FEATURE ENCODING IN BAND-LIMITED DISTRIBUTED SURVEILLANCE SYSTEMS

Alireza Rahimpour, Ali Taalimi, Hairong Qi

Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville, TN, USA 37996
{arahimpo, ataalimi, hqi}@utk.edu

ABSTRACT

Distributed surveillance systems have become popular in recent years due to security concerns. However, transmitting high dimensional data in bandwidth-limited distributed systems becomes a major challenge. In this paper, we address this issue by proposing a novel probabilistic algorithm based on the divergence between the probability distributions of the visual features in order to reduce their dimensionality and thus save the network bandwidth in distributed wireless smart camera networks. We demonstrate the effectiveness of the proposed approach through extensive experiments on two surveillance recognition tasks.

***Index Terms*—** Feature Encoding, Compact Data Representation, Distributed Recognition Systems, Band-limited Wireless Camera Network.

1. INTRODUCTION

As the concerns about public safety increase in recent years (due to some incidents such as Boston bombing, etc.), researchers have focused more on developing surveillance systems based on distributed wireless smart cameras. These cameras can cooperate, forming a wireless visual sensing network whose nodes besides visual sensing, also have processing, storage and communication capabilities. Because the smart camera networks have become increasingly more affordable and perform better in balancing the computational power and energy efficiency, they have been employed in many surveillance tasks including distributed object recognition [1], [2], [3], [4], cross view action recognition [5], [6] and person re-identification [7], [8] to name just a few.

However, a major challenge in visual sensor networks is limitation in terms of transmission bandwidth, storage and processing power. In the traditional system design for visual sensor networks, images are acquired and compressed locally at the camera nodes, and then transmitted to the base station which performs the specific analysis tasks (e.g., video surveillance, object recognition, etc.). However, recently, a new paradigm has emerged based on analyze-then-compress, where the visual content is processed locally at the camera nodes, to extract a concise representation constituted by local visual features (e.g., SIFT, SURF, HOG). Such features are then compressed and transmitted to the base station for further analysis. Since the feature-based representation is usually more compact than the pixel-based representation, the analyze-then-compress approach is particularly attractive for those scenarios for which the bandwidth is scarce [1].

Recently several works have been done towards the analyze-then-compress feature compression. For instance, in [9] and [10], authors explore approaches for scalability in large-scale camera networks using recent advances in Compressive Sensing (CS). In [11], the dimension of the features is reduced using an approach based on Sparse Principal Component Analysis (SPCA). Furthermore, [12] argued that reliable feature correspondence can be established in a much lower dimensional space between cameras, even if the feature vectors are linearly projected onto a random subspace. [2] studied a SIFT-feature selection algorithm, where the number of SIFT features that need to be transmitted to the base station is reduced by considering the joint distribution of the features among multiple camera views of a common object.

Another study [13], further considered using robust structure-from-motion techniques (e.g., RANSAC) to select strong object features between two camera views, and subsequently rejecting weak features from the final stage of object recognition. Moreover, some variant works such as [14] focused on multi-view feature engineering and learning (e.g., introducing some new descriptors such as Multi-View HOG). However, such solutions are known to break down easily when the camera transformation is large or when the features are extracted from low-quality images. Moreover, most of the existing approaches (e.g., [13], [2]) require the communication between the smart cameras for selecting the best visual features.

The contribution of this work is three-fold. First, we propose a probabilistic algorithm based on the divergence between the probability distributions of the visual features in order to select the most informative visual features and building a compact and physically meaningful model of the training set. Second, we introduce a scheme for calculating the low-dimensional codes for visual feature of each image, before its transmission to the base station and without any communications between the cameras. Third, we elaborate on the distributed recognition task and illustrate the performance of the proposed approach based on the experiments on two challenging and low-resolution multi-view datasets.

The remainder of this paper is organized as follows. Section 2 elaborates on the proposed method for obtaining a compact model of the feature histograms in the off-line training stage and then introduces a scheme which uses this compact representation of the training set to encode the histogram of features to low-dimensional codes. Section 3 describes the experimental setting, followed by detailed discussion and comparison of the results. The final section concludes the paper.

2. METHODOLOGY

In our distributed recognition system, dense SIFT feature descriptors [15] are computed at a grid of overlapped patches in the image and further quantized to form a dictionary of visual words using the bag-of-words (BoW) approach [16]. Using the hierarchical k-means, all the feature descriptors are clustered into visual words and a term-frequency visual histogram is defined for each image. After performing the feature encoding via the proposed method, the low dimensional encoded feature histogram of the test image is sent to the base station and a nearest neighbor search (using the chi-square distance) is conducted in the training set to find the closest histogram to the testing sample. The class label of the closest histogram in the training set is then used to label the histogram of the test data.

