kNN Based Image Classification Relying on Local Feature Similarity

Giuseppe Amato

Fabrizio Falchi

ISTI-CNR, Pisa

fabrizio.falchi@isti.cnr.it

Outline

1. **VISITO Tuscany Project**
   - exp. landmarks recognition

2. **Single-label distance weighted k-NN classifier**
   - based on image-to-image distance function
   - used as a baseline

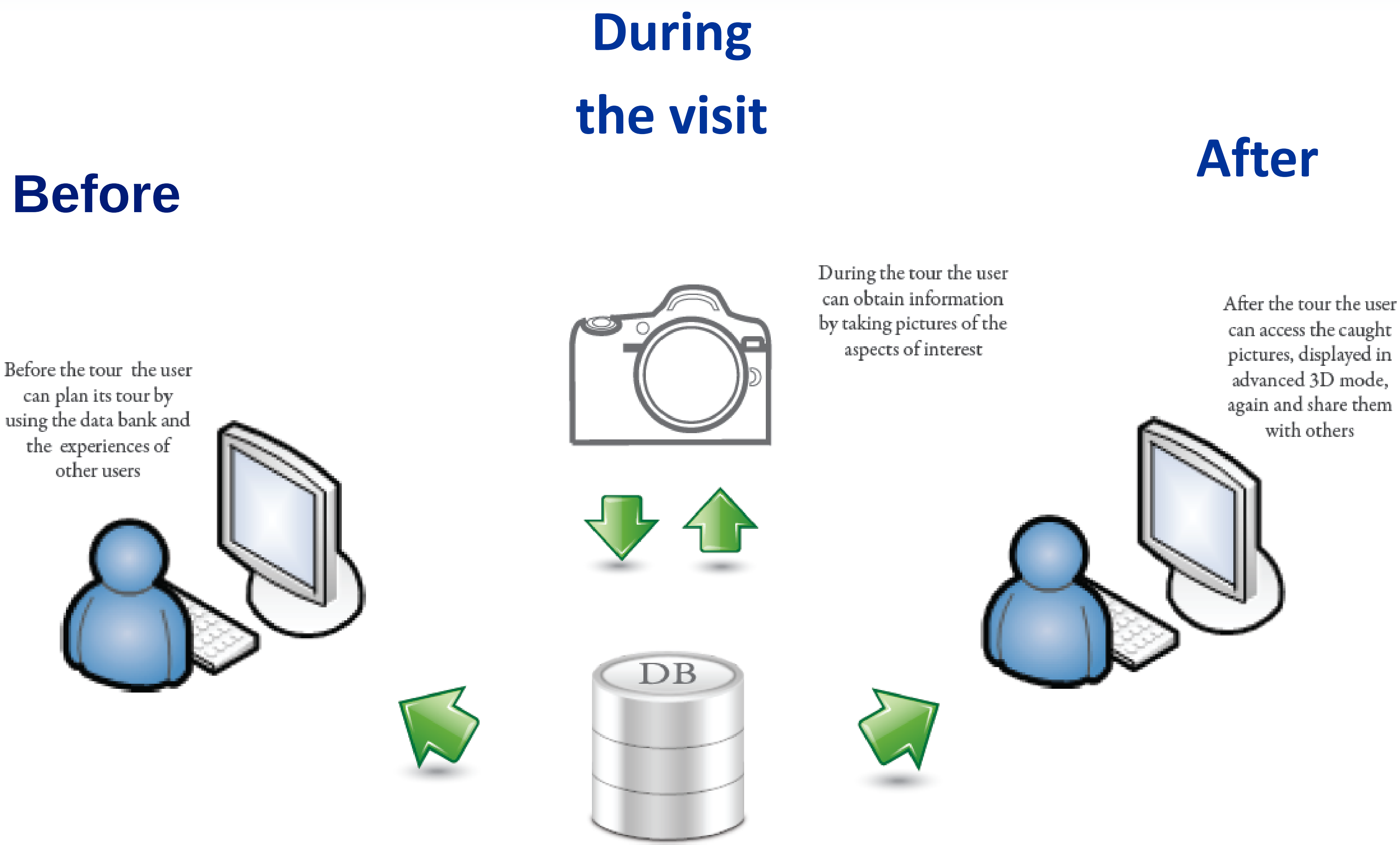
3. **4 novel classifiers based on NN search local features**
   - two steps classification process:
     1. local feature classification
     2. image classification

4. **Experimental results**

The VISITO Tuscany Project

- Started: October 2009 (**24 months**)
  - **helping tourists** visiting art cities
  - a geo-referenced **interactive guide** on sparthphones and Internet
  - **image-based** interaction trough images (pictures as queries)
  - management of **very large image archives**
  - offering functionalities to **professional users** too

