Surveillance Video Segmentation Using Sparse Recovery from Compressive Measurements

Gauri Jagatap

April 13, 2015

1 Introduction

Video data transmitted by surveillance cameras is generally processed to detect moving objects automatically. The video generally consists of a moving object that covers a small fraction of a video frame and majority of the frame is spanned by the background. If each frame is vectorized, and these vectors are concatenated, it is referred to this as the video volume. The video volume can be split into the background and the moving objects (background separation). An intuitive method to separate out the background is by using the fact that the background being stationary, will form the low rank part of the video volume. On the other hand, the moving objects constitute the sparse component. This decomposition is done using a low rank and sparse decomposition of the video volume.

The speed of this processing however, is slowed down by the abundance of data collected, which mostly consists of spells of inactivity. Compressive sensing is a technique used to acquire video data in a different basis, instead of the usual spatial basis, like Fourier or Wavelet, that involves acquiring a small fraction (up to 50% in this paper, lower fractions can be used for larger number of frames or higher resolution data) of the data that would have been acquired in the spatial basis. This method works when the given data is sparse in this basis, which validates the low sampling rate (less than Nyquist).

Sparse signal recovery from such compressive measurements is a process of minimizing the nuclear norm or l_0 norm of the signal (in this case, video volume) in the transformed basis (which ensures sparsity in the transformed basis). The whole video volume can be recovered from this norm minimization and further background separation techniques can be employed to separate out the moving objects.

However, one can formulate the minimization problem such that the background and moving objects are separated out during the recovery. Moreover, one can choose tight wavelet transforms, which have specific properties that help simplify the minimization problem significantly.

Section 2 discusses relevant related work that inspired us to work on this topic. Section 3 is divided into three subsections. Subsection 3.1, describes a simple low rank and sparse decomposition method, that can be used on on the video volume for background separation. Subsection 3.2 describes a standard sparse video volume recovery from compressive measurements.Subsection 3.3 combines these two methods to achieve background separation through sparse recovery. Section 4 covers simulations and results of employing these three methods on 2 different sets of video frames. Section 5, describes the suggested modifications to improve accuracy of the algorithm developed in subsection 3.3.

2 Related Work

Sparse and low rank matrix decompositions have been implemented using alternating directions augmented Lagrangian methods[YY09] and other convex optimization methods[Can+11] has been used popularly to implement background subtraction in problems like the Robust Batch Alignment of Images [Pen+10]. Such procedures have been increasingly employed in medical image processing problems [Mah+14].

Compressive sensing techniques[CRT06] have been utilized for background separation in [Cev+08] in which some background pixel values are known at the time of initialization. This work is mostly inspired by the work presented in [JDS13], with certain modifications in the way the compressive measurements are stored and the basis used for minimization.

3 Problem Formulation

3.1 Model 1: Low Rank and Sparse Decomposition

Let $X \in \mathbb{R}^{n\times J}$ represent the video volume where

$$X = [x_1\, x_2 \ldots\, x_J] \tag{1}$$

and $x_j \in \mathbb{R}^{n\times 1}$ are vectorized frames ($j = 1, 2...J$ is the index for frames with a total J frames). It can be split into a low rank X_1 and sparse component X_2 where

$$X = X_1 + X_2 \tag{2}$$

Given a video volume X, its decomposition into a low rank and sparse component involves the following minimization

$$\min_{X_1, X_2} ||X_1||_* + \mu ||X_2||_1 \tag{3}$$

$$s.t. \, X_1 + X_2 = X \tag{4}$$

The corresponding Lagrangian takes the form

$$L(X_1, X_2) = ||X_1||_* + \mu ||X_2||_1 - \langle \lambda, X_1 + X_2 - X \rangle + \frac{\beta}{2} ||X_1 + X_2 - X||_F^2 \tag{5}$$

