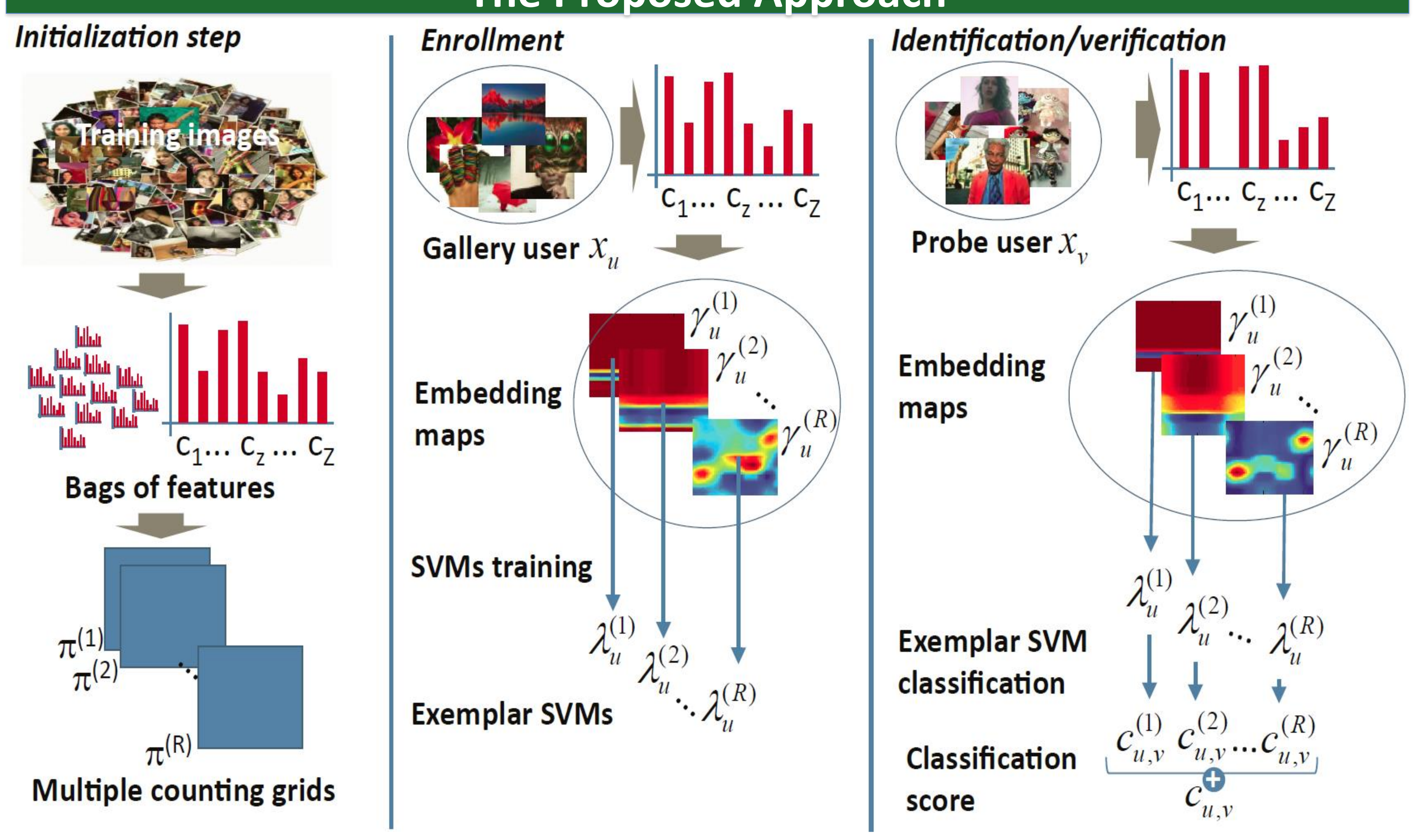
# Recognizing People by Their Personal Aesthetics: A Statistical Multi-level Approach

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).

