POP: Person Re-Identification Post-Rank Optimisation

Chunxiao Liu — Tsinghua University — lcx08@mails.tsinghua.edu.cn

Chen Change Loy — The Chinese University of Hong Kong — ccloy@ie.cuhk.edu.hk

Shaogang Gong — Queen Mary University of London — sgg@eecs.qmul.ac.uk

Guijin Wang — Tsinghua University — wangguijin@tsinghua.edu.cn

1 Introduction

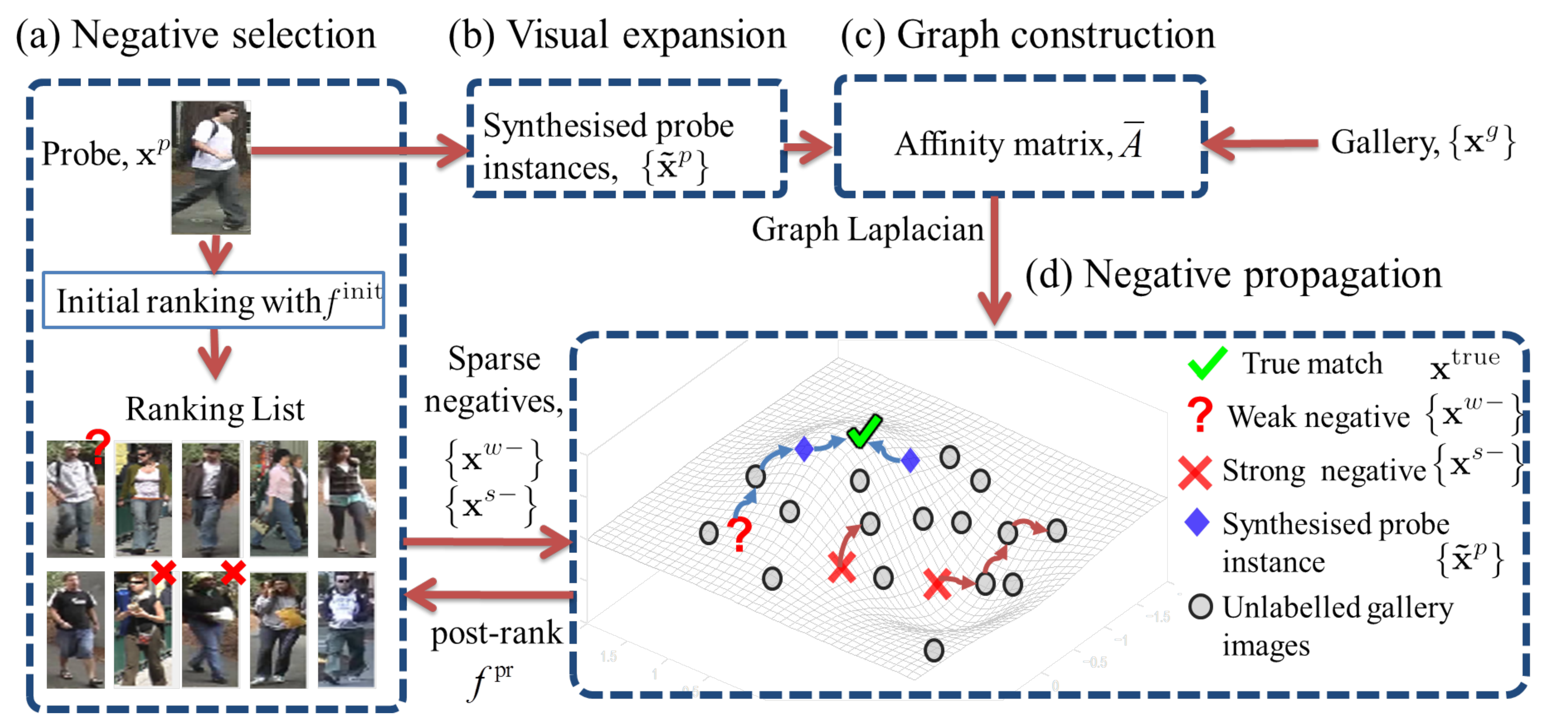
Problem:

- Visual ambiguities and disparities
- Off-line learning scalability (minimising user feedback clicks)

Contributions:

- Formulate a systematic framework for *fast* re-identification post-rank optimisation, with significant increase in recognition accuracy (*over 30% increase for rank-1 recognition rate on VIPeR dataset*).
- Minimise *human-in-the-loop* effort by *one-shot* negative feedback selection (weak/strong negative).
- Formulate a new visual expansion model for overcoming insufficient training samples.
- Incremental affinity graph construction for exploiting large quantities of unlabelled data.

