ENDOGENOUS CONVOLUTIONAL SPARSE REPRESENTATIONS FOR TRANSLATION INVARIANT IMAGE SUBSPACE MODELS

Brendt Wohlberg

Los Alamos National Laboratory
Los Alamos, NM 87545, USA

ABSTRACT

Subspace models for image data sets, constructed by computing sparse representations of each image with respect to other images in the set, have been found to perform very well in a variety of applications, including clustering and classification problems. One of the limitations of these methods, however, is that the subspace representation is unable to directly model the effects of non-linear transformations such as translation, rotation, and dilation that frequently occur in practice. In this paper it is shown that the properties of convolutional sparse representations can be exploited to make these methods translation invariant, thereby simplifying or eliminating the alignment pre-processing task. The potential of the proposed approach is demonstrated in two diverse applications: image clustering and video background modeling.

***Index Terms*—** Subspace Models, Translation Invariance, Convolutional Sparse Representation

1. INTRODUCTION

Subspace models, associated with Principal Component Analysis (PCA) or the Karhunen-Loève Transform (KLT), have a long history in signal and image processing. More recently, models based on a union of subspaces have found application in a wide variety of problems, examples including

- the Sparse Representation-based Classification (SRC) method for face recognition [1],
- the Sparse Subspace Clustering (SSC) method for motion segmentation and face clustering [2], and
- subspace estimation for video background modeling [3, 4, 5].

A common approach of many of these methods is the use of *endogenous* sparse representations [6]. A sparse representation of a signal $\mathbf{s}$ as a sparse linear combination $\mathbf{x}$ on a dictionary matrix D, computed via an optimization such as $\arg\min_{\mathbf{x}} \frac{1}{2} \left\| D\mathbf{x} - \mathbf{s} \right\|_2^2 + \lambda \left\| \mathbf{x} \right\|_1$ or, in the Multiple Measurement Vector (MMV) case in which the representations are jointly computed for multiple signals $\mathbf{s}_n$ concatenated as columns of matrix S,

$$\arg\min_X \frac{1}{2} \left\| DX - S \right\|_2^2 + \lambda \left\| X \right\|_1 \ . \tag{1}$$

The dictionary D is usually analytically defined (e.g. an overcomplete wavelet basis) or learned offline from a training data set; an endogenous sparse representation, in contrast, represents a member of a data set using the remaining members of the set as a dictionary. In the MMV case, it is essential to constrain the representation X so that a signal is not trivially represented in terms of its occurrence as an element of the dictionary. Defining $\mathcal{E}_n = \{m \mid \mathbf{d}_m = \mathbf{s}_n\}$ as the set of indices m for signal $\mathbf{s}_n$ such that a dictionary element $\mathbf{d}_m$ is the same datum as that signal, Eq. (1) must be solved subject to the constraint $X_{m,n} = 0 \, \forall m \in \mathcal{E}_n$.

One of the limitations of subspace models is that they are, in general[1], not able to represent the non-linear image transformations such as translation, rotation, and dilation that are commonly encountered in practical applications. As a result, these methods usually assume that the image data have been aligned in a pre-processing stage that is often difficult, and usually suboptimal with respect to the subspace model that follows. It is possible to perform automatic alignment by wrapping the subspace modeling within an iterative estimate of the non-linear transforms required for alignment [8], but this requires a complex and expensive optimization.

The contribution of the present paper is the observation that translation invariance, at least, can be directly integrated into these techniques by replacing the sparse representations with *convolutional sparse representations* [9]. Such a representation replaces the usual sparse representation of a signal $\mathbf{s}$ as a sparse linear combination $\mathbf{x}$ on a dictionary matrix D by the sum of convolutions of a set of dictionary filters $\{\mathbf{d}_m\}$ with a set of sparse coefficient maps $\{\mathbf{x}_m\}$, computed via a problem such as

$$\arg\min_{\{\mathbf{x}_m\}} \frac{1}{2} \left\| \sum_m \mathbf{d}_m * \mathbf{x}_m - \mathbf{s} \right\|_2^2 + \lambda \sum_m \left\| \mathbf{x}_m \right\|_1 \ . \tag{2}$$

This research was supported by the U.S. Department of Energy through the LANL/LDRD Program.