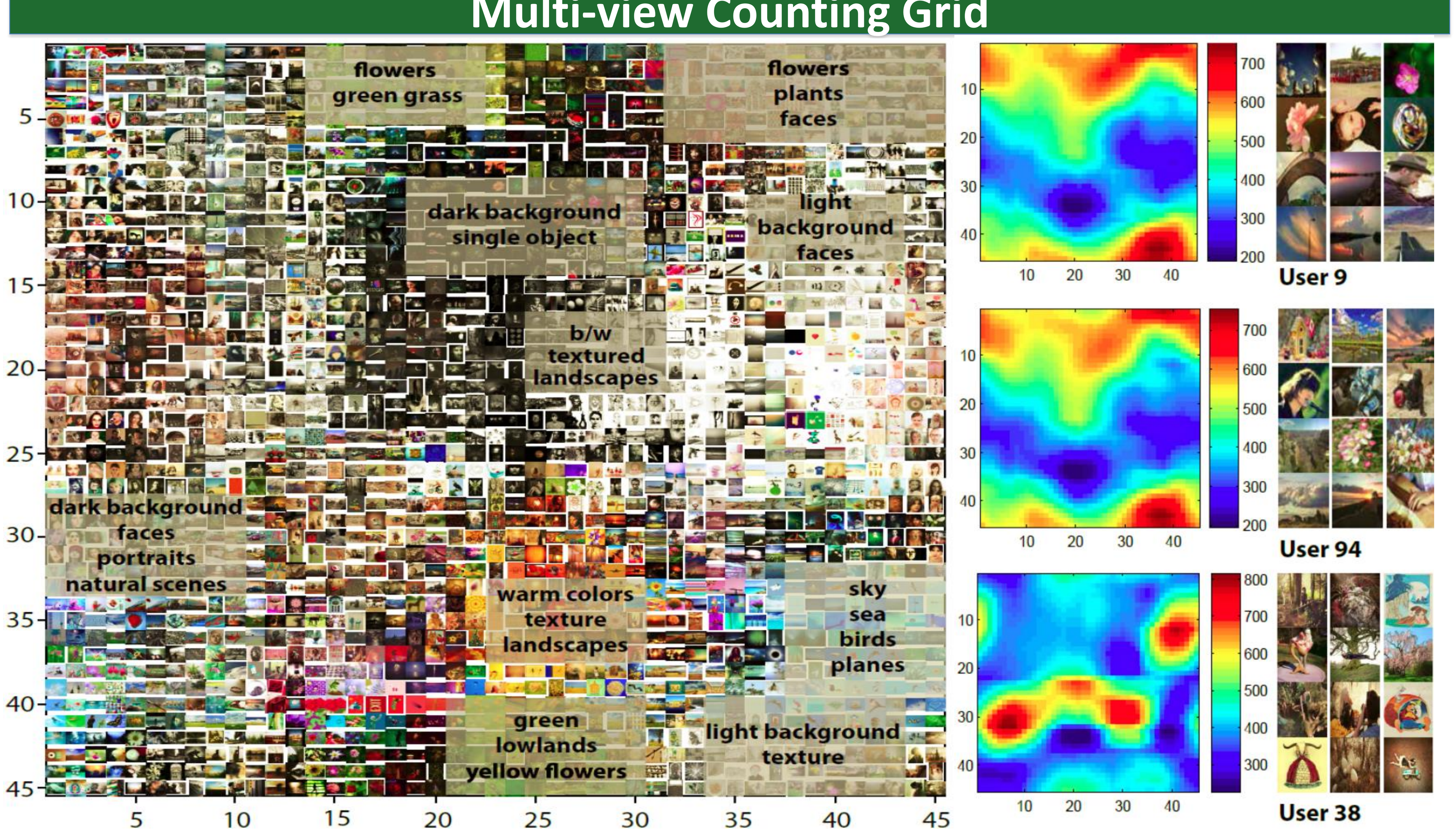
**ENROLLMENT STAGE**:

- Images of each gallery user projected within each different CG learnt, obtaining R *generative embedding maps* $\{y_u^{(r)}\}_{r=1,\ldots,R}$ per user
- Use embedding maps as ID template for user $u$ to learn a battery of exemplar SVMs $\{\lambda_u^{(r)}\}_{r=1,\ldots,R}$ (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

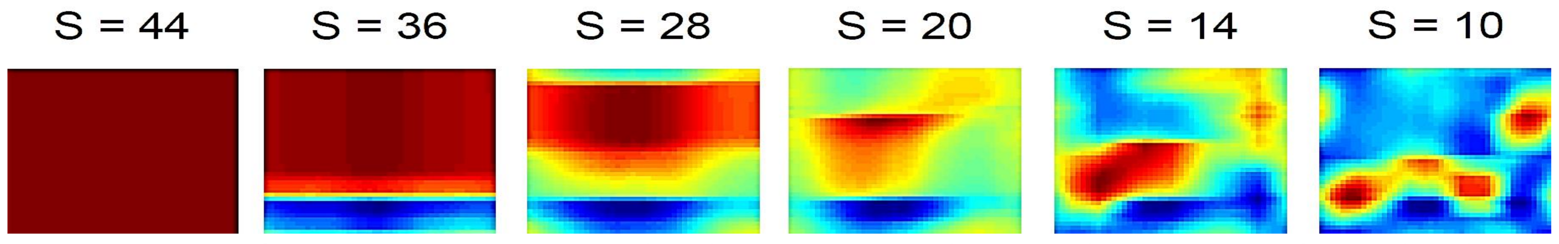
**IDENTIFICATION /VERIFICATION** – ***Classification***:

- All probe images of user $u$ are encoded as *BoF*
- Project them on the multi-resolution CG
- Resulting probe embedding maps $\{y_u^{(r)}\}_{r=1,\ldots,R}$ 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.