The basic idea (1)



The basic idea (2)

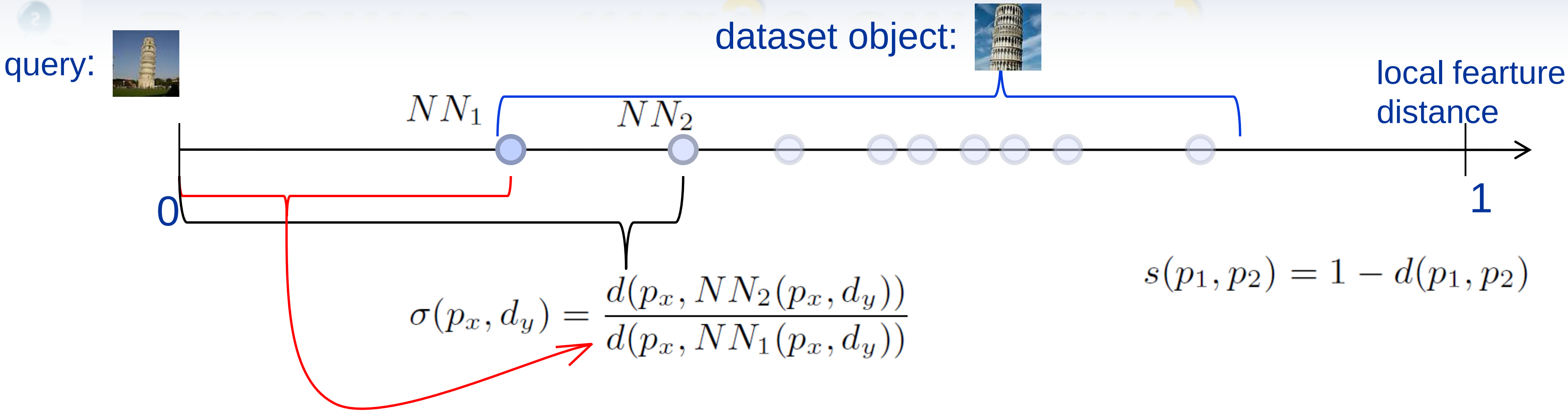
During the visit

- **Automatic localization** of the user and user interest
  - analyzing pictures taken by the user
- **Personalized** touristic information is then sent to the user
  - Web pages, phone calls, …

Pre-visit and post-visit:

- Virtual visit to interesting places using images and 3D models
- Possibility of obtaining information to plan the real visit

Baseline – Image Similarity



- Image Similarity based on Local Features
  - local features matching [Lowe 2001]

$$m(p_x, d_y) = \begin{cases} 1 & \text{if } \sigma(p_x, d_y) < c \\ 0 & \text{otherwise} \end{cases}$$

