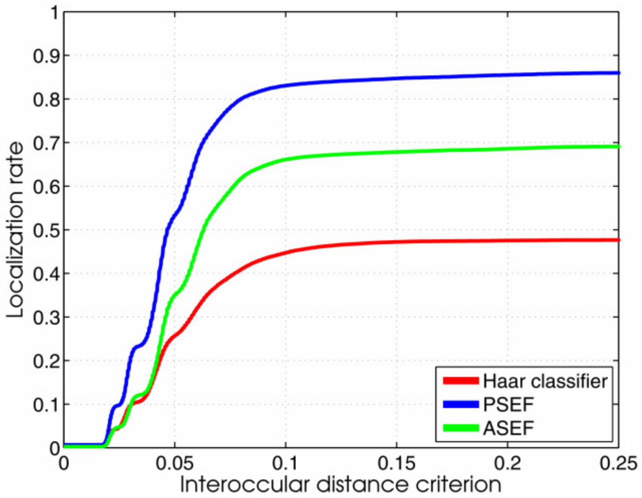
- Sign ambiguity
  - Results (we choose d_{eye}=0.25 for alleviating the sign ambiguity)

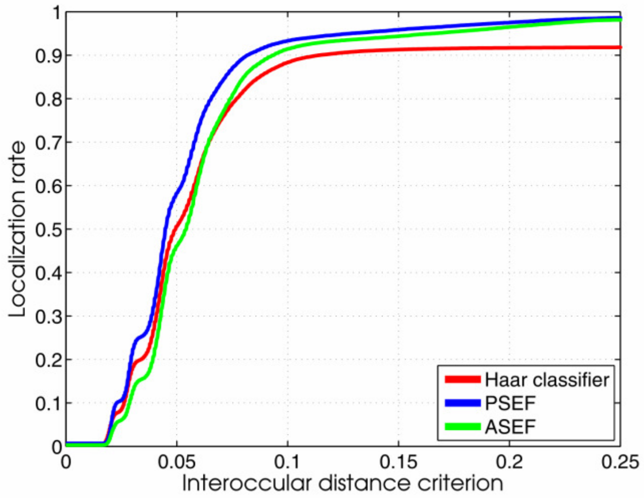


(a) Results for PSEF 1
(b) Results for PSEF 2
(c) Results for PSEF 3
(d) Results for PSEF 4
(e) Results for PSEF 5

EXPERIMENTS

- Comparison of PSEF, ASEF and Haar





Criterion	Unconstrained search space						Constrained search space					
	Left eye			Both eyes			Left eye			Both eyes		
	Haar	ASEF	PSEF	Haar	ASEF	PSEF	Haar	ASEF	PSEF	Haar	ASEF	PSEF
0.05	50.5	56.9	70.5	25.6	35.0	53.0	67.5	65.6	74.3	50.6	46.1	58.2
0.10	69.8	79.2	89.5	44.7	66.1	83.0	92.4	94.6	95.9	88.3	91.4	93.3
0.15	71.1	80.5	90.7	47.2	67.8	84.7	94.6	96.5	97.6	91.3	94.4	95.8
0.20	72.5	81.2	91.2	47.5	68.6	85.5	95.0	97.8	98.5	91.7	96.5	97.5
0.25	72.7	81.5	91.5	47.7	69.1	86.0	95.0	98.7	99.1	91.8	98.1	98.6

LUKS
Laboratory of artificial perception systems and cybernetics

EXPERIMENTS

- Time needed (Intel i7-2600 CPU @ 3.40 GHz, OpenCV, FFTW)

Face part	Without hard constraint		With hard constraint	
	Haar	ASEF, PSEF	Haar	ASEF, PSEF
Left eye	21.6 ms	0.65 ms	11.5 ms	0.66 ms
Right eye	24.8 ms	0.35 ms	13.6 ms	0.35 ms
Both eyes	46.4 ms	1.00 ms	25.1 ms	1.01 ms

- Training time
  - Haar – days (often weeks)
  - PSEF (hours)
  - ASEF (minutes)