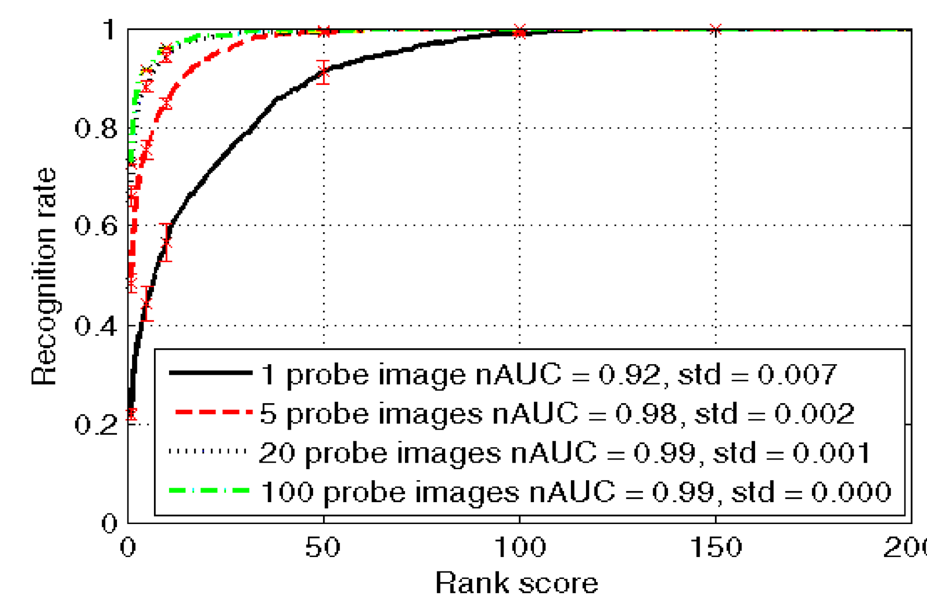


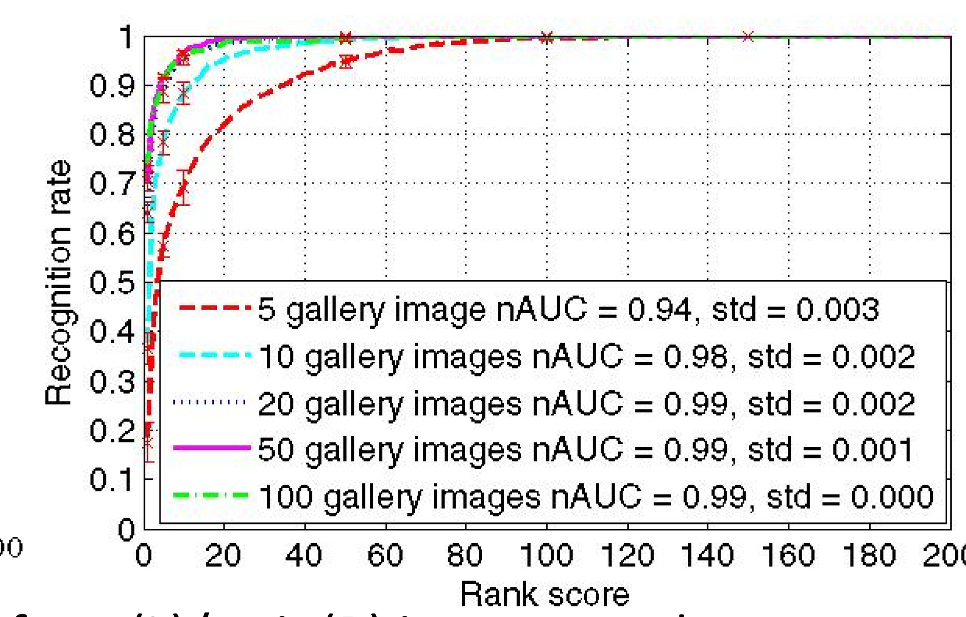

Embedding map for user 38. From the lowest resolution ($r$=1, S=44) to the higher, identifying semantic areas.

