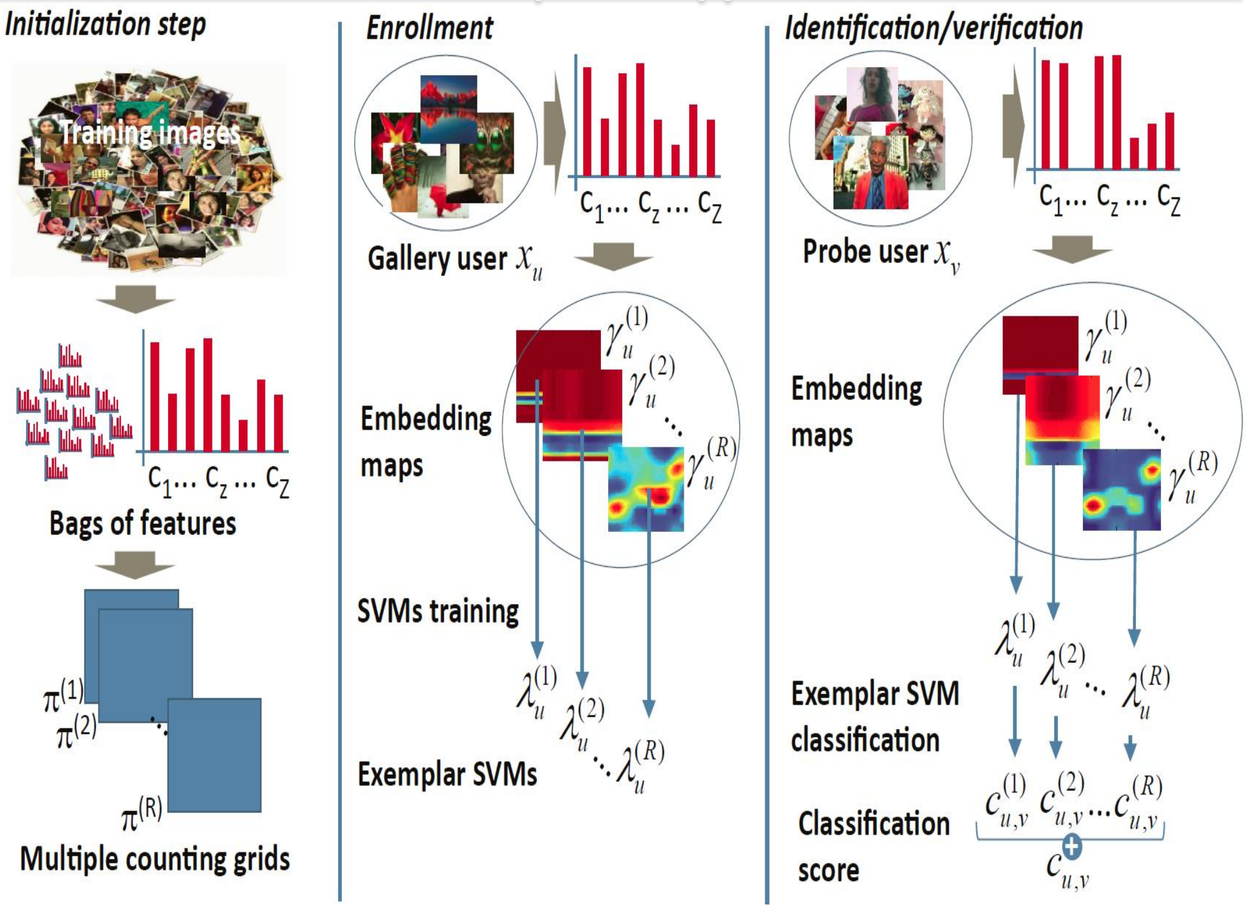
Recognizing People by Their Personal Aesthetics: A Statistical Multi-level Approach

Cristina Segalin[1], Alessandro Perina[2], Marco Cristani[1]
[1]Computer Science Department, University of Verona, Verona, Italy
[2]Istituto Italiano di Tecnologia (IIT), Genova, Italy

What is this work about?

We propose a soft biometric multi-level approach to recognize people by their personal aesthetics on a dataset of 200 users, 40K images. Given a set of preferred image of a user, it extracts a set of features which are discriminative for his/her.

The Proposed Approach



INITIALIZATION STAGE - ***Creating Bags of Features (BoF)***

Each image ***x*** is composed by the concatenation of features ranging from color statistics, image aesthetics cues, objects detections, each indicating the level of presence of a particular cue, i.e. an intensity ***count***, forming a BoF representation.