[1] Small transformations can sometimes be handled via tangent methods [7].

Convolutional sparse representations usually use dictionary filters consisting of small image patches constructed via a dictionary learning process, but in the proposed endogenous application the filters are themselves images extracted from the same image set as the target signal $\mathbf{s}$. The convolutional nature of the representation allows the position of non-zeros in the coefficient maps to determine the translational alignment between the selected dictionary images $\{\mathbf{d}_m\}$ and the target image $\mathbf{s}$ – in the ideal case, each map $\{\mathbf{x}_m\}$ has at most one non-zero entry – with $\mathbf{d}_m * \mathbf{x}_m$ representing the contribution of the aligned $\mathbf{d}_m$ to the representation of $\mathbf{s}$.

Convolutional sparse representations have not previously been proposed for such applications, presumably due to the computational expense of standard spatial-domain algorithms for solving Eq. (2), making use of filters with large support as required here completely infeasible. The recent emergence of Discrete Fourier Transform (DFT) domain algorithms [10, 11] for solving Eq. (2) enable the proposed applications since convolution implemented via the DFT has a computational cost that depends only on the size of the images. A further essential development is the modification of the original DFT domain algorithm [10] with $\mathcal{O}(M^3)$ computational cost in the number of filters M to one with $\mathcal{O}(M)$ cost [11]. It should be emphasized, however, that computing these representations remains very computationally expensive relative to standard sparse representations; while the $\mathcal{O}(M)$ algorithm makes use of these techniques feasible for research use, further computational improvements, e.g. via the use of multi-resolution techniques, are necessary for many practical applications.

2. CONVOLUTIONAL SPARSE REPRESENTATIONS

A common approach in methods based on endogenous sparse representations is to include an additive sparse component to represent outliers that are not well represented within the union of subspaces (e.g. [1, 2, 3]). Extending Eq. (2) in this way leads to the optimization

$$\underset{\{\mathbf{x}_m\},\mathbf{u}}{\arg\min} \frac{1}{2} \left\| \sum_m \mathbf{d}_m * \mathbf{x}_m + \mathbf{u} - \mathbf{s} \right\|_2^2 + \lambda \sum_m \|\mathbf{x}_m\|_1 + \mu \|\mathbf{u}\|_1 \;, \quad (3)$$

where $\{\mathbf{d}_m\}$, $\{\mathbf{x}_m\}$, and $\mathbf{s}$ are as before, and $\mathbf{u}$ represents outliers that cannot be efficiently represented via the convolutional sparse representation. In the MMV context this problem becomes

$$\underset{\{\mathbf{x}_{m,n}\},\{\mathbf{u}_n\}}{\arg\min} \frac{1}{2} \sum_n \left\| \sum_m \mathbf{d}_m * \mathbf{x}_{m,n} + \mathbf{u}_n - \mathbf{s}_n \right\|_2^2 + \\ \lambda \sum_n \sum_m \|\mathbf{x}_{m,n}\|_1 + \mu \sum_n \|\mathbf{u}_n\|_1 \;, \quad (4)$$

where, as before, in endogenous application, the constraint $\mathbf{x}_{m,n} = 0 \forall m \in \mathcal{E}_n$ is necessary to avoid trivial solutions. Since this problem is separable in index n, further mathematical development will address the Single Measurement Vector (SMV) case for notational simplicity, but it is emphasized that these results are trivial to extend to the MMV case.