## 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.



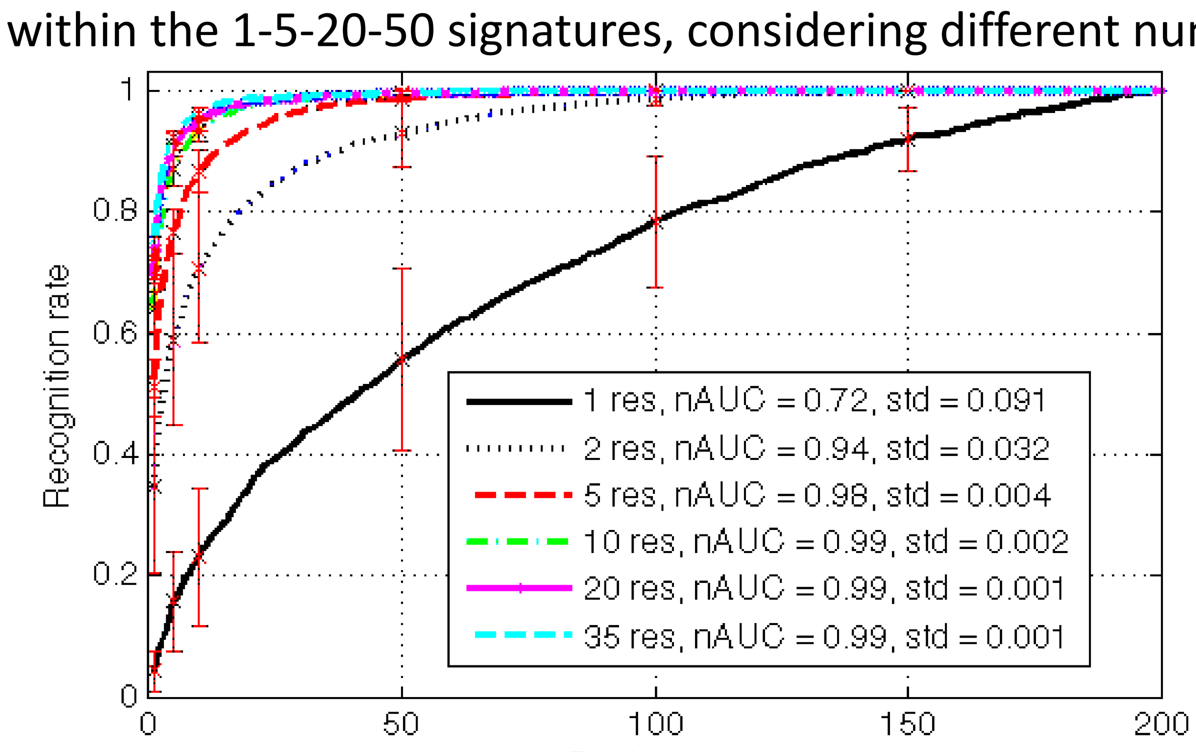




CMC curves while varying the num. of test(L)/train(R) images used to compose the probe(L)/gallery(R) signature. The normalized area under curve (nAUC) is also reported.

| $T_{te}$ | Met. | rank 1 | rank 5 | rank 20 | rank 50 | nAUC |
|---|---|---|---|---|---|---|
| 1 | [2] | 0.06 | 0.18 | 0.40 | 0.82 | 0.76 |
| | *our* | **0.22±0.01** | **0.44±0.04** | **0.70±0.04** | **0.91±0.02** | **0.92±0.007** |
| 5 | [2] | 0.14 | 0.39 | 0.68 | 0.96 | 0.89 |
| | *our* | **0.48±0.02** | **0.75±0.02** | **0.94±0.01** | **0.99±<0.01** | **0.98±0.002** |
| 20 | [2] | 0.25 | 0.62 | 0.88 | 0.99 | 0.96 |
| | *our* | **0.66±0.02** | **0.88±0.01** | **0.98±<0.01** | **1.00±<0.01** | **0.99±0.001** |
| 100 | [9] | 0.35 | 0.79 | 0.97 | 0.99 | 0.98 |
| | *our* | **0.73±<0.01** | **0.92±<0.01** | **0.98±<0.01** | **1.00±<0.01** | **0.99±0.000** |