INITIALIZATION STAGE – ***Multi-view Counting Grid (CG) Training***:

- Dataset divided in ***gallery*** and ***probe*** set (100 preferred img./user both)
- Learn a ***multi-resolution CG*** [1], given BoF of the training images , extent E and window size S of CG. R=*E-S* CGs are learnt starting from r=*1 (S=E-1=44)* to r=R *(S=10),* using at each resolution level r the CG learnt at the previous step (except for the first CG initialized randomly).

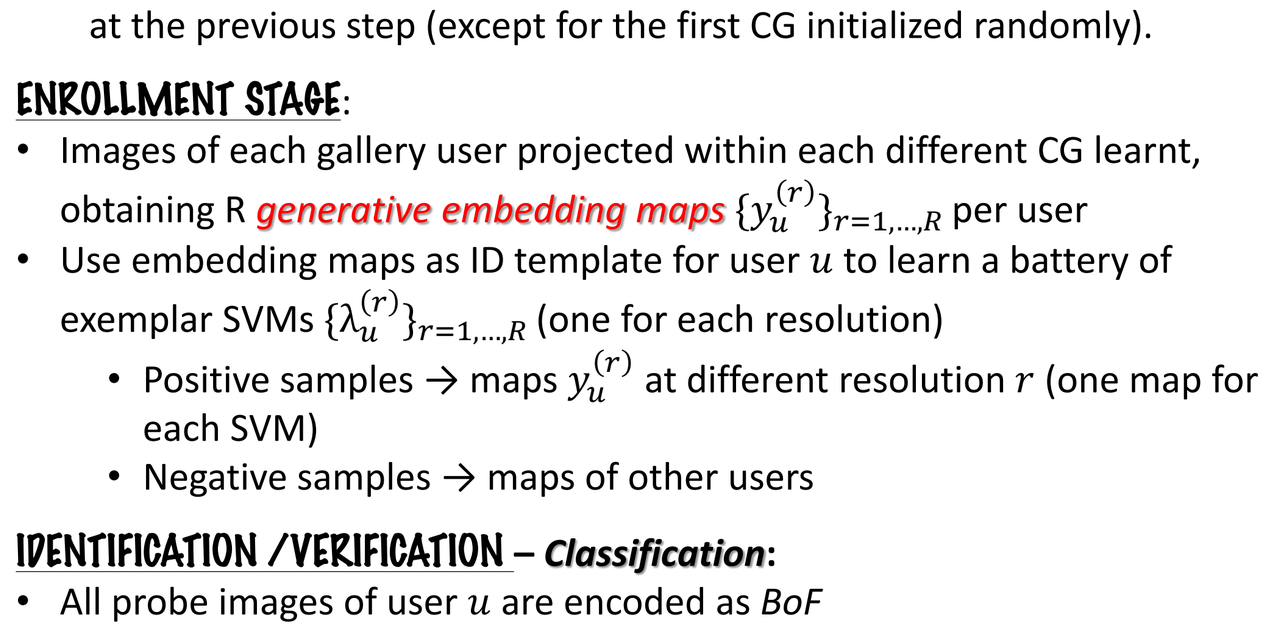
ENROLLMENT STAGE:

- Images of each gallery user projected within each different CG learnt, obtaining R ***generative embedding maps*** $\{y_u^{(r)}\}_{r=1,\dots,R}$ per user
- Use embedding maps as ID template for user u to learn a battery of exemplar SVMs $\{\lambda_u^{(r)}\}_{r=1,\dots,R}$ (one for each resolution)
  - Positive samples → maps $y_u^{(r)}$ at different resolution r (one map for each SVM)
  - Negative samples → maps of other users

IDENTIFICATION /VERIFICATION – ***Classification***:

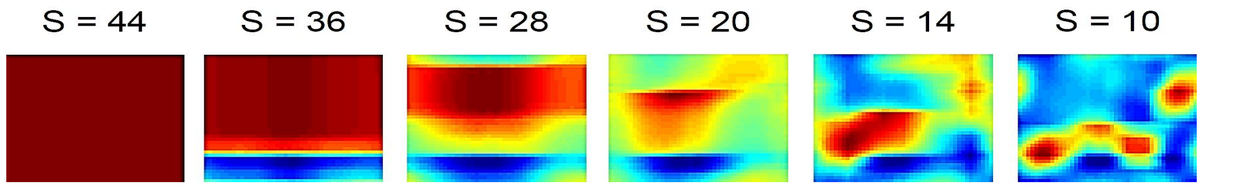
- All probe images of user u are encoded as *BoF*
- Project them on the multi-resolution CG
- Resulting probe embedding maps $\{y_u^{(r)}\}_{r=1,\dots,R}$ used as input of the SVMs related to gallery user u, producing R scores
- Average the R scores to provide a single classification score

Multi-view Counting Grid



(L) Collage of images of CG. (R) Embedding maps of a single $r = R$ with some images preferred by three users.

