Collaboratively Regularized Nearest Points for Set Based Recognition

Yang Wu
mm.media.kyoto-u.ac.jp/members/yangwu
Michihiko Minoh
mm.media.kyoto-u.ac.jp/members/minoh
Masayuki Mukunoki
mm.media.kyoto-u.ac.jp/members/mukunoki

Academic Center for Computing and Media Studies
Kyoto University
Kyoto, 606-8501, Japan

Set based recognition (*i.e*, using a set of instances together for recognition) has been attracting more and more attention in recent years, benefitting from two facts: the difficulty of collecting sets of images for recognition fades quickly, and set based recognition models generally outperform the ones for single instance based recognition. In the past few years various approaches have been proposed, which were reviewed in [4] and [7].

In this paper, we propose a novel model called collaboratively regularized nearest points (CRNP) which inherits the merits of simplicity, robustness, and high-efficiency from the very recently introduced regularized nearest points (RNP) method [7] on finding the set-to-set distance using the l_2-norm regularized affine hulls. Meanwhile, CRNP makes use of the powerful discriminative ability induced by collaborative representation, following the same idea as that in sparse recognition for classification (SRC) for image-based recognition [3] and collaborative sparse approximation (CSA) for set-based recognition [5]. However, CRNP uses l_2-norm instead of the expensive l_1-norm for coefficients regularization, which makes it much more efficient.

Given the test/query set $\mathbf{Q}$ and all the training/gallery sets $\mathbf{X}_i, i \in \{1,\ldots,n\}$, CRNP solves the following optimization problem:

$$\begin{aligned} &\min_{\alpha,\beta} \left\{ \|\mathbf{Q}\alpha - \mathbf{X}\beta\|_2^2 + \lambda_1 \|\alpha\|_2^2 + \lambda_2 \|\beta\|_2^2 \right\}, \\ &s.t. \quad \textstyle\sum_k \alpha_k = 1, \sum_{i=1}^n \sum_j \beta_i^j = 1, \end{aligned} \tag{1}$$

where $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_n]$ denotes all the training/gallery sets together; $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_n^T]^T$ are the corresponding coefficients for these sets; λ_1 and λ_2 are trade-off parameters.

This problem inherits the distance finding model from RNP, however, it performs collaborative set-to-sets distance finding using all the training/gallery sets instead of the independent set-to-set distance finding in RNP as illustrated in Figure 1.

The optimization problem (1) with equality constraints can be transformed to the following unconstrained optimization problem:

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left\{ \left\|\mathbf{z} - \hat{\mathbf{Q}}\boldsymbol{\alpha} - \hat{\mathbf{X}}\boldsymbol{\beta}\right\|_2^2 + \lambda_1 \|\boldsymbol{\alpha}\|_2^2 + \lambda_2 \|\boldsymbol{\beta}\|_2^2 \right\}, \tag{2}$$

where $\mathbf{z} = [\mathbf{0}_{1,m}, \sqrt{\gamma_1}, \sqrt{\gamma_2}]^T$ with m denoting the dimensionality of the image feature space and γ_1 and γ_2 denoting the Lagrangian multipliers. $\hat{\mathbf{Q}} = [\mathbf{Q}^T, \sqrt{\gamma_1}\mathbf{1}_{N_q,1}, \mathbf{0}_{N_q,1}]^T$ in which N_q is the number of samples in $\mathbf{Q}$, and $\hat{\mathbf{X}} = [-\mathbf{X}^T, \mathbf{0}_{N_x,1}, \sqrt{\gamma_2}\mathbf{1}_{N_x,1}]^T$ where N_x is the number of samples in $\mathbf{X}$. $\mathbf{0}_{i,j}$ and $\mathbf{1}_{i,j}$ denote the $i \times j$ zero matrix and the $i \times j$ dimensional matrix of ones, respectively.

We follow [7] on alternatively optimizing $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, which avoids the time-consuming matrix inverse operation of an integrated matrix containing both $\mathbf{Q}$ and $\mathbf{X}$ for each test/query set $\mathbf{Q}$. More concretely, when $\boldsymbol{\alpha}$ is fixed, $\boldsymbol{\beta}$ has a closed-form solution

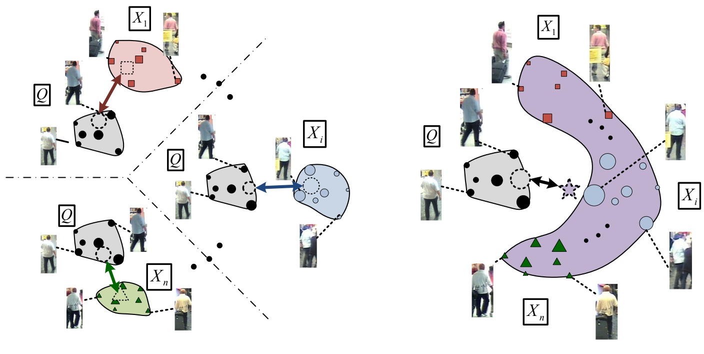
$$\boldsymbol{\beta}^* = \mathbf{P}_x \left(\mathbf{z} - \hat{\mathbf{Q}}\boldsymbol{\alpha}\right), \tag{3}$$