2.1. Compact Representation of the Features

In recent years several studies have been carried out in the context of compact dictionary learning [17], [18], [19], as an approach for finding a compact representation of the data [20], [21]. However, the lack of physical interpretation of the compact dictionary (i.e., physical meaning of each basis in the dictionary) has been a critical shortcoming of the standard dictionary learning techniques. Inspired by the method in [22] based on non-negative matrix factorization, in this paper, we address this issue by proposing a novel probabilistic approach for selecting a group of features as a compact representation of all the features in the training set. We believe that nothing is more meaningful for representing the data than the data itself.

Assume there are c classes in the training set and there are N feature histograms $\boldsymbol{h}_i \in R^m$ in each class (i.e., $H = \{\boldsymbol{h}_1, \dots, \boldsymbol{h}_N\} \in R^{m \times N}$). When each bin of histogram is divided by the number of visual words in each cluster, the probability density function (*pdf*) which represents a probability distribution is produced. Therefore, for each class we have N *pdf*s as: $F = \{\boldsymbol{f}_1, \dots, \boldsymbol{f}_N\} \in R^{m \times N}$.

The objective of this stage of the proposed approach is to compare all the *pdf*s in each class and select a few informative ones by solving the following optimization problem:

$$\begin{aligned} \min_{w_{ij}} \sum_{j=1}^{N} \sum_{i=1}^{N} (\sum_{k=1}^{m} ((\boldsymbol{f}_i(k) - \boldsymbol{f}_j(k)) \ln(\frac{\boldsymbol{f}_i(k)}{\boldsymbol{f}_j(k)})) w_{ij} \\ s.t. \sum_{i=1}^{N} w_{ij} = 1, \, \forall j; \;\; (\sum_{i=1}^{N} (\sum_{j=1}^{N} |w_{ij}|^q)^{p/q})^{1/p} \leq \lambda \\ w_{ij} \geq 0, \, \forall i, j, \end{aligned} \tag{1}$$

where $\sum_{k=1}^{m} ((\boldsymbol{f}_i(k) - \boldsymbol{f}_j(k)) \ln(\frac{\boldsymbol{f}_i(k)}{\boldsymbol{f}_j(k)})$ is the symmetric form of the KL divergence [23]. This term measures the difference between all the probability distribution pairs in F.

w_{ij} is defined as the probability of $\boldsymbol{f}_i$ being a representative for $\boldsymbol{f}_j$ (i.e., $w_{ij} \in [0, 1]$). Therefore, we must have $\sum_{i=1}^{N} w_{ij} = 1$, to assure that the probability of each $\boldsymbol{f}_j$ being represented via $F = \{\boldsymbol{f}_1, \dots, \boldsymbol{f}_N\}$ is equal to one. Hence, the first term in Eq. 1 is the cost of representing $\boldsymbol{f}_j$ via $\boldsymbol{f}_i$, which is defined as the divergence measure between them, times the probability of the occurrence of this event. We define $\boldsymbol{W} \in R^{N \times N}$ as the probability matrix for all the $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$ pairs (i.e., w_{ij} is the ith row and jth column entry of the $\boldsymbol{W}$ matrix). In other word, when $\boldsymbol{f}_i$ is a representative for $\boldsymbol{f}_j$, the corresponding row in the $\boldsymbol{W}$ matrix is non-zero. Since our goal is to find some few representations of $\boldsymbol{f}_j$ using $\boldsymbol{f}_i$, we impose a row-sparsity constraint on the $\boldsymbol{W}$ matrix in order to select only a few $\boldsymbol{f}_i$s and set the other rows of $\boldsymbol{W}$ entirely equal to zero.