The necessary constraints can be included in Eq. (3) via the use of indicator functions [12], leading to

$$\underset{\{\mathbf{x}_m\},\mathbf{u}}{\arg\min} \frac{1}{2} \left\| \sum_m \mathbf{d}_m * \mathbf{x}_m + \mathbf{u} - \mathbf{s} \right\|_2^2 + \lambda \sum_m \|\mathbf{x}_m\|_1 + \\ \mu \|\mathbf{u}\|_1 + \sum_{m \in \mathcal{E}} \iota(\mathbf{x}_m) \;, \quad (5)$$

where $\iota(\cdot)$ is zero if its argument is a zero vector and infinite otherwise. Rewriting in the appropriate form for solving within the Alternating Direction Method of Multipliers (ADMM) [13, 14] framework gives

$$\underset{\{\mathbf{x}_m\},\{\mathbf{y}_m\},\mathbf{u}}{\arg\min} \frac{1}{2} \left\| \sum_m \mathbf{d}_m * \mathbf{x}_m + \mathbf{u} - \mathbf{s} \right\|_2^2 + \lambda \sum_m \|\mathbf{y}_m\|_1 + \\ \mu \|\mathbf{u}\|_1 + \sum_{m \in \mathcal{E}} \iota(\mathbf{y}_m) \text{ s.t. } \mathbf{x}_m - \mathbf{y}_m = 0 \;, \quad (6)$$

for which the ADMM iterations are

$$\{\mathbf{x}_m\}^{(k+1)} = \underset{\{\mathbf{x}_m\}}{\arg\min} \frac{1}{2} \left\| \sum_m \mathbf{d}_m * \mathbf{x}_m - (\mathbf{s} - \mathbf{u}^{(k)}) \right\|_2^2 + \\ \frac{\rho}{2} \sum_m \left\| \mathbf{x}_m - (\mathbf{y}_m^{(k)} - \mathbf{z}_m^{(k)}) \right\|_2^2 \quad (7)$$

$$\{\mathbf{y}_m\}^{(k+1)} = \underset{\{\mathbf{y}_m\}}{\arg\min} \lambda \sum_m \|\mathbf{y}_m\|_1 + \sum_{m \in \mathcal{E}} \iota(\mathbf{y}_m) + \\ \frac{\rho}{2} \sum_m \left\| \mathbf{y}_m - (\mathbf{x}_m^{(k+1)} + \mathbf{z}_m^{(k)}) \right\|_2^2 \quad (8)$$

$$\mathbf{u}^{(k+1)} = \underset{\mathbf{u}}{\arg\min} \frac{1}{2} \left\| \mathbf{u} - \left(\mathbf{s} - \sum_m \mathbf{d}_m * \mathbf{x}_m^{(k+1)} \right) \right\|_2^2 + \\ \mu \|\mathbf{u}\|_1 \quad (9)$$

$$\mathbf{z}_m^{(k+1)} = \mathbf{z}_m^{(k)} + \mathbf{x}_m^{(k+1)} - \mathbf{y}_m^{(k+1)} \;. \quad (10)$$

Subproblems Eq. (8) and Eq. (9) can be solved via shrinkage/soft thresholding

$$\mathcal{S}_\gamma(\mathbf{u}) = \operatorname{sign}(\mathbf{u}) \odot \max(0, |\mathbf{u}| - \gamma) \quad (11)$$

as

$$\mathbf{y}_m^{(k+1)} = \begin{cases} \mathcal{S}_{\lambda/\rho}\left(\mathbf{x}_m^{(k+1)} + \mathbf{z}_m^{(k)}\right) & \text{if } m \notin \mathcal{E} \\ 0 & \text{if } m \in \mathcal{E} \;, \end{cases} \quad (12)$$

$$\mathbf{u}^{(k+1)} = \mathcal{S}_\mu \left(\mathbf{s} - \sum_m \mathbf{d}_m * \mathbf{x}_m^{(k+1)} \right) \;. \quad (13)$$

The only computationally expensive step is subproblem Eq. (7), which can be solved effectively in the DFT domain by applying the Sherman-Morrison formula [11].

