Collaborative Mean Attraction for Set Based Recognition

1. Introduction

Set based recognition concerns about the pattern recognition problems for which multiple instances of the same object/class are available and treated together as a set for classification. Such problems widely exist in real applications as getting multiple instances for an individual class is getting easier and easier. Take the visual recognition tasks as an example, it is more and more convenient to get a set of instances either from video records or individual images (shot by cameras or collected from the Internet), The attractiveness as getting multiple instances involved is that it can probably lead to a better recognition performance [10].

Recently, set based recognition has attracted a lot of attention, especially on the popular tasks like face recognition [2][6][13] and person re-identification [9][16][10]. Despite their great variations in detailed modeling, existing methods can be classified into two groups based on how they treat the training data in the classification process: independent models and collaborative models.

Independent models compute an independent set-to-set distance between the test set and each training set, and then classify the test set by such distances. Representative methods include Minimum Point-wise Distance (MPD) [4] which finds the minimum distance between any pair of points (samples in the feature space) with one from each set, Affine/Convex Hull based Image Set Distance (AHISD/CHISD) [2] which computes the geometric distance between two affine/convex hulls generated from the two sets, Sparse Approximated Nearest Points (SANP) [17] and its kernel version KSANP [6] which add sparsity constraints to the CHISD model, Set Based Discriminative Ranking (SBDR) [13] which iterates between CHISD or SANP and metric learning to find the geometric distance in a discriminative metric space, and Regularized Nearest points (RNP) [16] which adopts l_2-norm regularization terms to complement affine hulls and increase both effectiveness and efficiency.

Collaborative models treat all the training sets together as a large indiscriminate set and compute only one single geometric distance between it and the test set, which is also referred to as set-to-sets distance [10]. The collaboration of individual training sets in distance finding stimulates the competition among them which makes the linear combination coefficients discriminative. Existing collaborative models are so far only Collaborative Sparse Approximation (CSA) [9] and Collaboratively Regularized Nearest Points (CRNP) [10], as well as their extended versions with sample/set pre-selection [11][15]. Compared with similar independent models, collaborative ones are not only more effective but also much more efficient.

In this paper we propose a novel collaborative model named Collaborative Mean Attraction (CMA), which is simpler, faster and also more effective than both CSA and CRNP. Unlike them, CMA does not rely on affine/convex hulls but just uses simple l_2-norm based regularization terms to make the linear combinations over both sets (the indiscriminate training set and the test set) to be close to the means of them as much as possible. This strategy anchors the linear combinations whilst forcing them to attract each other by the minimum between-set distance. The model itself has only 2 parameters for balancing the attraction and pulling back, which are much fewer than those in CSA and CRNP. Meanwhile, CMA inherits the efficiency from CRNP, but goes beyond it with its compactness. In the rest part of the paper, we detail CMA in Section 2 and provide extensive experiments in Section 3 to demonstrate its superiority to all the related state-of-the-art methods.

Though it should be very easy to implement CMA according to the paper and reproduce our experimental results, we would like to share our code online for promoting potential applications and extensions once the paper gets accepted.

2. Collaborative Mean Attraction

2.1 Optimization of the coefficients

Given a test set $\mathbf{Q} \in \mathbb{R}^{m\times N_q}$ and all the training sets $\mathbf{X} \in \mathbb{R}^{m\times N_x}$ with $\mathbf{X} = \cup \mathbf{X}_i, i \in \{1,\dots,n\}$, where m is the feature dimensionality, N_q and N_x are the number of samples in $\mathbf{Q}$ and $\mathbf{X}$, respectively, and n is the number of classes, CMA solves the following problem:

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} f\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right), \tag{1}$$

with the objective function

$$f\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right) = \|\mathbf{Q}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda_1 \left\|\boldsymbol{\alpha} - \frac{\mathbf{1}_{N_q,1}}{N_q}\right\|_2^2 + \lambda_2 \left\|\boldsymbol{\beta} - \frac{\mathbf{1}_{N_x,1}}{N_x}\right\|_2^2, \tag{2}$$

where λ_1 and λ_2 are two trade-off parameters, and $\mathbf{1}_{i,j}$ denotes the $i\times j$ dimensional matrix of ones. $\|\mathbf{Q}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}\|_2^2$

is the distance between two linear combinations which can be viewed as the distance between two generalized means (with unequal weights), and the other two terms forces these two generalized means to be not too far away from the actual means. Since the attraction of generalized means occurs between the test set and the indiscriminate training set (all training sets together), this model is called collaborative mean attraction.