  - image similarity measure

$$s^m(d_x, d_y) = \frac{1}{|d_x|} \sum_{p_x \in d_x} m(p_x, d_y)$$

Distance Ratio Threshold

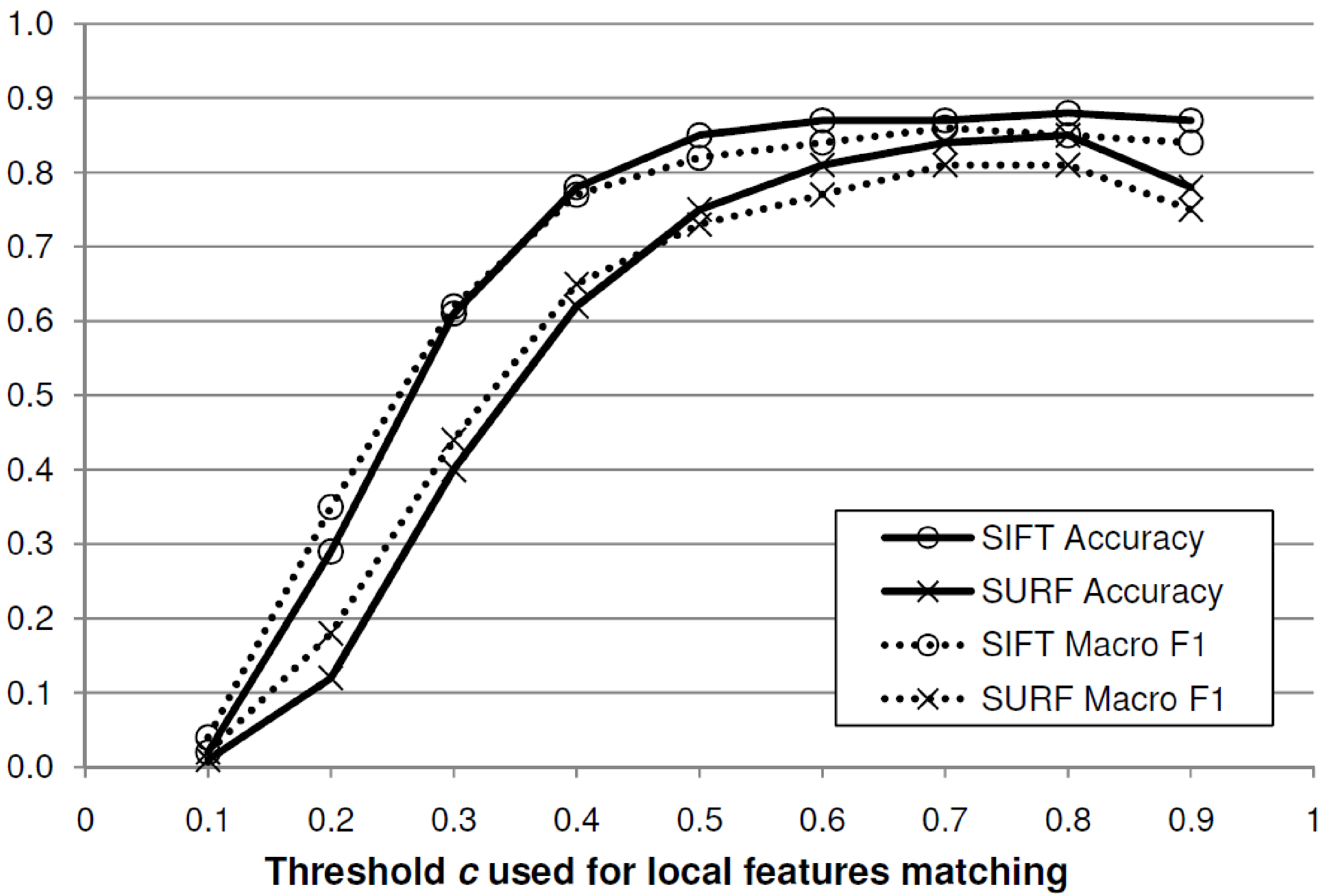


Image similarity based classifier

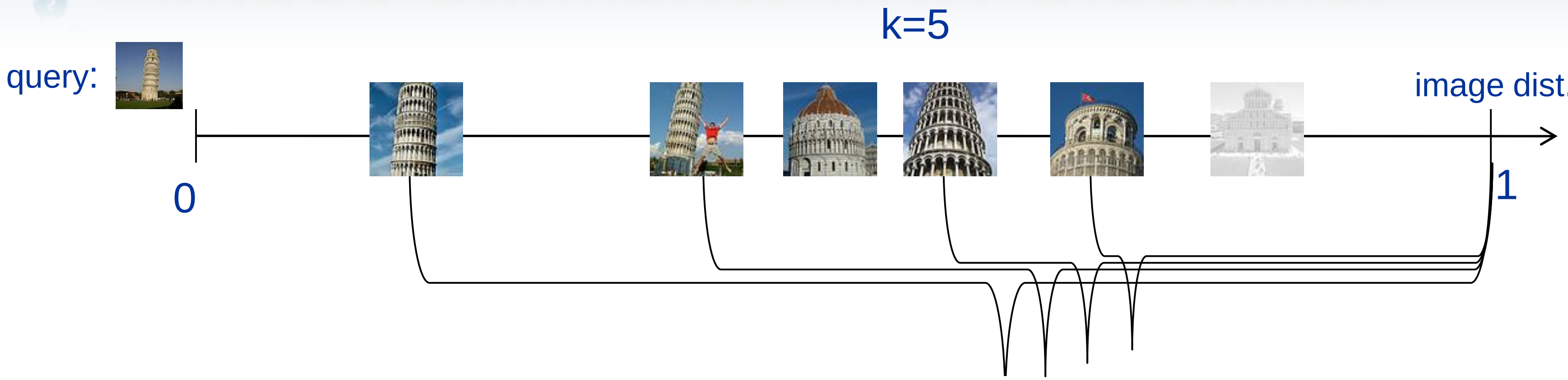
- ***single-label distance weighted k-NN:***
  - a **set of classes** $C=\{c_1,\ldots,c_m\}$,
    each associated with a training set of images
  - a similarity function $s(d_x,d_y)$
    is defined for any two images d_x and d_i

