Robust Object Recognition via Third-Party Collaborative Representation

ICPR 2012

Yang Wu[1], Michihiko Minoh[1], Masayuki Mukunoki[1], and Shihong Lao[2]

[1]*Academic Center of Computing and Media Studies, **Kyoto University**, Japan.*

[2]***Omron Social Solutions Co., LTD**, Kyoto, Japan.*

Motivation:

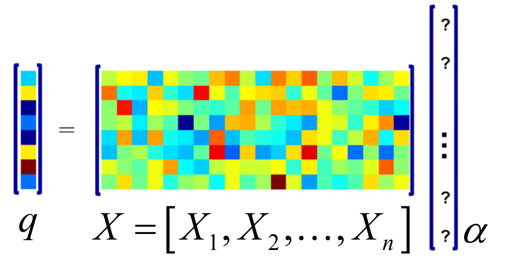
Sparse (Collaborative) Representation for Classification:

Effective and powerful but has two preconditions:

1. The training images have been carefully controlled.
2. The number of samples per class is sufficiently large.

Our goal:

1. To relax these preconditions for broader applications.
2. To be efficient, robust, and easily implementable.



$$\arg\min_{\alpha} \|\alpha\|_1 \quad \text{OR} \quad \arg\min_{\alpha} \|\alpha\|_2 .$$

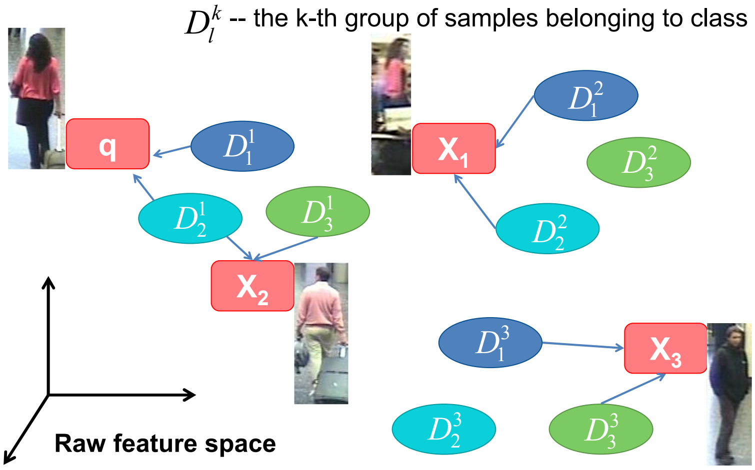
Idea:

Notations:

q -- query/test sample(s),

X_i -- training samples for class i;

D_l^k -- the k-th group of samples belonging to class l.



TPCR Algorithm:

1. Introduce a third-party dictionary

$$D = [D_1, D_2, \ldots, D_L], D_l = [D_l^1, D_l^2, \ldots, D_l^{K_l}], l = 1, \ldots, L.$$

2. Collaborative representation of q & X with D

$$\hat{\alpha}(q) = \arg\min_{\alpha} \{\|q - D\alpha\|_2^2 + \lambda\|\alpha\|_2^2\};$$

$$\hat{\alpha}(x_i^j) = \arg\min_{\alpha} \{\|x_i^j - D\alpha\|_2^2 + \lambda\|\alpha\|_2^2\}.$$

ℓ_2

3. Compute the class-wise summed coefficients

$$\hat{\beta}_l(q) = \sum_{m=1}^{K_l} \hat{\alpha}_{lk}(q), \hat{\beta}_l(x_i^j) = \sum_{m=1}^{K_l} \hat{\alpha}_{lk}(x_i^j), \quad \forall l, \forall i, \forall j.$$

4. Stack into a new feature vec. and normalize.

$$\hat{\beta}(q) = [\hat{\beta}_1(q), \ldots, \hat{\beta}_L(q)]^T, \hat{\beta}(x_i^j) = [\hat{\beta}_1(x_i^j), \ldots, \hat{\beta}_L(x_i^j)]^T.$$