3. CONSTRAINTS

When the problem is solved together with constraints (i.e. when $\mathcal{E} \neq \emptyset$), convergence can be very slow since the constraint on $\mathbf{x}_m$ demands a major perturbation in the solution that is only enforced via the $\mathbf{y}_m$ update. This issue can be addressed by the introduction of a weighted ℓ^2 term in Eq. (5)

$$\operatorname*{arg\,min}_{\{\mathbf{x}_m\},\mathbf{u}} \frac{1}{2} \left\| \sum_m \mathbf{d}_m * \mathbf{x}_m + \mathbf{u} - \mathbf{s} \right\|_2^2 + \lambda \sum_m \|\mathbf{x}_m\|_1 + \sum_m \frac{w_m}{2} \|\mathbf{x}_m\|_2^2 + \mu \|\mathbf{u}\|_1 + \sum_{m \in \mathcal{E}} \iota(\mathbf{x}_m) \;, \qquad (14)$$

where w_m is zero for $m \notin \mathcal{E}$ and very large for $m \in \mathcal{E}$. The only corresponding modification to the ADMM algorithm is in the $\mathbf{x}_m$ subproblem, which becomes

$$\operatorname*{arg\,min}_{\{\mathbf{x}_m\}} \frac{1}{2} \left\| \sum_m \mathbf{d}_m * \mathbf{x}_m - (\mathbf{s} - \mathbf{u}^{(k)}) \right\|_2^2 + \sum_m \frac{w_m}{2} \|\mathbf{x}_m\|_2^2 + \frac{\rho}{2} \sum_m \left\| \mathbf{x}_m - (\mathbf{y}_m^{(k)} - \mathbf{z}_m^{(k)}) \right\|_2^2 \;. \qquad (15)$$

This modified problem can also be solved in the DFT domain in $\mathcal{O}(M)$ time (with a small increase in the constant factor) via a minor modification of the Sherman-Morrison formula-based solution for Eq. (7) [11].

4. BOUNDARY ISSUES

We expect the convolutional representation to be able to align overlapping sets of images, but a good approximation of the target image will usually only be achievable in the central region common to all images. The boundary region exterior to this common central region cannot be well represented as either a convolutional sparse representation or a sparse additive component, and will therefore substantially perturb the solution if not explicitly considered in the model. The obvious approach is to modify the data fidelity term of Eq. (3) to include an operator P projecting out only the common central region

$$\frac{1}{2} \left\| P \left(\sum_m \mathbf{d}_m * \mathbf{x}_m + \mathbf{u} - \mathbf{s} \right) \right\|_2^2 \;, \qquad (16)$$

but such a projection does not have a simple representation in the DFT domain, so that such a modification would preclude the use of the efficient DFT domain method [11] for solving Eq. (7). A simple and effective alternative is to introduce a spatially-varying weighting in the ℓ^1 norm of $\mathbf{u}$ so that there is no penalty on $\mathbf{u}$ in the boundary region; the result is that the central region of $\mathbf{u}$ behaves as before, and the boundary region of $\mathbf{u}$ adapts without penalty to represent any error between the signal and its convolutional sparse representation. If $\mathbf{b}$ is a mask vector that is zero on the boundary region and unity on the central region, the modified $\mathbf{u}$ update can be expressed as

$$\mathbf{u}^{(k+1)} = \mathcal{S}_{\mu,\mathbf{b}} \left(\mathbf{s} - \sum_m \mathbf{d}_m * \mathbf{x}_m^{(k+1)} \right) \;, \qquad (17)$$

where

$$\mathcal{S}_{\gamma,\mathbf{b}}(\mathbf{u}) = \operatorname{sign}(\mathbf{u}) \odot \max(0, |\mathbf{u}| - \gamma \mathbf{b}) \;. \qquad (18)$$

5. RESULTS

The utility of the proposed method is demonstrated in two distinct applications. In all cases algorithm parameters were manually selected for best performance.

The first of these is a face clustering problem using images from the Extended Yale B dataset [15, 16]. A set of manually aligned and cropped images derived from this set was used for a demonstration of SSC performance [2]. For the experiments reported here, five aligned test sets of 45 images each were constructed by randomly selecting (after removal of unsuitable examples) a set of 15 aligned face images each for three distinct subjects. A set of five corresponding misaligned dataset was constructed by first selecting the corresponding images from the original unaligned image set (from which the aligned set was also derived) and then cropping 336×384 subimages with alignment to the centers of the faces randomly varied by 20 pixels in each direction.