- Given a test image d_x, a label from C is associated:
  1. k-NN seach in the training set using d_x as query
  2. the label is chosen by maximizing the sum of the similarity between d_x and the kNN search results:

$$z(d_x, c_j) = \sum_{d_i \in \chi^k(d_x)\ :\ \Phi(d_i)=c_j} s(d_x, d_i)$$

$$\hat{\Phi}^s(d_x) = \arg\max_{c_j \in C} z(d_x, c_j)$$

Image Similarity kNN Classifier



$$z^k(p_x, c_j) = \sum_{p_i \in \chi^k(p_x) \;:\; \Phi(p_i)=c_j} s(p_x, p_i)$$

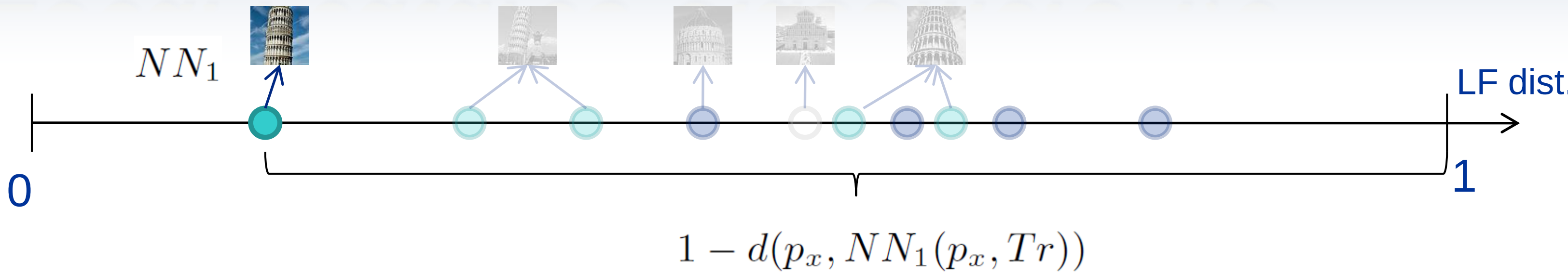
- predicted class label:

$$\hat{\Phi}^s(d_x) = \arg\max_{c_j \in C} z(d_x, c_j)$$

- confidence:

$$\nu_{doc}(\hat{\Phi}^s, d_x) = 1 - \frac{\arg \max_{c_j \in C - \hat{\Phi}^s(d_x)} z(d_x, c_j)}{\arg\max_{c_j \in C} z(d_x, c_j)}$$

Local Features Classifiers 1/3

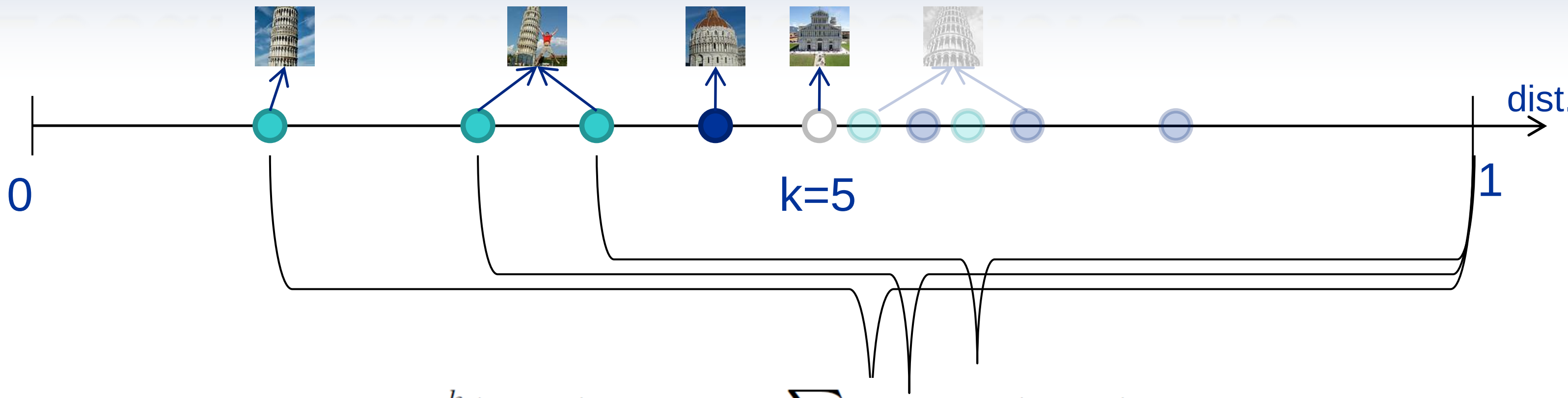


- We are now comparing the query local features with “all” the local features in the training set.
- LFs are labeled as the image in Tr they belong to
- LFs belonging to images with the same label have the same color