In order to achieve this goal, we exploit a joint $L_{p,q}$ norm regularization (the second constraint in Eq. 1), where λ is a regularization parameter and determines the number of non-zero rows of the $\boldsymbol{W}$ matrix. The L_{pq} norm, is convex for $p \geq 1$ and $q \geq 1$; otherwise it is a quasi-norm and is non-convex. In fact, we can consider any $q \geq 1$, however, $L_{p,\infty}$ (i.e., $q = \infty, p \leq 1$) has the property of giving us the real number of non-zero features which is the desired goal in our feature selection task. The $L_{p,\infty}$ penalty is a convex relaxation of a pseudo-norm which counts the number of non-zero rows in $\boldsymbol{W}$. Another consideration in L_{pq} norm is the choice of p. Some works such as [24] have investigated the L_p norm with $0 < p < 1$. It is worth noting that for $0 < p < 1$, Eq. 1 is not a convex problem and we cannot guarantee the global minimum and the solution is not unique and it highly depends on initialization. In other words, even though $0 < p < 1$ might lead to more sparse result, but the solution would not be consistent. Additionally, choosing an optimum initialization method is not straight forward. Hence, in this work, we consider $p = 1$ that leads to a convex problem and global minimum for the optimization problem in Eq. 1. As a result, the proposed optimization problem will have the following form:

$$\min_{w_{ij}} \sum_{j=1}^{N} \sum_{i=1}^{N} (\sum_{k=1}^{m} ((\boldsymbol{f}_i(k) - \boldsymbol{f}_j(k)) \ln(\frac{\boldsymbol{f}_i(k)}{\boldsymbol{f}_j(k)}))w_{ij}$$
$$s.t. \ \sum_{i=1}^{N} w_{ij} = 1, \ \forall j; \quad \sum_{i=1}^{N} (\max_{1 \leq j \leq N} |w_{ij}|) \leq \lambda \tag{2}$$
$$w_{ij} \geq 0, \ \forall i, j,$$

We refer to Eq. 2 as the Divergence-based Feature Selection (DFS) method. We select the feature histograms, corresponding to indices of non-zero rows of $\boldsymbol{W}$ as our representative features in each class and we repeat this process for all the classes in the training set. The number of selected features for each class is determined by the regularization parameter λ (i.e., λ is roughly the number of non-zero rows in the $\boldsymbol{W}$ matrix). It is important to note that the value of λ should satisfy $\lambda \leq N$ (i.e., N is the number of training data in each class), otherwise the $\boldsymbol{W}$ matrix would be the identity matrix, since each probability distribution $\boldsymbol{f}_i$ is the best representation for itself. The convex optimization problem in Eq. 2 is solved using the Alternating Direction Method of Multipliers framework in [25] (more details about the optimization steps can be found in [22]).

2.2. Generating Low Dimensional Feature Codes

After constructing a compact representation of the feature histograms for all the classes in the training set, it will be saved in the smart cameras' memory (we refer to this compact representation as $\boldsymbol{D}$). In the on-line testing stage in each camera, a feature histogram $\boldsymbol{h}_i$ is extracted for the ith test image at each of the p local cameras independently, and we calculate the corresponding code (i.e., $\boldsymbol{s}_i$) for each feature histogram, using a supervised constrained non-negative matrix factorization scheme:

$$\min_{\boldsymbol{s}_i} \{\|\boldsymbol{h}_i - \boldsymbol{D}\boldsymbol{s}_i\|_F^2\}, \ \ i = (1, ..., p),$$
$$s.t. \ \ \boldsymbol{s}_i(j) \geq 0, \ j = 1, \ldots, k, \qquad \sum_{j=1}^{k} \boldsymbol{s}_i(j) = 1 \tag{3}$$

where $\boldsymbol{h}_i \in R^{m \times 1}$, $\boldsymbol{D} \in R^{m \times k}$, $\boldsymbol{s}_i \in R^{k \times 1}$ and $k << m$. k is the number of features that have been selected for the whole training set in the previous step (i.e., number of columns of $\boldsymbol{D}$), and m is the dimension of the original feature histograms (i.e., 1000 in our setting). The optimization problem in Eq. 3 is developed from the nonnegative constrained least squares (NCLS) method [26] in conjunction with the sum-to-one constraint. The objective is to minimize the least squares error:

$$\min_{\boldsymbol{s}_i} \left\| \hat{\boldsymbol{h}}_i - \hat{\boldsymbol{D}}\boldsymbol{s}_i \right\|_F^2, \ \ i = (1, ..., p),$$
$$s.t. \ \ \boldsymbol{s}_i(j) \geq 0, \ j = 1, \ldots, k, \tag{4}$$

where $\hat{\boldsymbol{h}}_i$ and $\hat{\boldsymbol{D}}$ are the augmented matrices

$$\hat{\boldsymbol{h}}_i = \begin{bmatrix} \delta \boldsymbol{h}_i \\ \boldsymbol{1} \end{bmatrix}, \ \ \hat{\boldsymbol{D}} = \begin{bmatrix} \delta \boldsymbol{D} \\ \boldsymbol{1}^T \end{bmatrix} \tag{5}$$