2 Post-Rank Optimsation by Negative Mining



Negative selection : A user selects one (any) strong negative from the top *N* ranked instances, denoted as $\mathbf{x}^{s-}$.

3 Cross-Camera View Visual Expansion

Motivation:

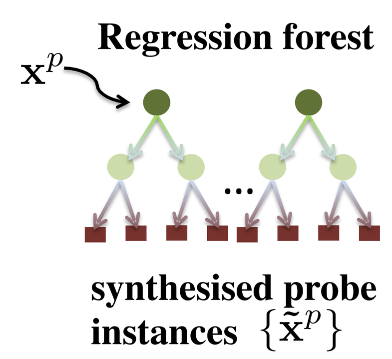
- A single strong negative selected by user is insufficient for learning a post-rank function.
- Due to large feature inconsistency between different camera views, a probe image from the probe camera view cannot be readily used as positive sample in the gallery view.

Method:

☐ Step 1 : **Learning an appearance mapping space:**

$$\mathrm{M}: \mathbf{x}^p \rightarrow \mathbf{x}^g \in \mathbb{R}^d$$

The visual variations between a probe and a gallery camera view are accounted by multi-output regression forest.



☐ Step 2 : **Generating synthesised probe instance:**

$$\tilde{\mathbf{x}}^p = \sum_{t=1}^{T_s} M_{\pi_t}(\mathbf{x}^p) \qquad T_s = \frac{2}{3} T_r$$

- M_t is the regression predictor for the *t*-th regression tree.
- The subscript $\pi = \{\pi_1, \ldots, \pi_{T_s}\}$ is a randomly sampled index.

This process can be repeated to generate more synthesised probe instances if desired.

4 Incremental Construction of Affinity Graph

Motivation:

Propagate the sparse labelled samples to the large quantity of unlabelled set.

Method:

a) **Clustering forest for graph construction**

- Its implicit feature selection mechanism is beneficial to mitigating noisy visual features.
- It offers scalable and tractable solution to our incremental graph construction requirement so to accommodate varying number of selected negatives accumulating on-the-fly.

b) **Incremental construction**

$$\bar{A} \in \mathbb{R}^{n \times n} \rightarrow \bar{A} \in \mathbb{R}^{(n+\tilde{n}) \times (n+\tilde{n})} \qquad \tilde{n} = \Omega(\tilde{\mathbf{x}}^p)$$

☐ Step 1: construct a graph for the gallery instances $\{\mathbf{x}^g\}$.

☐ Step 2: include synthesised positives in the construction of the affinity graph.

5 Negative Propagation over Graph

$$\mathcal{L} = \{\mathbf{x}^{s-}\} \cup \{\mathbf{x}^{w-}\} \cup \{\tilde{\mathbf{x}}^p\}$$

$$\mathcal{U} = \{\mathbf{x}^g\} \setminus (\{\mathbf{x}^{s-}\} \cup \{\mathbf{x}^{w-}\})$$