CG allows to project images in a latent space where similar pictures are mapped nearby, so that semantic areas can emerge. Each CG is characterized by a particular resolution: the higher the resolution, the stronger the visual similarity of close images.



Embedding map for user 38. From the lowest resolution (r=1, S=44) to the higher, identifying semantic areas. The main idea is to show all the mappings of the images belonging to a given user, thus highlighting the zones of the latent space where the images have been located.

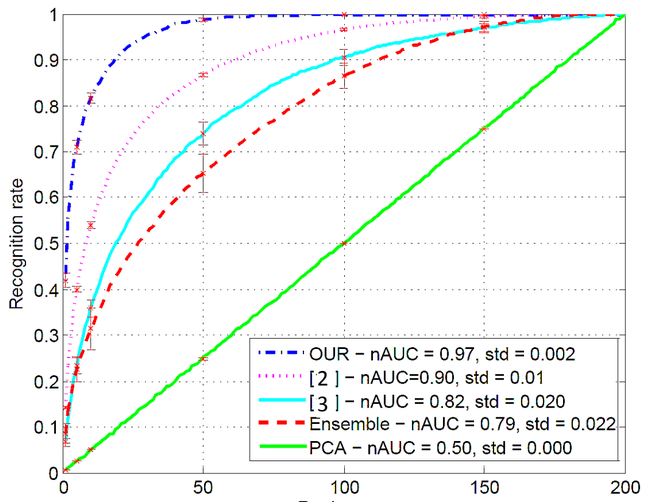
Identification Results and Feature Analysis

The identification task serves to guess the identity of a subject. Probe embedding maps are given as input to all the U gallery classifier, producing U identification scores. Fixing a gallery user, the average of the confidence scores produced by the exemplar SVMs (one for each resolution) is calculated. The gallery user with the highest averaged score corresponds to the probe user.

Build a ***CMC curve***: given a probe signature of a user and the matching confidence score, the curves tell the rate at which the correct user is found within the first k matches, with all possible k spanned on the x-axis.

category	rank 1	rank 5	rank 20	rank 50	nAUC
color	0.38±0.21	0.65±0.01	0.86±0.01	0.97±0.01	0.96± <0.01
composition	0.11±0.01	0.25±0.02	0.45±0.02	0.69±0.12	0.81±0.01
texture	0.10±0.01	0.21±0.01	0.39±0.03	0.64±0.02	0.79±0.01
content	0.10±0.01	0.20±0.01	0.38±0.03	0.61±0.03	0.78±0.01
all	0.36±0.02	0.64±0.02	0.86±0.01	0.97± <0.01	0.96± <0.01
color + composition + textures	0.37±0.02	0.64±0.02	0.86±0.01	0.98±0.01	0.96± <0.01
color + composition	**0.42±0.02**	**0.71±0.02**	**0.91±0.01**	**0.99±0.01**	**0.97± <0.01**

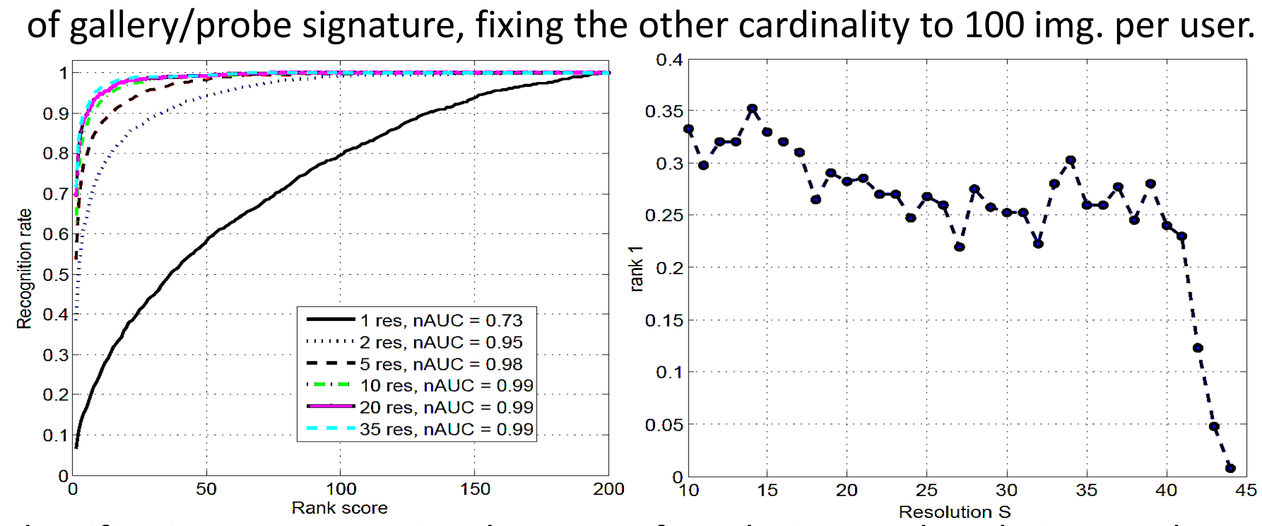
CMC scores, 100 img. for each gallery user and 5 img. for the probe user.