with δ being a small weight and $\boldsymbol{1}^T$ is a row vector of all 1s. This augmentation is used to incorporate the sum-to-one constraint. The constrained minimization problem in Eq. 4 is solved by a standard active set method [27]. This process is simple and can be done fast inside each smart camera. After finding $\boldsymbol{s}_i \in R^{k \times 1}$ in each camera ($i = 1, \ldots, p$), these low dimensional codes will be sent to the base station for performing the intended recognition task. Transmitting codes with k dimension instead of feature histograms with m dimension leads to major saving in bandwidth of the wireless network (the compression ratio: m/k, $k << m$) as well as better recognition accuracy.

3. EXPERIMENTS AND RESULTS

In this work, we validate our proposed feature encoding scheme on two multi-view recognition tasks including pedestrian recognition in surveillance video and distributed object recognition in smart camera networks.

3.1. Datasets

The Person Re-ID (PRID) dataset [28] is one of the few multi-view datasets which includes multi image frames for each pedestrian recorded from two different, static surveillance cameras. Images from these cameras contain a viewpoint change and a stark difference in illumination, background and camera characteristics. Since images are extracted from trajectories, several different poses per pedestrian are available in each camera view. It contains recorded frames of 475 person trajectories from one view and 856 from the other one, with 245 persons appearing in both views [28]. Figure 1 illustrates some sample frames of this dataset. The second dataset is the Berkeley Multi-view Wireless (*BMW*) database [10] that consists of multiple-view images of 20 landmark buildings on the Berkeley campus. We employ this dataset in order to evaluate the performance of our proposed algorithm on multi-view object recognition. It is important to note that the image quality in this database is considerably lower than many existing high-resolution databases, which is intended

Fig. 1. PRID dataset samples

Table 1. Recognition accuracy comparison on BMW dataset. Compression ratio: 2.4

Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
DFS	**99**	86	**86**	**100**	86	**93**	**98**	**94**	**91**	**76**	**96**	**99**	**86**	**100**	**86**	**100**	**100**	**100**	**100**	93	**93.55**
SPCA	94.44	**91.66**	66.66	81.94	**91.66**	88.88	93.05	91.66	73.61	65.27	76.38	83.33	72.22	93.05	80.55	79.16	90.27	93.05	83.33	100	84.51
SfM	83.33	90.27	58.33	65.27	81.94	87.50	86.11	72.22	63.88	61.11	69.44	70.83	52.77	90.27	75.00	80.55	84.72	**100**	86.11	**95.83**	77.77

to reproduce realistic imaging conditions for surveillance applications [10].

3.2. Pedestrian Recognition in Surveillance Video

In the first experiment on PRID dataset, we consider 30 different frames for each pedestrian in each camera and we randomly choose 20 persons in the dataset for the recognition task. Hence, there are 1200 images for which we randomly pick half of it as the training set and the rest as the testing set. The dimension of the original feature histograms is 1000. Figure 2 shows the recognition accuracy versus the compression rate using the proposed DFS method. In this figure, we can observe that with compression rate of 2.94 (i.e., features with dimension of 340), the accuracy is slightly better than using the original features. The reason is that our feature selection scheme omits those features which are closer to the features from other classes than the features in their own class. We define the compression ratio as dimension of the original features divided by dimension of the encoded features (i.e., k in Eq. 3). For instance, in Figure 2 at the marked point on the curve with compression ratio of 2.94, the feature dimension is equal to $1000/2.94 = 340$. Figure 3 illustrates the recognition accuracy for classification of all the 20 persons in the experiment with 340-D feature histograms ($\lambda = 17$ in Eq. 2). It is worth noting that the recognition task in this experiment is different from person re-identification task which is based on image matching and retrieval.