The performance of SSC and Convolutional SSC was compared on both aligned and misaligned data sets. For standard SSC spectral clustering was performed, as advocated in [2], on the the coefficient matrix X obtained by solving

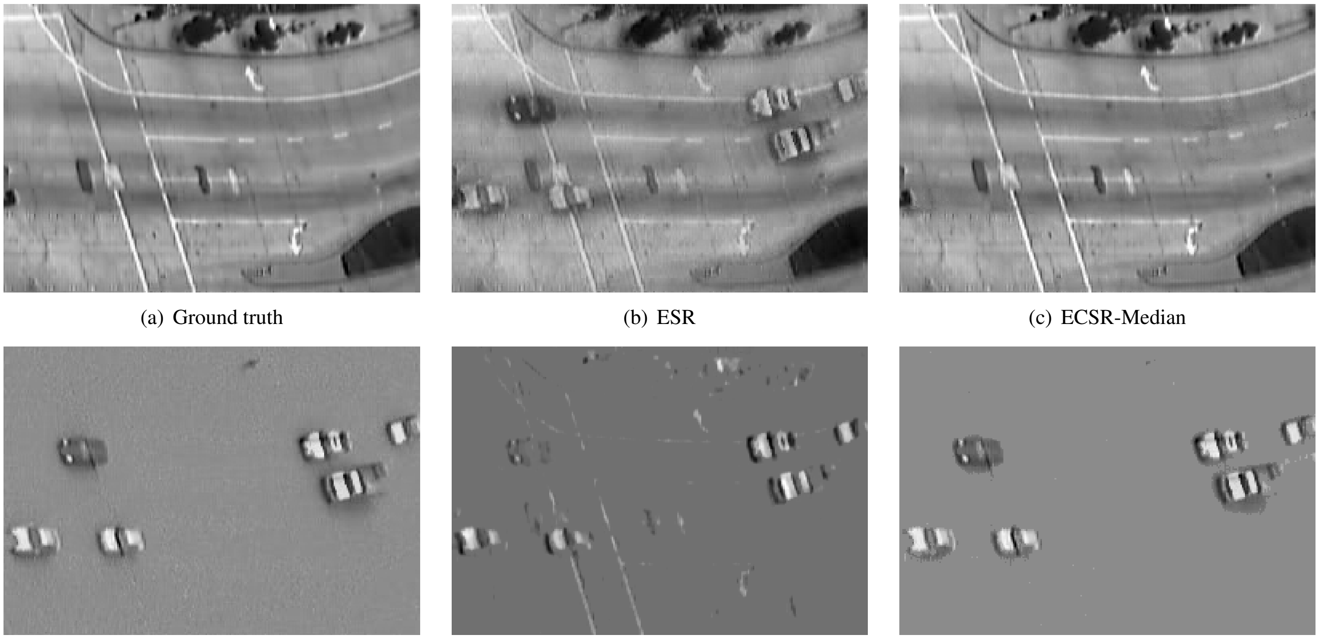
$$\operatorname*{arg\,min}_{X,U} \frac{1}{2} \|DX + U - S\|_2^2 + \lambda \|X\|_1 + \mu \|U\|_1 \qquad (19)$$

with $S = D$ and $\mathcal{E}_n = \{n\}$, with $\lambda = 1$ and $\mu = 0.1$ for both data sets. In the Convolutional SSC case, a set of coefficient maps $\mathbf{x}_{m,n}$ was obtained by solving Eq. (4) with $\mathbf{s}_n = \mathbf{d}_n$ and $\mathcal{E}_n = \{n\}$ with $\lambda = 50$ and $\mu = 0.1$ for the aligned data set and $\lambda = 50$ and $\mu = 0.05$ for the unaligned data set. A coefficient matrix X independent of the spatial location of non-zero coefficient was constructed from these maps by setting each entry $X_{m,n}$ as the spatial sum of the absolute values of the corresponding $\mathbf{x}_{m,n}$, i.e. $X_{m,n} = \mathbb{1}^T |\mathbf{x}_{m,n}|$ where $\mathbb{1}$ denotes a vector with unit entries, and $|\cdot|$ is applied elementwise. Spectral clustering based on X was performed as for the SSC case. Clustering errors are compared in Table 1; note that Convolutional SSC performs well in both cases, but standard SSC performs very poorly when the images are misaligned.