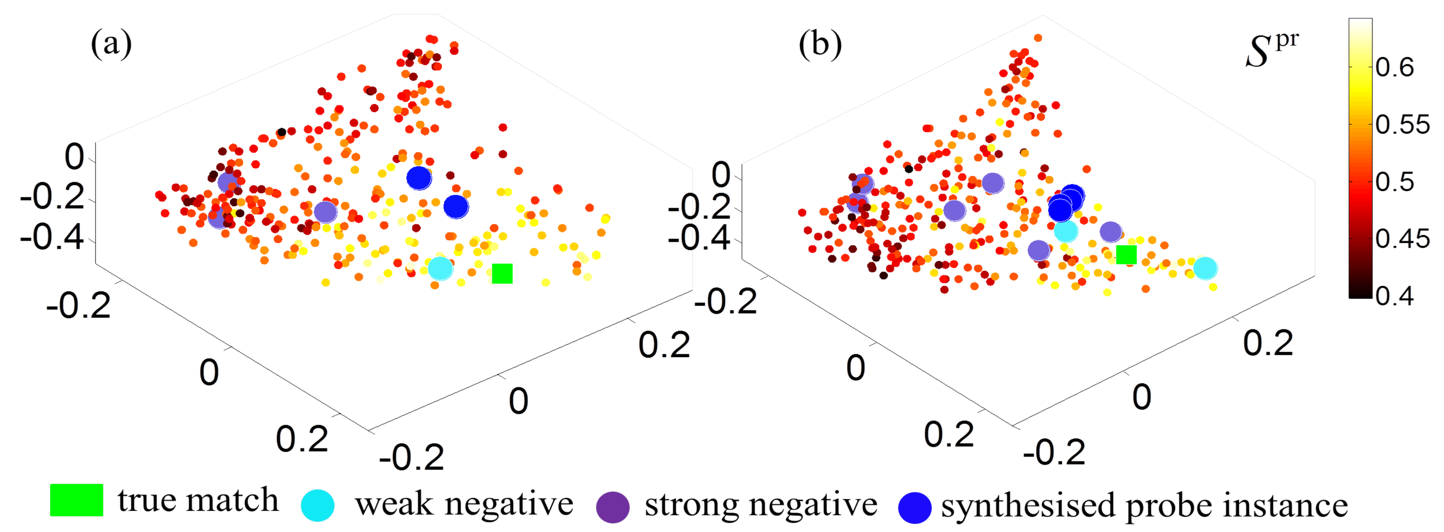
$$y = \begin{cases} +1 & \text{if } \mathbf{x} \in \{\mathbf{x}^{w-}\} \cup \{\tilde{\mathbf{x}}^p\} \\ -1 & \text{if } \mathbf{x} \in \{\mathbf{x}^{s-}\} \end{cases}$$

Information propagation

$$f^{\text{pr}} = \underset{f \in \mathcal{H}_K}{argmin} \frac{1}{l} \sum_{i=1}^{l} \max(1 - y_i f(\mathbf{x}_i), 0) + \lambda_A \|f\|_K^2 + \lambda_I \|f\|_I^2$$

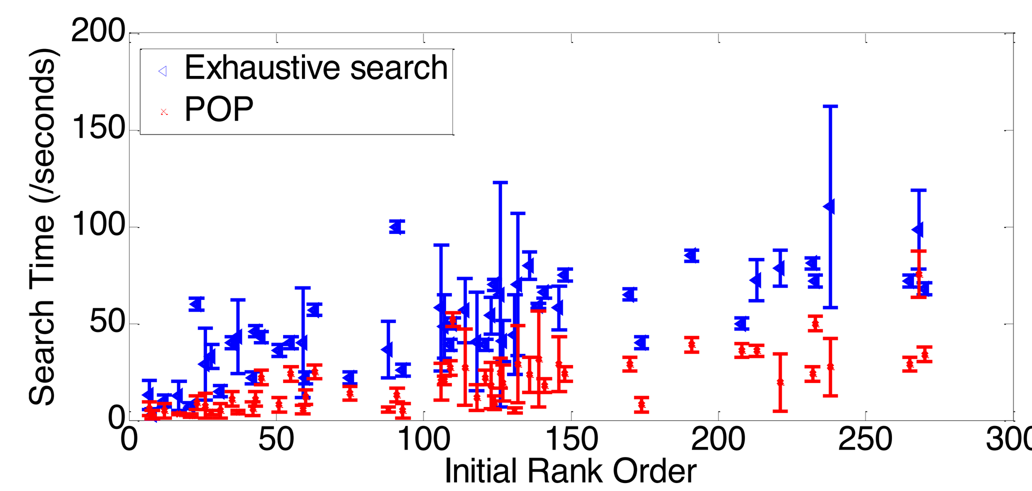
Final score: $s = (1-\beta)s^{\text{init}} + \beta s^{\text{pr}}$

6 Negative Accumulation



(a) Three-dimensional embedding of gallery images obtained using multi-dimensional scaling after the first round of negative selection.
(b) The embedding after the second round.