| $T_{tr}$ | Met. | rank 1 | rank 5 | rank 20 | rank 50 | nAUC |
|---|---|---|---|---|---|---|
| 5 | *our* | **0.17±0.04** | **0.57±0.02** | **0.82±0.02** | **0.95±0.01** | **0.94±0.003** |
| | [2] | 0.07 | 0.23 | 0.49 | 0.88 | 0.81 |
| 10 | *our* | **0.37±0.03** | **0.78±0.02** | **0.95±0.01** | **0.99±<0.01** | **0.98±0.002** |
| | [2] | 0.11 | 0.32 | 0.62 | 0.94 | 0.87 |
| 20 | *our* | **0.64±0.02** | **0.89±0.02** | **0.98±0.01** | **1.00±<0.01** | **0.99±0.002** |
| | [2] | 0.15 | 0.44 | 0.74 | 0.97 | 0.91 |
| 50 | *our* | **0.70±0.02** | **0.99±<0.01** | **1.00±<0.01** | **0.99±0.001** | **0.92±<0.01** |
| | [2] | 0.22 | 0.57 | 0.83 | 0.99 | 0.94 |
| 100 | *our* | **0.73±<0.01** | **0.92±<0.01** | **0.98±<0.01** | **1.00±<0.01** | **0.99±<0.01** |
| | [2] | 0.35 | 0.79 | 0.97 | 0.99 | 0.98 |

Recognition results varying the num. $T_{te}$(top) / $T_{tr}$(bottom) of images that compose the probe (top) / gallery (bottom) signature and fixing the num. of gallery images $T_{tr}$ (top) / probe image $T_{te}$(top) to 100 per each user. The rank numbers are the $x$-axis value of the CMC curve: they represent the average probability of having the correct match within the 1-5-20-50 signatures, considering different number of probe images.

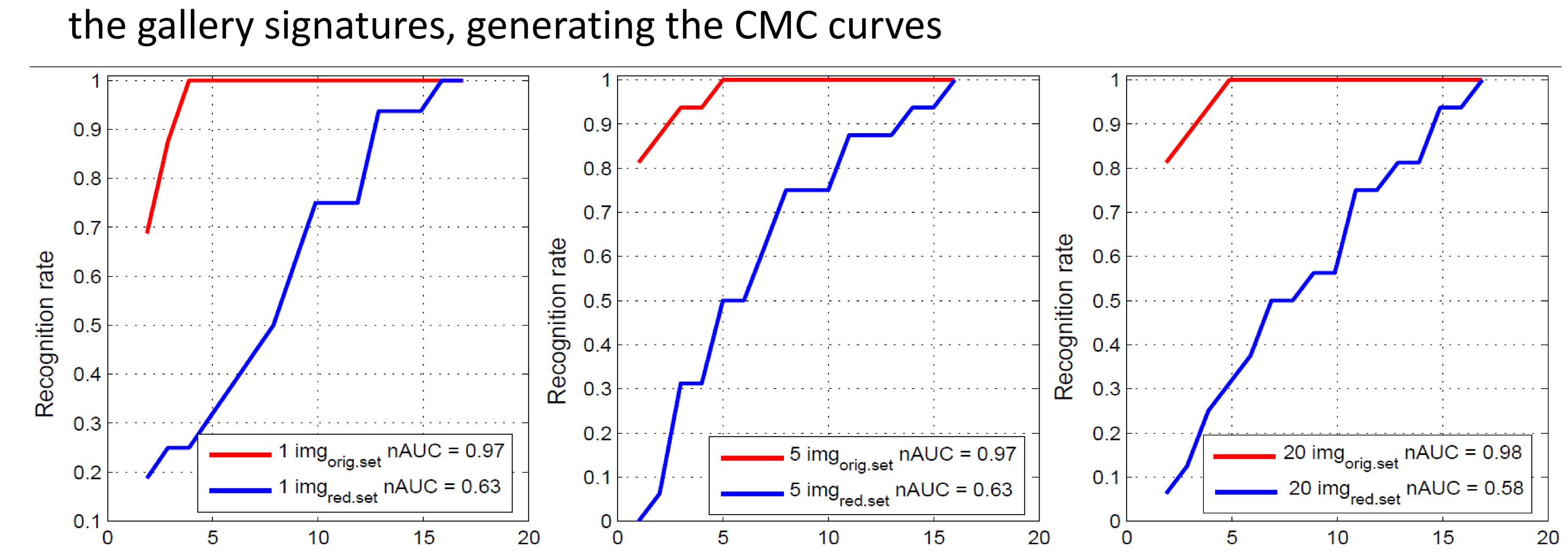



Identification scores while varying the number of resolutions employed

## Limitation of our approach

***Experiment***:

- Build a reduced dataset, by sampling one image from each pool of 200 liked images of all 200 users
- Organize the 200 images on a web interface
- Ask to 16 users of the dataset to select from this interface 5,10,20 images
- Use the selected images as test signature for our approach, and compare them with the gallery signatures, generating the CMC curves




## 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

 cristina.segalin@univr.it