1-NN LF Classifier

$$\begin{cases} \hat{\Phi}^f(p_x) = \Phi(NN_1(p_x, Tr)). \\ \nu(\hat{\Phi}^f, p_x) = 1 - d(p_x, NN_1(p_x, Tr)) \end{cases}$$

Local Features Classifiers 2/3

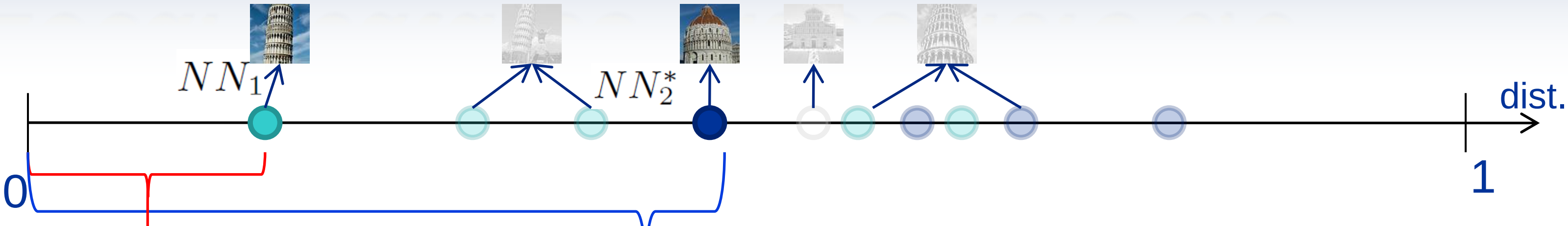


$$z^k(p_x, c_j) = \sum_{p_i \in \chi^k(p_x)\ :\ \Phi(p_i) = c_j} s(p_x, p_i)$$

Weighted kNN LF Classifier

$$\begin{cases} \hat{\Phi}^k(p_x) = \arg\max_{c_j \in C} z^k(p_x, c_j) \\ \nu(\hat{\Phi}^k, p_x) = 1 - \dfrac{\arg\max_{c_j \in C - \hat{\Phi}^k(p_x)} z^k(p_x, c_j)}{\arg\max_{c_i \in C} z^k(p_x, c_i)} \end{cases}$$

Local Features Classifiers 3/3



$$\dot{\sigma}(p_x, T_r) = \frac{d(p_x, NN_2^*(p_x, Tr))}{d(p_x, NN_1(p_x, Tr))}$$

LF Matching Classifier

$$\hat{\Phi}^m(p_x) = \Phi(NN_1(p_x, Tr))$$

$$\nu(\hat{\Phi}^m, p_x) = \begin{cases} 1 & \text{if } \dot{\sigma}(p_x, t_r) < c \\ 0 & \text{otherwise} \end{cases}$$

Weighted LF Distance Ratio Classifier

$$\begin{cases} \hat{\Phi}^w(p_x) = \Phi(NN_1(p_x, Tr)) \\ \nu(\hat{\Phi}^w, p_x) = (1 - \dot{\sigma}(p_x, t_r))^2 \end{cases}$$