T_{te}	*rank 1*	*rank 5*	*rank 20*	*rank 50*	*nAUC*
1	0.19	0.42	0.66	0.86	0.90
5	0.42	0.71	0.71	0.99	0.97
20	0.63	0.87	0.7	1.00	0.99
100	0.73	0.92	0.98	1.00	0.99

T_{tr}	*rank 1*	*rank 5*	*rank 20*	*rank 50*	*nAUC*
5	0.29	0.59	0.83	0.94	0.95
10	0.46	0.80	0.95	0.99	0.98
20	0.63	0.89	0.98	1.00	0.99
50	0.71	0.93	0.99	1.00	0.99
100	0.71	0.92	0.98	1.00	0.99

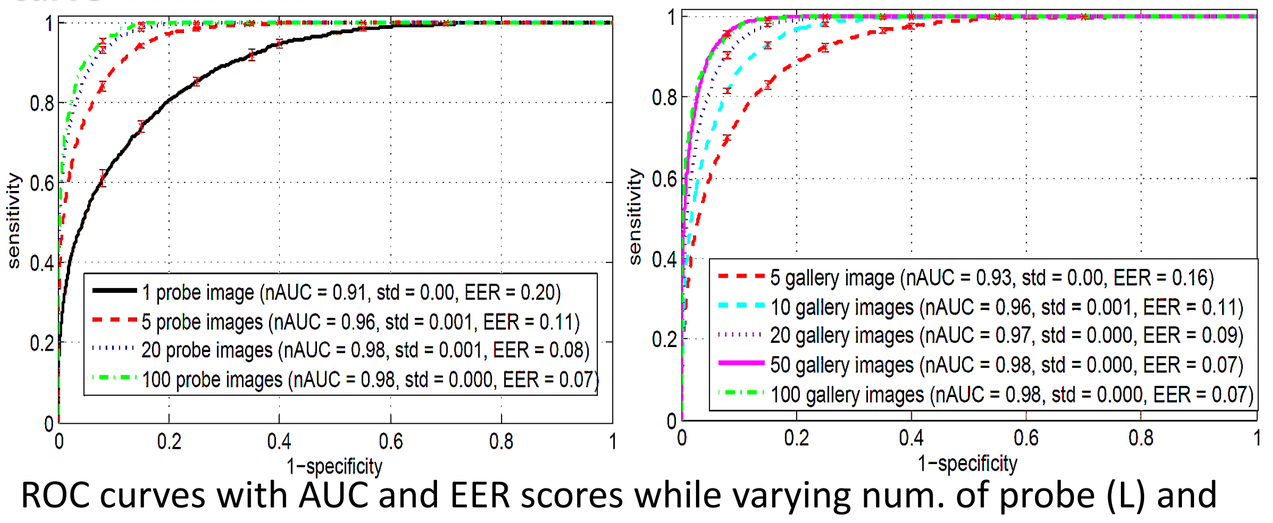
(L) Results with 5 img./probe set and 100 img/gallery set, comparing with other baselines. (R) Results of our approach varying the num. T_{te}/T_{tr} img. of gallery/probe signature, fixing the other cardinality to 100 img. per user.



Identification scores varying the num. of resolutions and analysis at rank 1.

Verification Results

Verify the identity of a user by his/her preferred images. For every user u, probe set of user u are taken as ***client*** images, all the other probe sets as ***impostor*** images. The maps are given to a single classifier, which accepts or rejects the signature considering a given ***"authentication threshold"***, i.e. a value over which the subject is authenticated. Sensitivity (TPR) and specificity (TNR) can now be computed. By varying this threshold, the ***ROC curve*** is obtained.



ROC curves with AUC and EER scores while varying num. of probe (L) and gallery (R) images.

References

[1]Perina, A., Jojic, N.: Image analysis by counting on a grid. IEEE CVPR 2011
[2]Lovato, P., et al: Faved! biometrics: Tell me which image you like and I'll tell you who you are. IEEE Trans. on Information Forensics and Security 2014
[3] Segalin, C., Perina, A., Cristani, M.: Biometrics on visual preferences: A "Pump and Distill" regression approach ICIP 2014

 cristina.segalin@univr.it