5. Do matching/classification in the feature space of $\hat{\beta}$.

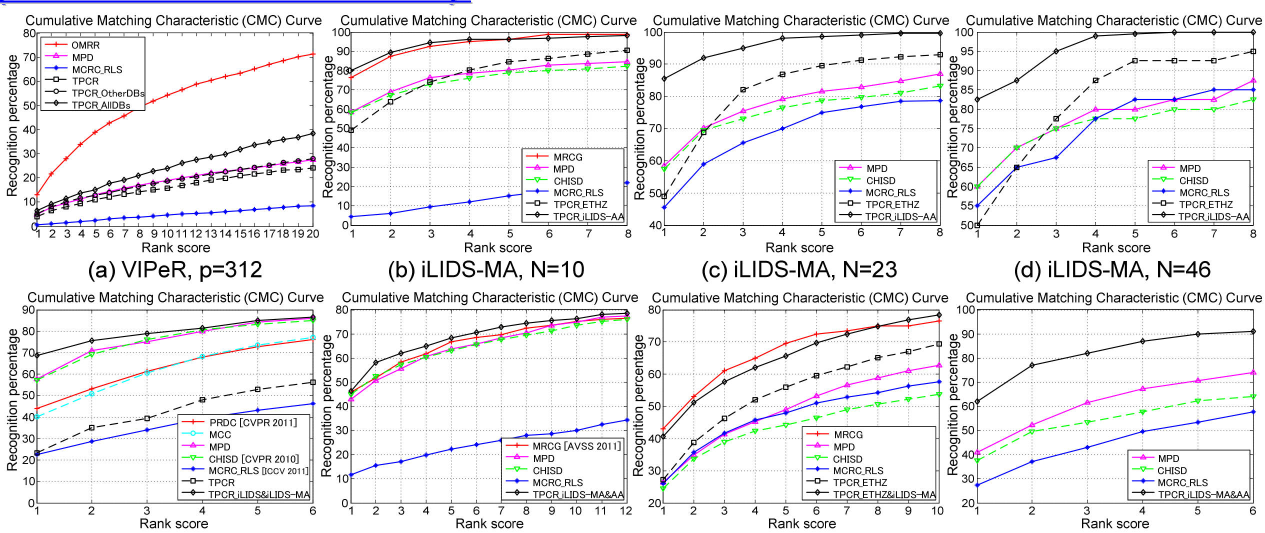
Experiments (on Person Re-identification):

Effectiveness:

TPCR with both similar and non-similar datasets as its dictionary gets very encouraging performance (outperforming the state-of-the-art).

Failure case:

No good dictionary for VIPeR dataset.



(a) VIPeR, p=312 (b) iLIDS-MA, N=10 (c) iLIDS-MA, N=23 (d) iLIDS-MA, N=46

(e) iLIDS, p=30 (f) iLIDS, p=119 (g) iLIDS-AA, N=10 (h) iLIDS-AA, p=30, N=46

Figure 1. Experimental results in CMC curves. p is the number of randomly selected persons for testing. N is the number of randomly selected samples for each query/gallery set for multiple-shot re-identification. When p or N is not specified, it takes the maximum feasible value.

Exp. settings:

Standard color and texture features;
10 times cross-validation when applicable.

Table 1. Dataset properties.

Dataset	NS	NP	NSPP	Seq?	WCV
$VIPeR$ [5]	1264	632	1×2	No	CVIPOB
$iLIDS$ [6]	476	119	2 to 8	Partly	CVIPOB
$iLIDS-MA$ [1]	3680	40	46×2	Yes	CVIPOB
$iLIDS-AA$ [1]	10329	100	21 to 243	Yes	CVIPOBL
$ETHZ-Seq1$ [2]	4857	83	7 to 226	Yes	POB
$ETHZ-Seq2$ [2]	1961	35	6 to 206	Yes	POB
$ETHZ-Seq3$ [2]	1762	28	5 to 356	Yes	POB

Future Directions:

- Evaluation criteria for a good third-party dictionary.
- Selecting and/or learning strategies for obtaining a good dictionary.
- Integration with metric learning techniques.