Whole Image Classification

- Class Label evaluation

$$z(d_x, c_i) = \sum_{p_x \in d_x, \hat{\Phi}(p_x) = c_i} \nu(\hat{\Phi}, p_x)$$

- Predicted Class Label

$$\hat{\Phi}(d_x) = \arg\max_{c_j \in C} z(d_x, c_j)$$

- Whole Image Classifiaction Confidence

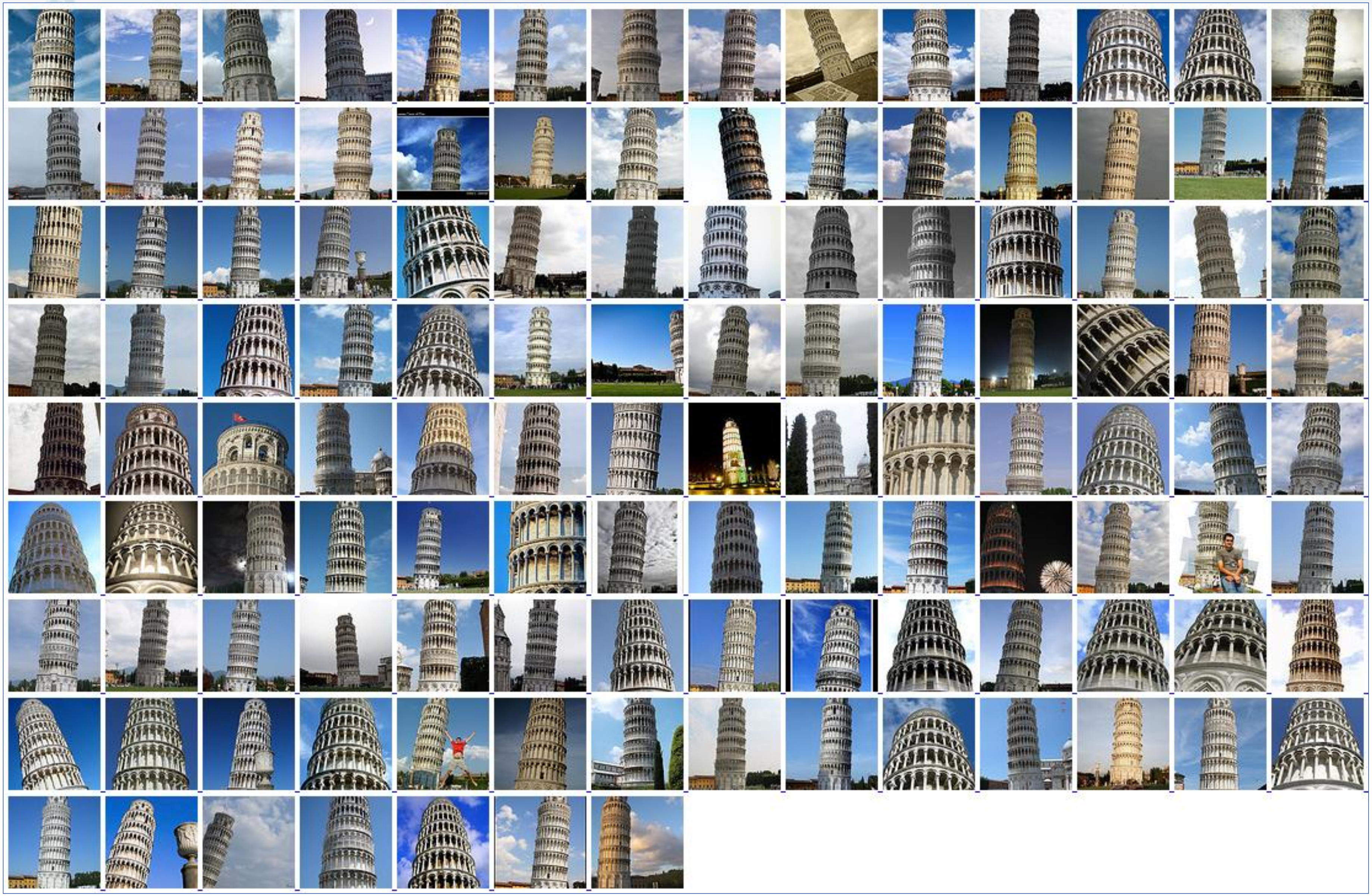
$$\nu_{img}(\hat{\Phi}, d_x) = 1 - \frac{\arg \max_{c_j \in C - \hat{\Phi}(p_x)} z(d_x, c_j)}{\arg\max_{c_i \in C} z(d_x, c_i)}$$