7 Behavioural Studies



2.6 times reduction of searching time

8 Comparative Evaluations

(a) POP vs. L1-norm, RankSVM, PRDC, MCC

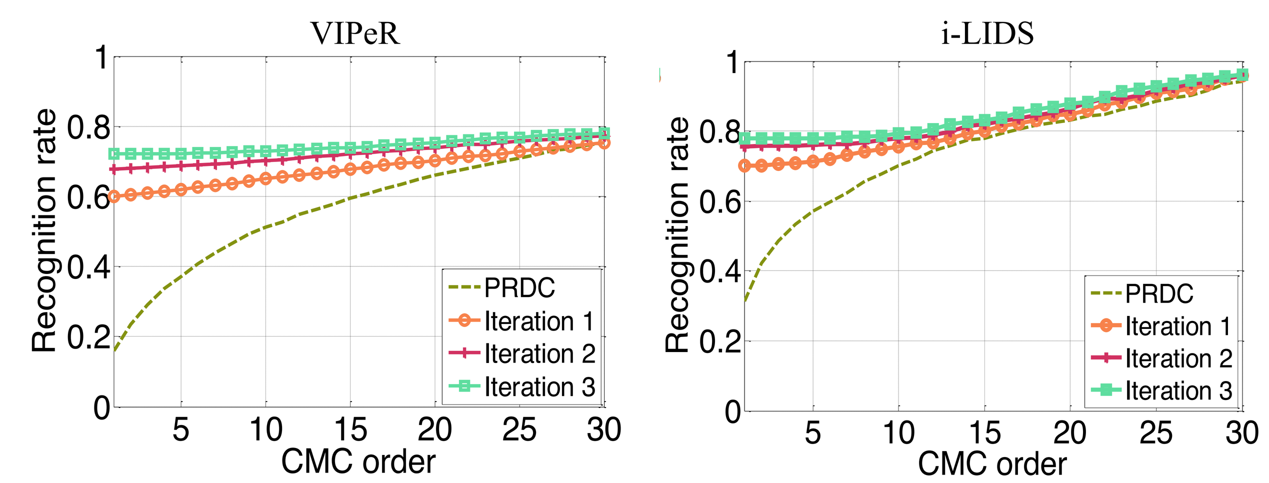
✓ The rank-1 average recognition rates are boosted by *38.33%* and *40.05%* on VIPeR and i-LIDS respectively for all four different initial ranking models.

(b) POP vs. other Post-Rank Models

✓ Improvements of over *12.00%* and *3.70%* at rank-5 recognition rate on VIPeR and i-LIDS respectively, after 4 rounds feedback.

(c) Benefits from Visual Expansion

✓ Improving POP from 37.66% to *51.39%* after 4 rounds feedback on average.



Initial Ranking	VIPeR								i-LIDS							
	one-shot				multi-shots				one-shot				multi-shots			
	0	R1	R2	R3	0	R1	R2	R3	0	R1	R2	R3	0	R1	R2	R3
*L*1-norm	9.43	31.90	42.88	47.56	9.43	30.41	44.21	50.13	29.60	67.60	73.20	75.60	29.60	67.60	75.40	81.80
RankSVM	14.87	59.05	67.06	71.08	14.87	58.48	67.85	71.58	29.80	73.40	77.20	79.80	29.80	73.40	79.40	82.40
PRDC	16.01	59.91	67.88	72.03	16.01	59.49	68.35	72.22	31.40	70.20	75.60	78.00	31.40	70.20	77.00	80.00
MCC	17.85	60.13	64.08	66.87	17.85	60.06	63.64	66.20	30.00	69.80	73.60	76.60	30.00	69.20	74.80	80.40

[1] B. Prosser, et al, "Person re-identification by support vector ranking", BMVC, 2010
[2] W. Zheng, et al, "Re-identification by relative distance comparison", TPAMI, 2012.
[3] A. Globerson, et al, "Metric learning by collapsing classers", NIPS, 2005.

Project page:
http://personal.ie.cuhk.edu.hk/~ccloy/project_pop/