	SSC					CSSC				
Test Set	1	2	3	4	5	1	2	3	4	5
Aligned	0	0	1	2	0	3	0	0	5	2
Misaligned	12	17	14	8	5	2	0	0	2	0

Table 1. Clustering errors (number of faces out of a total of 45 assigned to the wrong cluster) for SSC and Convolutional SSC applied to the aligned and misaligned datasets for the same faces. The expected value for uniform random cluster assignment is 25.65 errors, and the maximum possible is 30.

The second demonstration applies the proposed method to align a set of frames for video background modeling in a simulated moving-camera video sequence. This represents



(a) Ground truth

(b) ESR

(c) ECSR-Median

(d) Ground truth

(e) ESR

(f) ECSR-Median

Fig. 1. Background (first row) and foreground (second row) estimates for frame 15 from the fast-panning test video sequence, computed via the endogenous sparse representation method of [3] (ESR) and the proposed method with median filtering post-processing (ECSR-Median).

	RPCA	ESR	ECSR	ECSR-R	ECSR-M
Back	4.6dB	7.6dB	10.9dB	14.7dB	19.3dB
Fore	-0.5dB	1.9dB	3.6dB	9.6dB	14.2dB

Table 2. Comparison of background/foreground separation performance, measured as SNR against ground truth, of RPCA, the endogenous sparse representation method of [3] (ESR), the proposed method without post-processing (ECSR), and the proposed method with RPCA and median filtering post-processing (ECSR-R and ECSR-M respectively).

a difficult background modeling problem that cannot be directly addressed using Robust Principal Component Analysis (RPCA) [17, 18] methods. The test sequence was constructed from the Lankershim Boulevard Dataset [19, camera 4, 8:45–9:00 AM] by moving a 240×320 pixel cropping window within the original sequence at a rate of 3 pixels/frame, selecting a total of 30 consecutive frames. Since the panning view is simulated, the RPCA background/foreground separation result in the original stationary view sequence can be used to construct an approximate ground truth.

Problem Eq. (4) was solved with both $\{\mathbf{d}_m\}$ and $\{\mathbf{s}_n\}$ set to the test video sequence (with $\lambda = 50$ and $\mu = 0.06$), the resulting $\sum_m \mathbf{d}_m * \mathbf{x}_{m,n}$ representing the background estimate for frame n, and $\mathbf{u}_n$ representing the foreground estimate. Examination of the set of $\mathbf{d}_m * \mathbf{x}_{m,n}$ contributing to the background estimate for frame n reveals that the convolutional sparse representation achieves good performance in aligning these estimates from different frames, but their sum contains artifacts from the vehicles that are in different relative locations in each $\{\mathbf{d}_m\}$. Two post-processing strategies were applied in order to exploit the robust frame alignment to obtain improved results. Both of these operate on the set $\mathbf{d}_m * \mathbf{x}_{m,n}$, which can be considered as a set of distinct background estimates of frame n based on the filters m for which $\mathbf{x}_{m,n}$ contains a coefficient of significant absolute value. The first strategy simply normalizes each member of the set and applies median filtering in the temporal direction, and the second applies RPCA independently to the set of estimates for each frame n. Once the background is re-estimated, a foreground re-estimate is obtained by subtracting from the original sequence. The performance of these methods with the respect to the approximate ground truth is compared in Table 2, and selected examples are displayed in Fig. 1.

6. CONCLUSION

A variety of recent algorithms for subspace modeling, clustering, and classification of images can be made invariant to image translation by replacing the sparse representations in these methods with their convolutional equivalents. A novel ADMM-based algorithm has been developed for solving the endogenous convolutional sparse representation problem. The experimental results in two simple problems presented here provide initial evidence of the potential of the proposed method for translation invariant subspace modeling.