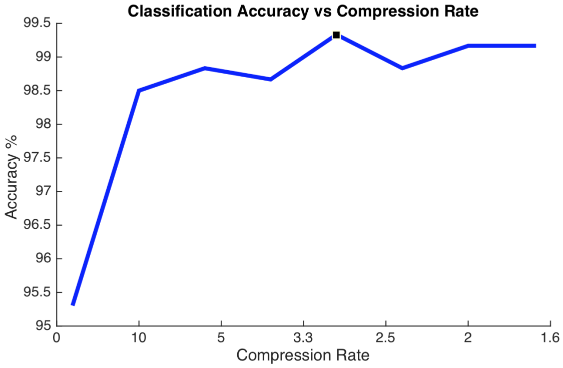
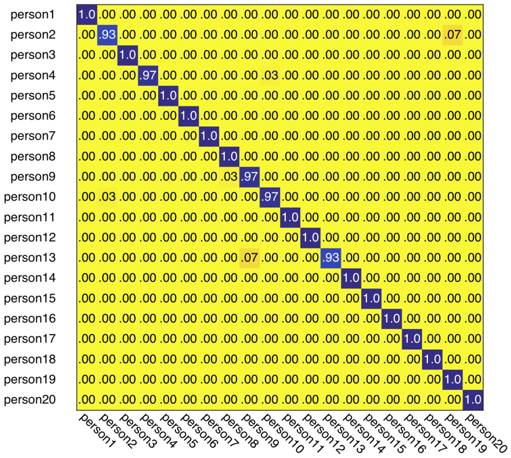


Fig. 2. Recognition accuracy for different dimensions of the feature histograms using the DFS method

3.3. Building Recognition in BMW Dataset

In the second experiment (on BMW), there exist 16 different vantage points, and at each vantage point, images are taken by five cameras simultaneously, thereby summing to 80 images per category. In this section, we compare the recognition accuracy of DFS method with two other existing works. Table 1 demonstrates the classification accuracy of different methods based on Sparse PCA (SPCA) [11] and Structure from Motion (SfM) [13]. To have a fair comparison, we set up the same experimental environment as the other two works. In fact, we only considered 8 images (even vantage points of camera #2) from each object for training and the rest of images from other cameras for testing (and compression ratio: 2.4). For most of the object categories our proposed method, outperforms SPCA and SfM based approaches. One important reason for outperforming the proposed DSF method compared to other two methods is that in contrast to SPCA and SfM methods, the physical interpretation of the reduced space is preserved during the dimensionality reduction procedure which is critical in the recognition task.



	person1	person2	person3	person4	person5	person6	person7	person8	person9	person10	person11	person12	person13	person14	person15	person16	person17	person18	person19	person20
person1	1.0	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
person2	.00	.93	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.07	.00
person3	.00	.00	1.0	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
person4	.00	.00	.00	.97	.00	.00	.00	.00	.00	.03	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
person5	.00	.00	.00	.00	1.0	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
person6	.00	.00	.00	.00	.00	1.0	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
person7	.00	.00	.00	.00	.00	.00	1.0	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
person8	.00	.00	.00	.00	.00	.00	.00	1.0	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
person9	.00	.00	.00	.00	.00	.00	.00	.03	.97	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
person10	.00	.03	.00	.00	.00	.00	.00	.00	.00	.97	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
person11	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	1.0	.00	.00	.00	.00	.00	.00	.00	.00	.00
person12	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	1.0	.00	.00	.00	.00	.00	.00	.00	.00
person13	.00	.00	.00	.00	.00	.00	.00	.00	.07	.00	.00	.00	.93	.00	.00	.00	.00	.00	.00	.00
person14	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	1.0	.00	.00	.00	.00	.00	.00
person15	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	1.0	.00	.00	.00	.00	.00
person16	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	1.0	.00	.00	.00	.00
person17	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	1.0	.00	.00	.00
person18	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	1.0	.00	.00
person19	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	1.0	.00
person20	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	1.0

Fig. 3. Confusion matrix of pedestrian recognition via DFS on PRID dataset using 340-D features.

4. CONCLUSION

In this paper, we introduced a probabilistic encoding approach based on divergence of the probability distributions of the visual features in limited bandwidth distributed camera networks. The performance of the proposed approach was discussed in two surveillance recognition tasks. The proposed DFS approach is applicable to the variety of distributed computer vision tasks based on transmission of the visual features in a network (e.g., cross view action recognition, person reidentification, etc.).

5. REFERENCES

[1] Alessandro Enrico Redondi, Luca Baroffio, Matteo Cesana, and Marco Tagliasacchi, "Cooperative features extraction in visual sensor networks: a game-theoretic approach," in *2015-Proceedings of the 9th International Conference on Distributed Smart Camera*. ACM, 2015.

[2] C Mario Christoudias, Raquel Urtasun, and Trevor Darrell, "Unsupervised feature selection via distributed coding for multi-view object recognition," in *Computer Vision and Pattern Recognition. CVPR 2008. IEEE Conference on*. IEEE, 2008.