Though the above problem becomes a linear least squares problem with an algebraic solution if we pack $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for optimization, we follow [10] on alternatively optimizing them, which avoids the time-consuming matrix inverse (or pseudo-inverse) operation of an integrated matrix containing both $\mathbf{Q}$ and $\mathbf{X}$ for each test/query set $\mathbf{Q}$. In the alternative optimization, the matrix inverse operation on the training data $\mathbf{X}$ is independent of $\mathbf{Q}$, so it can be pre-computed before testing and reused for each $\mathbf{Q}$. More concretely, when $\boldsymbol{\alpha}$ is fixed, the optimization problem becomes

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} \left\| \hat{\mathbf{X}}\boldsymbol{\beta} - \mathbf{Z_q} \right\|_F^2, \tag{3}$$

where $\hat{\mathbf{X}} = [\mathbf{X}^T, \sqrt{\lambda_2}\mathbf{I}_{N_x}]^T$ with $\mathbf{I}_{N_x}$ denoting an $N_x \times N_x$ dimensional identity matrix, and $\mathbf{Z_q} = [(\mathbf{Q}\boldsymbol{\alpha})^T, \sqrt{\lambda_2}\mathbf{1}_{N_x,1}/N_x]^T$. This is a simple least squares problem with an algebraic solution

$$\boldsymbol{\beta}^* = \left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}^T\mathbf{Z_q}. \tag{4}$$

The solution can be further decomposed to

$$\begin{aligned} \boldsymbol{\beta}^* &= \left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\right)^{-1}\mathbf{X}^T\mathbf{Q}\boldsymbol{\alpha} + \lambda_2\left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\right)^{-1}\mathbf{1}_{N_x,1}/N_x && (5)\\ &= \mathbf{P_x}\mathbf{Q}\boldsymbol{\alpha} + \mathbf{C_x}, && (6) \end{aligned}$$

where the projection matrix $\mathbf{P_x} = \left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\right)^{-1}\mathbf{X}^T$ and the constant $\mathbf{C_x} = \lambda_2\left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\right)^{-1}\mathbf{1}_{N_x,1}/N_x$ can be pre-computed once the training data $\mathbf{X}$ is given.

Similarly, when $\boldsymbol{\beta}$ is fixed, $\boldsymbol{\alpha}$ also has an algebraic solution

$$\begin{aligned} \boldsymbol{\alpha}^* &= \left(\hat{\mathbf{Q}}^T\hat{\mathbf{Q}}\right)^{-1}\mathbf{Q}^T\mathbf{X}\boldsymbol{\beta} + \lambda_1\left(\hat{\mathbf{Q}}^T\hat{\mathbf{Q}}\right)^{-1}\mathbf{1}_{N_q,1}/N_q && (7)\\ &= \mathbf{P_q}\mathbf{X}\boldsymbol{\beta} + \mathbf{C_q} && (8) \end{aligned}$$

with $\mathbf{P_q} = \left(\hat{\mathbf{Q}}^T\hat{\mathbf{Q}}\right)^{-1}\mathbf{Q}^T$ and the constant $\mathbf{C_q} = \lambda_1\left(\hat{\mathbf{Q}}^T\hat{\mathbf{Q}}\right)^{-1}\mathbf{1}_{N_q,1}/N_q$, where $\hat{\mathbf{Q}} = [\mathbf{Q}^T, \sqrt{\lambda_1}\mathbf{I}_{N_q}]^T$. $\mathbf{P_q}$ and $\mathbf{C_q}$ can be computed when $\mathbf{Q}$ is available.

The objective $f(\boldsymbol{\alpha}, \boldsymbol{\beta})$ 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 optimum. In our experiments to be presented, it always terminates in only a few steps. To make the optimization as efficient as possible, we propose to follow the procedure of Algorithm 1.