7. REFERENCES

[1] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009. doi:10.1109/tpami.2008.79

[2] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013. doi:10.1109/tpami.2013.57

[3] B. Wohlberg, R. Chartrand, and J. Theiler, "Local principal component pursuit for nonlinear datasets," in *Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc. (ICASSP)*, Mar. 2012, pp. 3925–3928. doi:10.1109/ICASSP.2012.6288776

[4] J. He, D. Zhang, L. Balzano, and T. Tao, "Iterative online subspace learning for robust image alignment," in *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Apr. 2013, pp. 1–8. doi:10.1109/fg.2013.6553759

[5] ——, "Iterative Grassmannian optimization for robust image alignment," 2013, arXiv:1306.0404. [Online]. Available: http://arxiv.org/abs/1306.0404

[6] E. L. Dyer, A. C. Sankaranarayanan, and R. G. Baraniuk, "Greedy feature selection for subspace clustering," *Journal of Machine Learning Research*, vol. 14, pp. 2487–2517, 2013. [Online]. Available: http://jmlr.org/papers/v14/dyer13a.html

[7] P. Y. Simard, Y. A. LeCun, J. S. Denker, and B. Victorri, "Transformation invariance in pattern recognition: Tangent distance and propagation," *International Journal of Imaging Systems and Technology*, vol. 11, no. 3, pp. 181–197, 2000. doi:10.1002/1098-1098(2000)11:3<181::aid-ima1003>3.0.co;2-e

[8] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," in *Proc. IEEE Conf. Comp. Vis. Patt. Recog. (CVPR)*, Jun. 2010, pp. 763–770. doi:10.1109/cvpr.2010.5540138

[9] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Conf. Comp. Vis. Patt. Recog. (CVPR)*, Jun. 2010, pp. 2528–2535. doi:10.1109/cvpr.2010.5539957

[10] H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," in *Proc. IEEE Conf. Comp. Vis. Patt. Recog. (CVPR)*, Jun. 2013, pp. 391–398. doi:10.1109/CVPR.2013.57

[11] B. Wohlberg, "Efficient convolutional sparse coding," in *Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc. (ICASSP)*, May 2014, pp. 7173–7177. doi:10.1109/ICASSP.2014.6854992

[12] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An Augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 681–695, 2011. doi:10.1109/tip.2010.2076294

[13] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. 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, 2010. doi:10.1561/2200000016

[14] J. Eckstein, "Augmented Lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results," Rutgers Center for Operations Research, Rutgers University, Rutcor Research Report RRR 32-2012, Dec. 2012. [Online]. Available: http://rutcor.rutgers.edu/pub/rrr/reports2012/32_2012.pdf

[15] "The Extended Yale Face Database B." [Online]. Available: http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html

[16] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001. doi:10.1109/34.927464

[17] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Adv. in Neural Inf. Proc. Sys. (NIPS) 22*, 2009, pp. 2080–2088.

[18] T. Bouwmans and E.-H. Zahzah, "Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance," *Computer Vision and Image Understanding*, 2014, (Special Issue on Background Models Challenge). doi:10.1016/j.cviu.2013.11.009

[19] "Lankershim Boulevard dataset," U.S. Department of Transportation Publication FHWA-HRT-07-029, Jan. 2007, data available from http://ngsim-community.org/.

Addendum

The algorithm outlined in Sec. 2 for solving the convolutional sparse coding problem with an additive term, Eq. (3), can take a large number of iterations to converge, and is rather sensitive to suitable choice of parameters λ, μ, and ρ. Subsequent to submission of the final version of this manuscript for inclusion in the official proceedings, it was found that a more effective solution to Eq. (3) is to map it to the standard convolutional sparse coding problem of Eq. (2) (which can be solved as described in [11]), by absorbing the additive term into the sum of convolutions with an impulse filter as the corresponding dictionary entry.