DEMONSTRATION

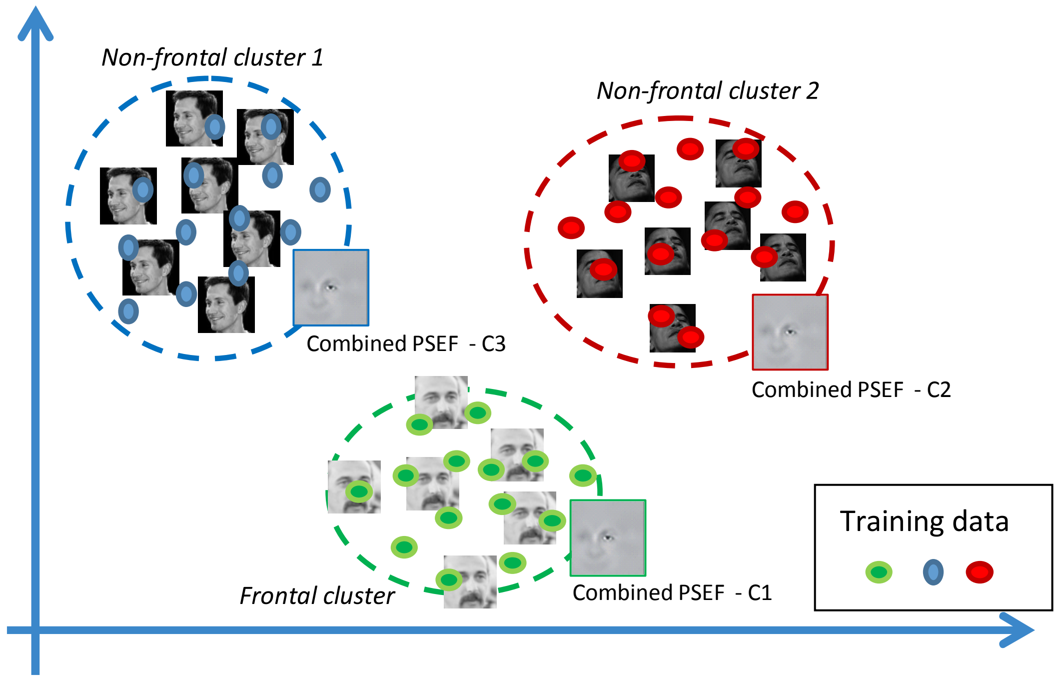
CONCLUSION

- We have presented a new class of correlation filters and applied them for the task of eye localization
- The filters exhibit some desirable properties
  - *Fast training* (hours) – ASEFs (minutes), Haar (days), which is important for optimization purposes
  - *Robust performance* – unlike window-based classifiers, correlation filters consider the entire image (holistic appraoch) – partial robustness to occlusions
  - *Rapid localization* – requires only one forward and one inverse Fourier transform, one element-wise multiplication, and maximum detection

LUKS
Laboratory of artificial perception systems and cybernetics

CONCLUSION

- Directions for future work
  - **A**daptive **m**ixtures of **PS**EFs - AMPS (for multi-modal SEF distributions, e.g., pose-specific correlation filters)