2.2 Classification

The collaborative distance finding implicitly makes $\boldsymbol{\beta}^* = [\boldsymbol{\beta}_1^*, \ldots, \boldsymbol{\beta}_n^*]$ discriminative. Following [10], we define the dissimilarity between $\mathbf{Q}$ and $\mathbf{X}_i, i \in \{1, \ldots, n\}$ as

$$d_{CMA}^i = \left(\|\mathbf{Q}\|_* + \|\mathbf{X}_i\|_*\right) \cdot \left\|\mathbf{Q}\boldsymbol{\alpha}^* - \mathbf{X}_i\boldsymbol{\beta}_i^*\right\|_2^2 / \left\|\boldsymbol{\beta}_i^*\right\|_2^2, \tag{9}$$

where $\|\mathbf{Q}\|_*$ and $\|\mathbf{X}_i\|_*$ are the nuclear norms (i.e. the sum of the singular values) of $\mathbf{Q}$ and $\mathbf{X}_i$, respectively. Then, $\mathbf{Q}$ is classified by

$$C(\mathbf{Q}) = \arg\min_i \left\{ d_{CMA}^i \right\}. \tag{10}$$

This classification model benefits from the discriminative power of $\boldsymbol{\beta}^*$, which tends to make the class-specific reconstruction residual $\|\mathbf{Q}\boldsymbol{\alpha}^* - \mathbf{X}_i\boldsymbol{\beta}_i^*\|_2^2$ smaller and $\|\boldsymbol{\beta}_i^*\|_2^2$ larger for the ground-truth label i than any other labels $j \in \{1, \ldots, n\}, j \neq i$.

Algorithm 1 COLLABORATIVE MEAN APPROXIMATION (CMA):

Require: The training/gallery sets $\mathbf{X} \in \mathbb{R}^{m \times N_x}$, an arbitrary test/query set $\mathbf{Q} \in \mathbb{R}^{m \times N_q}$, two trade-off parameters $\{\lambda_1, \lambda_2\}$, and two iteration termination thresholds $\{R_{th}, T_{th}\}$.

Ensure: The representation coefficients for distance finding: $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$.

Construct $\hat{\mathbf{X}} = [\mathbf{X}^T, \sqrt{\lambda_2}\mathbf{I}_{N_x}]^T$ and pre-compute $\mathbf{P_x} = \left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\right)^{-1}\mathbf{X}^T$ and $\mathbf{C_x} = \lambda_2\left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\right)^{-1}\mathbf{1}_{N_x,1}/N_x$. Note that $\left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\right)^{-1}$ can be shared in the pre-computation. The pre-computed $\mathbf{P_x}$ and $\mathbf{C_x}$ shall be saved for re-usage in testing any new test/query set.
Compute $\mathbf{P_q} = \left(\hat{\mathbf{Q}}^T\hat{\mathbf{Q}}\right)^{-1}\mathbf{Q}^T$ and the constant $\mathbf{C_q} = \lambda_1\left(\hat{\mathbf{Q}}^T\hat{\mathbf{Q}}\right)^{-1}\mathbf{1}_{N_q,1}/N_q$, where $\hat{\mathbf{Q}} = [\mathbf{Q}^T, \sqrt{\lambda_1}\mathbf{I}_{N_q}]^T$.
Compute $\mathbf{M_{qx}} = \mathbf{P_q}\mathbf{X}$ and $\mathbf{M_{xq}} = \mathbf{P_x}\mathbf{Q}$.
Initialize $\boldsymbol{\alpha}_0 = 1/N_q$, $\boldsymbol{\beta}_0 = 1/N_x$, the change rate of the objective value $r = 1$, and the iteration index $t = 1$.
Compute the initial objective value: $f(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$.
while $r \geq R_{th}$ **or** $t \leq T_{th}$ **do**
    Update the representation coefficients:
        $\boldsymbol{\alpha}_t = \mathbf{M_{qx}}\boldsymbol{\beta}_{t-1} + \mathbf{C_q}$;
        $\boldsymbol{\beta}_t = \mathbf{M_{xq}}\boldsymbol{\alpha}_t + \mathbf{C_x}$.
    Update the change rate of the objective value:
        $r = |f(\boldsymbol{\alpha}_t, \boldsymbol{\beta}_t) - f(\boldsymbol{\alpha}_{t-1}, \boldsymbol{\beta}_{t-1})| / f(\boldsymbol{\alpha}_{t-1}, \boldsymbol{\beta}_{t-1})$.
    t = t+1;
end while
Return $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$.

3. Experiments and Results

In order to show the superiority of CMA to all the related state-of-the-art methods when applicable, we make our experiments exactly follow those presented in [10] as they are so far most comprehensive and up-to-date.