Classification Task

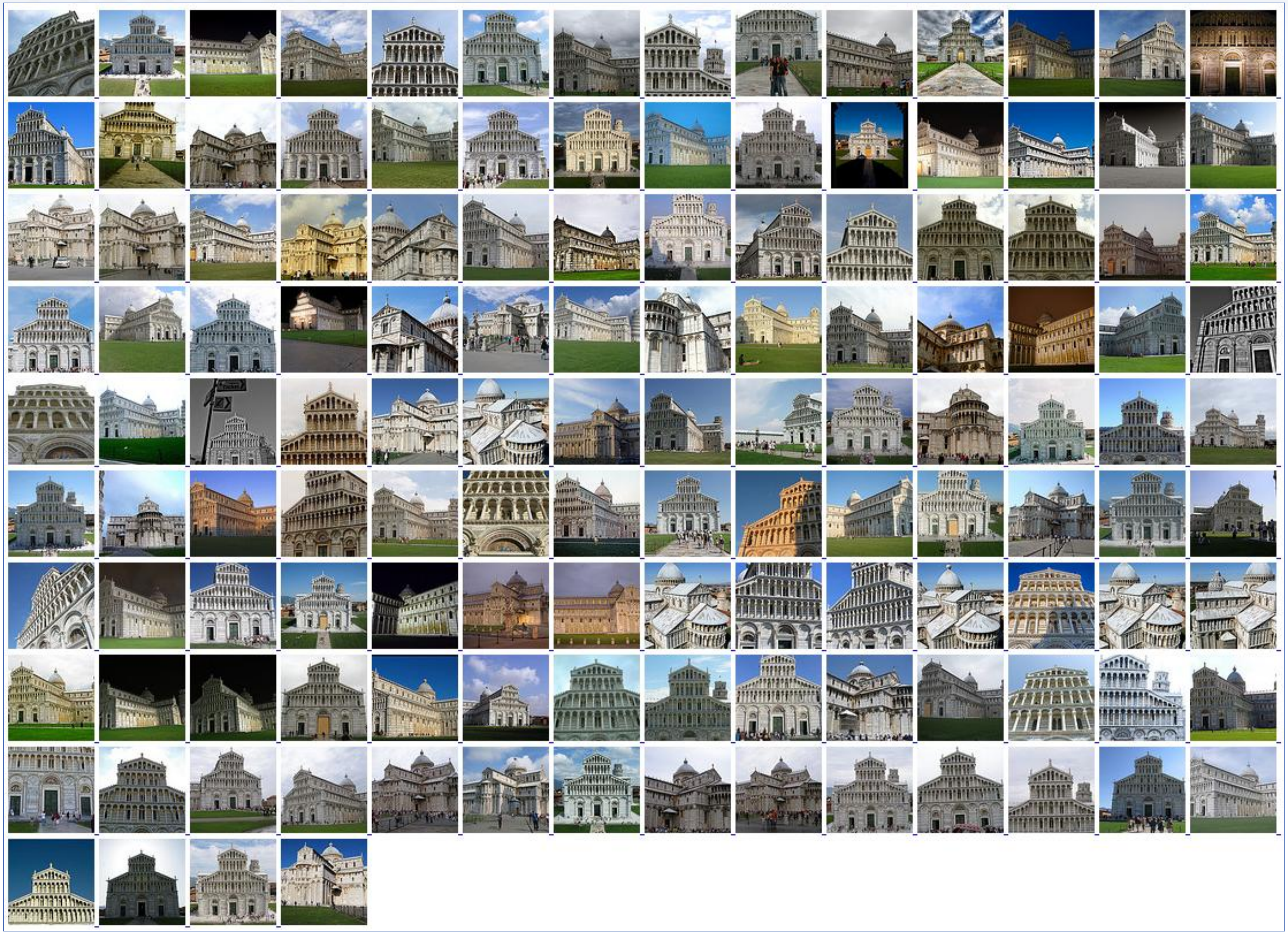
- We created a dataset composed of 1227 image of artifacts in Pisa
  - http://www.fabriziofalchi.it/pisaDataset/
- Contains images related to the following artifacts
  - Leaning Tower (119 pictures)
  - Duomo (130 pictures)
  - Baptistery (104 pictures)
  - Monumental Cemetery - Exterior (46 pictures)
  - Monumental Cemetery – Field (113 pictures)
  - Monumental Cemetery – Portico (138 pictures)
  - Chiesa della Spina (112 pictures)
  - Palazzo della carovana (101 pictures)
  - Palazzo dell’orologio (92 pictures)
  - Guelph Tower – Cittadella (71 pictures)
  - Basilica di San Piero (48 pictures)
  - Certosa di Calci (53 pictures)
- Task: associating images in the above classes

Classes:

Pisa-Dataset



Pisa-Dataset



Pisa-Dataset

Tests:

- Dataset partitioned in:
  - 20% **Training Set**
  - 80% **Test Set**

- Performance measures
  - **Recall**: TP / (TP+FN)
  - **Precision**: TP / (TP+FP)
  - **F_1**: harmonic mean between Recall and Precision
  - Micro-averaged **Accuracy**
    - equals micro-averaged Recall, Precision and F_1 in case of single-label classification