[3] Vittorio Ferrari, Tinne Tuytelaars, and Luc Van Gool, "Integrating multiple model views for object recognition," in *Computer Vision and Pattern Recognition. Proceedings of the IEEE Computer Society Conference on*. IEEE, CVPR-2004.

[4] Alireza Rahimpour, Ali Taalimi, Jiajia Luo, and Hairong Qi, "Distributed object recognition in smart camera networks," in *IEEE International Conference on Image Processing, Phoenix, Arizona, USA*. IEEE, 2016.

[5] Hossein Rahmani and Ajmal Mian, "Learning a non-linear knowledge transfer model for cross-view action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, CVPR-2015.

[6] Jingjing Zheng, Zhuolin Jiang, and Rama Chellappa, "Cross-view action recognition via transferable dictionary learning," *IEEE Transactions on Image Processing*, TIP-2016.

[7] Giuseppe Lisanti, Iacopo Masi, Andrew D Bagdanov, and Alberto Del Bimbo, "Person re-identification by iterative re-weighted sparse ranking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, TPAMI-2015.

[8] Ejaz Ahmed, Michael Jones, and Tim K Marks, "An improved deep learning architecture for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, CVPR-2015.

[9] Kaushik Mitra, Ashok Veeraraghavan, Aswin C Sankaranarayanan, and Richard G Baraniuk, "Toward compressive camera networks," *Computer*, 2014.

[10] Nikhil Naikal, Allen Y Yang, and S Shankar Sastry, "Towards an efficient distributed object recognition system in wireless smart camera networks," in *Information Fusion (FUSION),13th Conference on*. IEEE, 2010.

[11] Nikhil Naikal, Allen Y Yang, and S Shankar Sastry, "Informative feature selection for object recognition via sparse pca," in *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2011.

[12] Chuohao Yeo, Parvez Ahammad, and Kannan Ramchandran, "Rate-efficient visual correspondences using random projections," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008.

[13] Panu Turcot and D Lowe, "Better matching with fewer features: The selection of useful features in large database recognition problems," in *ICCV workshop on emergent issues in large amounts of visual data (WS-LAVD)*, 2009, vol. 4.

[14] Jingming Dong, Nikolaos Karianakis, Damek Davis, Joshua Hernandez, and Balzer, "Multi-view feature engineering and learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[15] David G Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. Ieee, 1999, vol. 2, pp. 1150–1157.

[16] John J Lee, "Libpmk: A pyramid match toolkit," 2008.

[17] Zhuolin Jiang, Zhe Lin, and Larry S Davis, "Label consistent k-svd: Learning a discriminative dictionary for recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, TPAMI-2013.

[18] Shu Kong and Donghui Wang, "A dictionary learning approach for classification: separating the particularity and the commonality," in *Computer Vision–ECCV 2012*. Springer, 2012.

[19] Zhuolin Jiang, Guangxiao Zhang, and Larry S Davis, "Submodular dictionary learning for sparse coding," in *Computer Vision and Pattern Recognition, 2012 IEEE Conference on*. IEEE, CVPR-2012.

[20] Mostafa Rahmani and George Atia, "Randomized robust subspace recovery for high dimensional data matrices," *arXiv preprint arXiv:1505.05901*, 2015.

[21] Mostafa Rahmani and George Atia, "A subspace learning approach for high dimensional matrix decomposition with efficient column/row sampling," in *Proceedings of The 33rd International Conference on Machine Learning*, 2016, pp. 1206–1214.

[22] Ernie Esser, Michael Moller, Stanley Osher, Guillermo Sapiro, and Jack Xin, "A convex model for nonnegative matrix factorization and dimensionality reduction on physical space," *IEEE Transactions on Image Processing, TIP*, vol. 21, no. 7, pp. 3239–3252, 2012.

[23] Solomon Kullback and Richard A Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[24] Rick Chartrand and Valentina Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, pp. 035020, 2008.

[25] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[26] Chein-I Chang and Daniel C Heinz, "Constrained subpixel target detection for remotely sensed imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 38, no. 3, pp. 1144–1159, 2000.

[27] Rasmus Bro and Sijmen De Jong, "A fast non-negativity-constrained least squares algorithm," *Journal of chemometrics*, vol. 11, no. 5, pp. 393–401, 1997.

[28] Martin Hirzer, Csaba Beleznai, Peter M. Roth, and Horst Bischof, "Person re-identification by descriptive and discriminative classification," in *Proc. Scandinavian Conference on Image Analysis (SCIA)*, 2011.