3.1 Experimental settings

Datasets and Experiments. The datasets are selected for the two most representative and popular set-based recognition tasks: face recognition and person re-identification. For face recognition, the widely adopted Honda/UCSD dataset [7] and CMU MoBo dataset [5] with the first 50 or 100 frames of each sequence are used (following [17], [13], [16] and [10]). Honda/UCSD contains gray images with 20×20 pixels and they are directly used as features. The specified 20 sequences of 20 subjects are used for training and the rest 39 sequences are left for testing. For CMU MoBo, the (1475-dimensional) LBP features are adopted, and we perform a 10-time result averaging by randomly choosing one sequence from the four candidates for each of the 24 subjects and have the unselected ones left for testing.

For person re-identification, three representative and complementary datasets: iLIDS-MA, iLIDS-AA [1] and CAVIAR4REID [3] are tested on. The first two were captured by two non-overlapping cameras at an airport with large viewpoint changes. The iLIDS-MA dataset has 40 persons with manually cropped images, while iLIDS-AA contains as many as 100 individuals collected by an automatic tracking algorithm which brings significant localization errors. For both datasets, we perform 10-time result averaging by randomly choosing 10 images for each training/test set. Unlike iLIDS-MA and iLIDS-AA, CAVIAR4REID consists of several sequences filmed in a shopping centre. Besides viewpoint changes, it has broad resolution changes and severe pose variations. We follow [3] on training with 22 specified subjects and testing on the other 50. Each set (either for training or test) contains 5 randomly sampled images, and 10-time result averaging is performed. We use the same 400-dimensional color and texture histograms based features as mentioned in [12] for all of them.

Methods. Following [10], all the related state-of-the-art methods, including MPD[4], SRC[8], CRC[18], CHISD[2], SANP[17], KSANP[6], SBDR[13], CSA[9], RNP[16] and CRNP[10], are compared with using exactly the same experimental settings if possible. As mentioned in [10], here SRC and CRC stand for the extended versions for set-based recognition from their original models. For KSANP and SBDR, only the results listed for the same tasks in their original papers are included, while for the others, we get the results by running their codes either from their authors (such as SRC, CHISD, SANP, CSA, and CRNP) or implemented by ourselves (such as MPD, CRC and RNP).

Parameters. We used the parameters recommended in their original papers for all the other methods, while for CMA, the detailed settings are shown in Table 1. Note that λ_1 and λ_2 are set to fit each dataset with the observations that they should be larger (looser constraints on the closeness to means) if we have more samples per class, and that larger within-class variations should lead to larger λ_1 and smaller λ_2 to make CMA robust to noises/errors in $\mathbf{Q}$ whilst tolerate diverse samples in $\mathbf{X}$ a bit more (e.g. iLIDS-AA). Values in Table 1 are just found to be good enough without fine tuning, so there might be better choices. The actual performances are not sensitive to these parameters (including R_{th} and T_{th}) in large ranges, and it should be easy to find good values for them given any new dataset or task.

3.2 Experimental results and analysis

The results for all the concerned methods on the seven experiments are listed in Table 2. To make the comparison as fair as possible, we tried our best to generate the results for other methods with exactly the same data splits as those for CMA, when it is doable. Meanwhile, we also cite the originally reported results for those methods which have been tested on the same datasets (though the experimental data may be slightly different). It is worth noticing that the referred results for SBDR on CMU MoBo were generated with the training set fixed to be the frontal sequence, which is unlike the completely random sequence sampling in this paper and also the other papers. Therefore, they are not counted for performance competition. Though person re-identification is widely treated as a ranking problem and evaluated by the Cumulative Matching Characteristic (CMC) curve [14], here we only report the rank-1 accuracy to make it consistent with that for face recognition.

Table 1 Parameter setting of CMA for all the experiments.

Experiment	λ_1	λ_2	R_{th}	T_{th}
Honda/UCSD (50 frames)	8	16	0.01	15
Honda/UCSD (100 frames)	12	24	0.01	15
CMU MoBo (50 frames)	0.1	0.2	0.01	15
CMU MoBo (100 frames)	0.4	0.8	0.01	15
iLIDS-MA	4	20	0.01	15
iLIDS-AA	10	2	0.01	15
CAVIAR4REID	2	10	0.01	15

It is clear that the proposed CMA method generally outperforms all the other methods in terms of recognition accuracy. The only weakness (more precisely it should be just not a superiority) is that it cannot make the best out of rich samples for each class/set (such as 100 frames for Honda/UCSD and CMU MoBo). Like CRNP, it indicates that too large set size may weaken the discriminative power of the l_2-norm based collaborative distance finding as it is more likely that some samples from different classes may replace the correct class in collaborative representation [10] with a smaller distance to the test set. Therefore, a pre-selection step may be helpful for reducing such a risk, as suggested in [11] and [15]. It could be an interesting future work.