where $\mathbf{P}_x = \left(\hat{\mathbf{X}}^T\hat{\mathbf{X}} + \lambda_2\mathbf{I}\right)^{-1}\hat{\mathbf{X}}^T$ (with $\mathbf{I}$ denoting the identity matrix) only depends on $\mathbf{X}$, so it can be pre-computed. When $\boldsymbol{\beta}$ is fixed, $\boldsymbol{\alpha}$ also has a closed-form solution

$$\boldsymbol{\alpha}^* = \mathbf{P}_q \left(\mathbf{z} - \hat{\mathbf{X}}\boldsymbol{\beta}\right), \tag{4}$$

where $\mathbf{P}_q = \left(\hat{\mathbf{Q}}^T\hat{\mathbf{Q}} + \lambda_1\mathbf{I}\right)^{-1}\hat{\mathbf{Q}}^T$.

As claimed in [7], the objective function in Formula (2) has a lower bound of 0 and it is jointly convex w.r.t. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Since in the alternative optimization, each step on updating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ decreases the objective, the iteration will converge to the global optimal solution. In our experiments to be presented, the iteration usually terminates in no more than 10 steps.



(a) Set-to-set distances (b) Set-to-sets distance

Figure 1: Motivation illustration. (a) The set-to-set distances generated by traditional independent distance finding approaches; (b) the set-to-sets distance got by collaborative distance finding methods.

Besides that, CRNP has a different classification model from that of RNP which makes use of the discriminative coefficients generated by collaborative distance finding. We define the dissimilarity between $\mathbf{Q}$ and $\mathbf{X}_i, i \in \{1,\ldots,n\}$ as

$$d_{CRNP}^i = (\|\mathbf{Q}\|_* + \|\mathbf{X}_i\|_*) \cdot \left\|\mathbf{Q}\boldsymbol{\alpha}^* - \mathbf{X}_i\boldsymbol{\beta}_i^*\right\|_2^2 / \left\|\boldsymbol{\beta}_i^*\right\|_2^2, \tag{5}$$

where $\|\mathbf{Q}\|_*$ is the nuclear norm of $\mathbf{Q}$, i.e. the sum of the singular values of $\mathbf{Q}$. Then, $\mathbf{Q}$ is classified by

$$C(\mathbf{Q}) = \arg\min_i \left\{ d_{CRNP}^i \right\}. \tag{6}$$

Extensive experiments on five benchmark datasets for face recognition and person re-identification demonstrate that CRNP is not only more effective but also significantly faster than other state-of-the-art methods.

Table 1: Computational cost comparison with all the related methods on all of the recognition tasks. The results are averaged over 10 random trials if applicable, and we report them in the "milliseconds per sample" manner to eliminate the influence of dataset size variation. The best results for the methods excluding the less robust "CRC" model are shown in bold.

Dataset	MPD[2]	SRC[3]	CRC[8]	CHISD[1]	SANP[6]	SBDR[4]	CSA[5]	RNP[7]	CRNP
Honda/UCSD (50)	3.2	1.2×10^3	*0.28*	77.7	19.6	259	17.4	11.5	**0.32**
Honda/UCSD (100)	6.4	4.1×10^3	0.55	330	17.3	97.8	32.6	14.5	**0.46**
CMU MoBo (50)	12.4	7.6×10^3	*0.94*	89.0	47.2	85.0	29.0	3.5	**2.1**
CMU MoBo (100)	71.4	2.7×10^4	*1.8*	394	53.0	79.3	39.1	5.9	**2.5**
iLIDS-MA	3.9	741	*0.51*	58.7	121	N/A	9.6	24.5	**3.3**
iLIDS-AA	9.9	2337	*1.2*	150	344	N/A	36.8	83.4	**7.2**
CAVIAR4REID	**3.8**	214	*0.35*	55.3	249	N/A	15.8	30.8	8.0

[1] H. Cevikalp and B. Triggs. Face recognition based on image sets. In *CVPR*, pages 2567–2573, 2010.

[2] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, 2010.

[3] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE TPAMI*, 31(2):210–227, 2009.

[4] Y. Wu, M. Minoh, M. Mukunoki, and S. Lao. Set based discriminative ranking for recognition. In *ECCV*, volume 7574, pages 497–510. 2012.

[5] Y. Wu, M. Minoh, M. Mukunoki, W. Li, and S. Lao. Collaborative sparse approximation for multiple-shot across-camera person re-identification. In *AVSS*, pages 209 –214, 2012.

[6] Y. Hu and A. S. Mian and R. Owens. Sparse Approximated Nearest Points for Image Set Classification. In *CVPR*, pages 121–128, 2011.

[7] M. Yang, P. Zhu, L.V. Gool, and L. Zhang. Face recognition based on regularized nearest points between image sets. In *FG*, 2013.

[8] L. Zhang, M. Yang, and X. Feng. Sparse representation or collaborative representation: which helps face recognition? In *ICCV*, 2011.