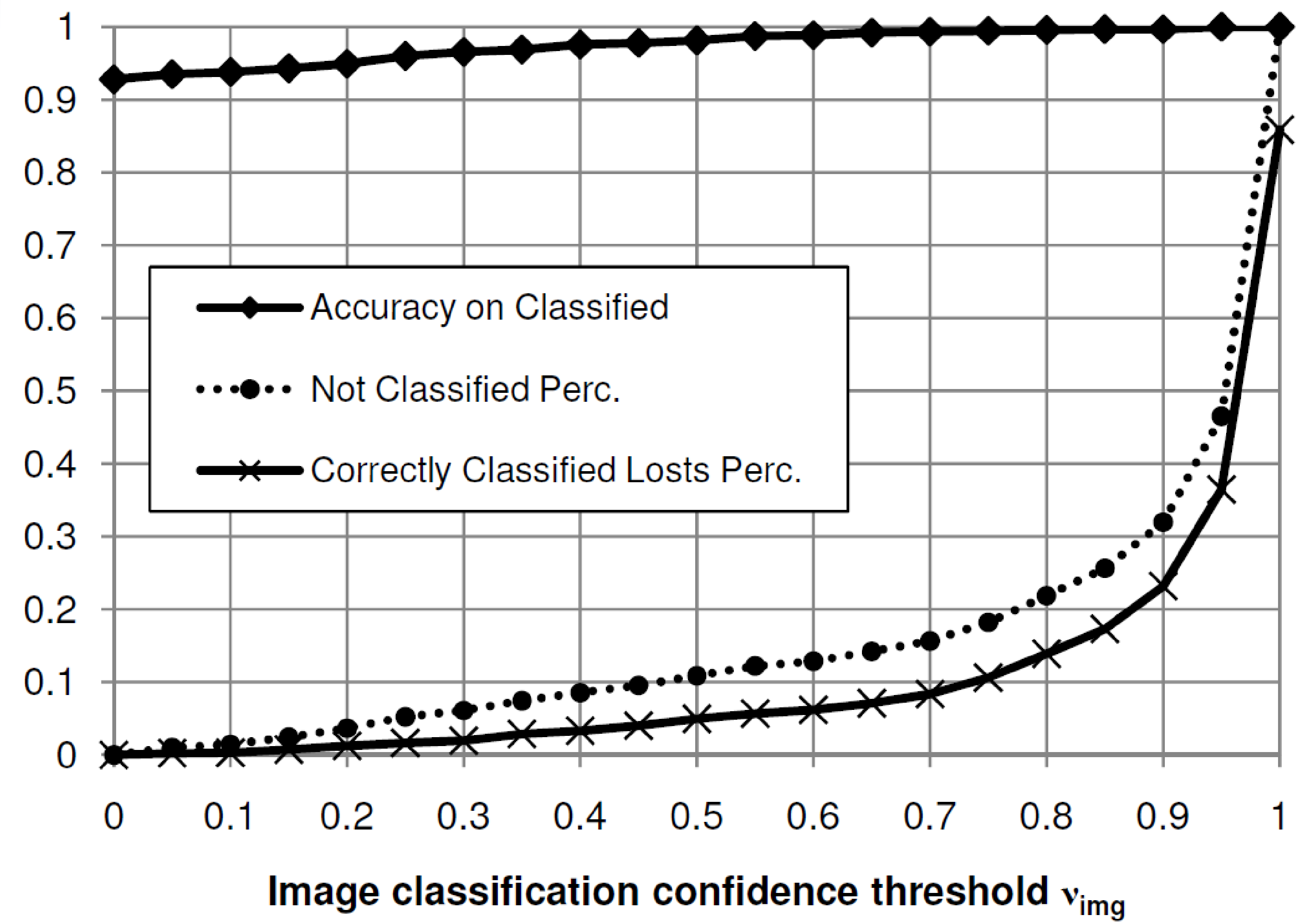
Results

									best	baseline
	classifier	$\hat{\Phi}^{f}$	$\hat{\Phi}^{1}$	$\hat{\Phi}^{5}$	$\hat{\Phi}^{10}$	$\hat{\Phi}^{25}$	$\hat{\Phi}^{50}$	$\hat{\Phi}^{m}$	$\hat{\Phi}^{w}$	$\hat{\Phi}^{s}$
Accuracy	**SIFT**	0.901	0.901	0.855	0.818	0.756	0.691	0.945	**0.952**	0.877
	SURF	0.883	0.881	0.841	0.794	0.714	0.668	0.927	**0.928**	0.851
F_1 Macro	**SIFT**	0.806	0.883	0.809	0.748	0.657	0.575	0.940	**0.947**	0.864
	SURF	0.791	0.866	0.804	0.727	0.606	0.542	0.915	**0.922**	0.828

- Local feature based classifiers perform better
- NN_1 / NN^*_2 distance ratio is relevant
- Weighted approach is better than binary
- Relative efficacy of classifier is the same for both SURF and SIFT
- Good performance of local features based classifiers relying on only 1-NN search betwen local features

Image label confidence

$\hat{\Phi}^{w}$



Conclusions

- Classifying first each local features result in **efficacy** advantage for the overall image classification
- LF based image classification relying on **NN search** over LFs can exploit efficiency advantages of index structure for similarity search given that local features are typically compared using Euclidean distance
- Next step is considering geometric consistency during the whole image classification phase

www.visitotuscany.it

- **Android APP: VISITO Tuscany**

- ***Contact:***
  - ***NeMIS Lab, ISTI -CNR***

fabrizio.falchi@isti.cnr.it
giuseppe.amato@isti.cnr.it