3.3 Computational cost

All the methods compared on efficiency are implemented in Matlab and run on a 2.67 GHz dual-core machine with 20GB memory. Since some of them can have (parts of) their models pre-computed before testing, we report the precomputation time for them in Table 4. The results show that the training phase of SBDR is very time-consuming, while the pre-computation time for other three methods are ignorable. In greater detail, CMA needs only slightly more time (less than 2.5s in the worst case) than RNP and CRNP. The testing time for all the methods is listed in Table 3, showing that CMA is generally more efficient than all the others (times faster than CRNP when the number of classes are large, as for person re-identification) except CRC because it just performs MPD in a low-dimensional projected space.

It is worth noticing that when there is only one concerned object in a single frame of video records, CMA itself can fin-

Table 3 Computational cost comparison. Averaged over 10 random trials if applicable and measured in "milliseconds per sample" to eliminate the influence of dataset size variation. The best results for methods excluding "CRC" are shown in bold.

Experiment	MPD [4]	SRC [8]	CRC [18]	CHISD [2]	SANP [17]	SBDR [13]	CSA [9]	RNP [16]	CRNP [10]	CMA
Honda/UCSD (50)	3.2	1.2×10^3	*0.28*	77.7	19.6	259	17.4	11.5	**0.32**	0.41
Honda/UCSD (100)	6.4	4.1×10^3	0.55	330	17.3	97.8	32.6	14.5	**0.46**	**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	**1.96**
CMU MoBo (100)	71.4	2.7×10^4	*1.8*	394	53.0	79.3	39.1	5.9	2.5	**2.1**
iLIDS-MA	3.9	741	*0.51*	58.7	121	N/A	9.6	24.5	3.3	**1.3**
iLIDS-AA	9.9	2337	*1.2*	150	344	N/A	36.8	83.4	7.2	**2.7**
CAVIAR4REID	3.8	214	*0.35*	55.3	249	N/A	15.8	30.8	8.0	**2.2**

Table 2 Recognition accuracy (%) comparison. The results with stars are directly copied from their original papers, while those without stars are got from our experiments. The best ones are shown in bold (results from our experiments and results from cited papers are compared separately, and if the best ones from cited results are better than those from our experiments they are marked in bold as well). The ones which do not exist in cited papers are stated as "not available (N/A)".

Experiment	Honda/UCSD		CMU MoBo		iLIDS-	iLIDS-	CAVIAR-
	50fs	100fs	50fs	100fs	MA	AA	4REID
MPD[4]	79.5	87.2	92.2	94.3	50.0	23.8	19.0
SRC[8]	76.9	**94.9**	88.9	92.4	57.3	36.0	25.4
CRC[18]	76.9	82.1	89.7	93.1	28.5	24.7	16.6
CHISD	79.5	79.5	90.8	94.2	52.5	24.6	25.4
CHISD*[2]	82.1*	84.6*	N/A	N/A	N/A	N/A	N/A
SANP	84.6	89.7	90.1	93.6	46.8	19.2	25.2
SANP*[17]	84.6*	92.3*	N/A	N/A	N/A	N/A	N/A
KSANP*[6]	87.2*	94.9*	N/A	N/A	N/A	N/A	N/A
SBDR*[13]	87.7*	89.2*	*95.0**	*96.1**	N/A	N/A	N/A
CSA[9]	84.6	92.3	86.3	94.4	59.0	22.5	24.6
RNP	66.7	92.3	91.8	**94.6**	53.3	25.5	24.0
RNP*[16]	87.2*	94.9*	91.9*	**94.7***	N/A	N/A	N/A
CRNP	89.7	**94.9**	93.5	94.4	59.3	35.5	27.0
CRNP*[10]	89.7*	**97.4***	93.3*	94.4*	59.0*	35.4*	26.8*
CMA	**92.3**	**94.9**	**94.4**	**94.6**	**61.3**	**36.1**	**28.4**

Table 4 For those methods which can have (parts of) their models pre-computed using the training data, the total precomputation time (in seconds) is listed for comparison.