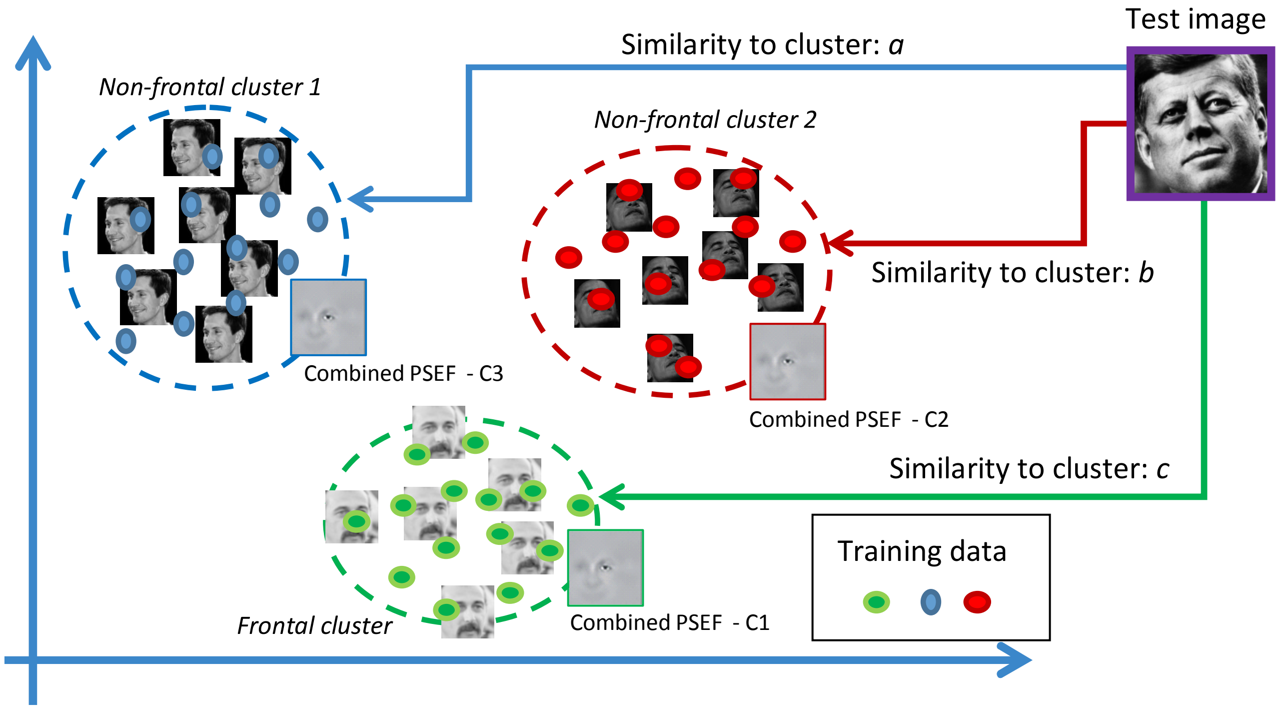


-

LUKS
Laboratory of artificial perception systems and cybernetics

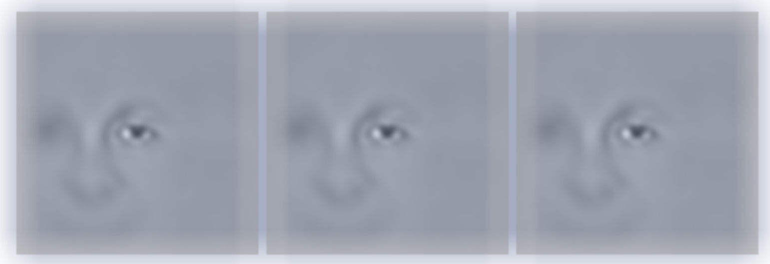
CONCLUSION

- Directions for future work
  - **A**daptive **m**ixtures of **PS**EFs - AMPS (for multi-modal SEF distributions, e.g., pose-specific correlation filters)



  - AMPS = *a*C3+*b*C2+*c*C1

LUKS
Laboratory of artificial perception systems and cybernetics



Investing in your future
OPERATION PART FINANCED BY THE EUROPEAN UNION
European Social Fund

Investing in your future
OPERATION PART FINANCED BY THE EUROPEAN UNION
European Regional Development Fund

THANK YOU FOR YOUR ATTENTION!

Contact Information
vitomir.struc@fe.uni-lj.si

Acknowledgements

The presented work has been ysupported in parts by the national researcg program P2-0250 (C) Metrology and Biometric Systems, the postdoctoral project **BAMBI** (ARRS ID Z2-4214), by the European Union, European Regional Fund, in scope of the framework of the Operational Programme for Strengthening Regional Development Potentials for the Perios 2007-2013, contract No. 3211-10-000467 (**Kc Class**) and the **BioID** project financed by the European Union from the European Social Fund, contract No. PP11/2010-(1-2009).

alpineon

The BAMBI Project
Biometric face recognition in ambient intelligence environments

KC Class

ADDITIONAL SLIDES

- Comparison between ASEF and combined PSEF filter

LUKS
Laboratory of artificial perception systems and cybernetics

ADDITIONAL SLIDES

- Localization performance for one (left) eye