Experiment	Honda/UCSD		CMU MoBo		iLIDS-	iLIDS-	CAVIAR-
	50fs	100fs	50fs	100fs	MA	AA	4REID
SBDR [13]	9230	14600	12300	31400	N/A	N/A	N/A
CSA [9]	0.59	0.74	28.7	50.2	0.39	0.62	0.26
RNP [16]	0.06	0.20	0.17	0.64	0.02	0.05	0.02
CRNP [10]	0.22	0.87	0.64	2.66	0.04	0.22	0.02
CMA	0.26	1.20	0.73	3.03	0.04	0.27	0.02

ish recognizing more than 370 frames per second with feature vectors of 400 dimensions or even more, which is about 15 times faster than real time (suppose the video has a 25 fps frame rate). This is a good news for people who care about processing speed in real applications.

4. Conclusion and Future Work

This paper presents a novel set based recognition model named CMA which is generally more effective and significantly faster than other related methods. Experimental results on various benchmark datasets for both face recognition and person re-identification have demonstrated its superiority. A possible future work will be pre-selecting samples/sets before applying CMA for the cases when large amounts samples/sets are available.

References

[1] Bak, S., Corvee, E., Bremond, F. and Thonnat, M.: Boosted human re-identification using Riemannian manifolds, *Image and Vision Computing*, Vol. 30, No. 6-7, pp. 443 – 452 (2012).

[2] Cevikalp, H. and Triggs, B.: Face recognition based on image sets, *CVPR*, pp. 2567–2573 (2010).

[3] Cheng, D. S., Cristani, M., Stoppa, M., Bazzani, L. and Murino, V.: Custom Pictorial Structures for Re-identification, *British Machine Vision Conference (BMVC)*, pp. 68.1–8.11 (2011). http://dx.doi.org/10.5244/C.25.68.

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

[5] Gross, R. and Shi, J.: The CMU Motion of Body (MoBo) Database, Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (2001).

[6] Hu, Y., Mian, A. S. and Owens, R.: Face Recognition Using Sparse Approximated Nearest Points between Image Sets, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 34, No. 10, pp. 1992–2004 (2012).

[7] Lee, K., Ho, J., Yang, M.-H. and Kriegman, D.: Video-Based Face Recognition Using Probabilistic Appearance Manifolds, *CVPR*, p. 313–320 (2003).

[8] Wright, J., Yang, A., Ganesh, A., Sastry, S. and Ma, Y.: Robust Face Recognition via Sparse Representation, *IEEE TPAMI*, Vol. 31, No. 2, pp. 210–227 (2009).

[9] Wu, Y., Minoh, M., Mukunoki, M., Li, W. and Lao, S.: Collaborative Sparse Approximation for Multiple-Shot Across-Camera Person Re-identification, *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, pp. 209 –214 (2012).

[10] Wu, Y., Minoh, M. and Mukunoki, M.: Collaboratively Regularized Nearest Points for Set Based Recognition, *Proceedings of the British Machine Vision Conference*, BMVA Press, pp. 1–10 (2013).

[11] Wu, Y., Minoh, M. and Mukunoki, M.: Locality-constrained Collaborative Sparse Approximation for Multiple-shot Person Re-identification, *The 2nd IAPR Asian Conference on Pattern Recognition (ACPR2013)*, pp. 140–144 (2013).

[12] Wu, Y., Minoh, M., Mukunoki, M. and Lao, S.: Robust Object Recognition via Third-Party Collaborative Representation, *21st International Conference on Pattern Recognition (ICPR)* (2012).

[13] Wu, Y., Minoh, M., Mukunoki, M. and Lao, S.: Set Based Discriminative Ranking for Recognition, *Computer Vision - ECCV 2012* (Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y. and Schmid, C., eds.), Lecture Notes in Computer Science, Vol. 7574, Springer Berlin Heidelberg, pp. 497–510 (2012).

[14] Wu, Y., Mukunoki, M., Funatomi, T., Minoh, M. and Lao, S.: Optimizing Mean Reciprocal Rank for person re-identification, *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, pp. 408 –413 (2011).

[15] Wu, Y., Mukunoki, M. and Minoh, M.: Locality-constrained Collaboratively Regularized Nearest Points for Multiple-shot Person Re-identification, *Proc. of The 20th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)* (2014).

[16] Yang, M., Zhu, P., Gool, L. V. and Zhang, L.: Face Recognition based on Regularized Nearest Points between Image Sets, *The 10th IEEE International Conference on Automatic Face and Gesture Recognition (FG)* (2013).

[17] Yiqun Hu and Ajmal S. Mian and Robyn Owens: Sparse Approximated Nearest Points for Image Set Classification, *CVPR*, pp. 121–128 (2011).

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