Taking a gradient with respect to X_1 and X_2,

$$\nabla_{X_1} L = \delta ||X_1||_* - \lambda + \beta(X_1 + X_2 - X) = 0 \tag{6}$$

$$= \delta ||X_1||_* + G_1(X_1, X_2) \tag{7}$$

$$\nabla_{X_2} L = \mu\delta ||X_2||_1 - \lambda + \beta(X_1 + X_2 - X) = 0 \tag{8}$$

$$= \mu\delta ||X_2||_1 + G_1(X_1, X_2) \tag{9}$$

where δ represents *subgradient* of a convex function. From (7), using the Singular Value Thresholding,

$$G_1(X_1, X_2) = 0 \tag{10}$$

$$\implies \bar{X_1} = \frac{\lambda^k}{\beta} + X - X_2^k \tag{11}$$

$$X_1^{k+1} = U^k max(\Sigma^k - \frac{1}{\beta}\mathbb{I}, 0)(V^T)^k \tag{12}$$

$$where \, \bar{X_1} = U^k \Sigma^k (V^T)^k \tag{13}$$

From (9), using Shrinkage formula,

$$G_2(X_1, X_2) = 0 \tag{14}$$

$$\implies \bar{X_2} = \frac{\lambda^k}{\beta} + X - X_1^{k+1} \tag{15}$$

$$X_2^{k+1} = \text{sign}(\bar{X_1}). \max(|\bar{X_1}| - \frac{\mu}{\beta}, 0) \tag{16}$$

The update for the Lagrangian multiplier

$$\lambda^{k+1} = \lambda^k - \gamma\beta(X_1^{k+1} + X_2^{k+1} - X) \tag{17}$$

$$where \, \beta = \frac{0.25}{||X||_1} \tag{18}$$

$$and \, 0 \leq \gamma \leq \frac{\sqrt{5} + 1}{2} \tag{19}$$

Note: The intuition for both Shrinkage and Singular Value Thresholding (SVT) lies in the definition of l_1 and *nuclear* norms. An l_1 norm sums up absolute values of the elements of a matrix. If a small threshold value is subtracted from the absolute value of each element, the l_1 norm reduces. This is called *shrinkage.* Similarly, nuclear norm involves summing up all singular values of the matrix. In SVT, the threshold value is subtracted from each singular value, hence reducing the nuclear norm.

3.2 Model 2:Sparse Recovery From Compressive Measurements

Compressive measurements on video frames can be in the form of $\Phi \in \mathbb{R}^{m \times nJ}$ where

$$\Phi \equiv [\phi_1\, \phi_2 \ldots \phi_J] \tag{20}$$

$$where\, m < n \tag{21}$$

$$and\, \phi_j^T \phi_j = \mathbb{I} \tag{22}$$

$\phi_j \in \mathbb{R}^{m \times n}$ are randomly permuted rows of an $(n \times n)$ Walsh Hadamard Transform. These are performed on the video volume X to give a set of measurement vectors $Y \in \mathbb{R}^{m \times J}$ such that

$$\Phi \circ X \equiv Y \tag{23}$$

$$where\, \Phi \circ X = [\phi_1\, \phi_2 \ldots \phi_J][X_1\, X_2 \ldots X_J] \tag{24}$$

$$= [\phi_1 X_1\, \phi_2 X_2 \ldots \phi_J X_J] \tag{25}$$

$$and\, Y = [y_1\, y_2 \ldots y_J] \tag{26}$$

$y_j \in \mathbb{R}^{m \times 1}$ are measurement vectors corresponding to each vectorized frame x_j where $j = \{1, 2 \ldots J\}$.

Compressive measurements are performed, assuming that the given video volume is *sparse* in a specific basis. The framelet basis is chosen, which consists of wavelets that are tight frames. Framelet 2X toolbox[SA04] is used to implement this problem. The set of framelets $F \in \mathbb{R}^{n' \times nJ}$can be denoted as

$$F = [F_1\, F_2 \ldots F_J] \tag{27}$$

where $F_j \in \mathbb{R}^{n' \times n}$, $n' \geq n$. Hence, the recovery from sparse video volume as measurement in framelet basis, is expressed as

$$\min_X ||F \circ X||_1 \tag{28}$$

$$where\, F \circ X = [F_1\, F_2 \ldots F_J][x_1\, x_2 \ldots x_J] \tag{29}$$

$$= [F_1 x_1\, F_2 x_2 \ldots F_J x_J] \tag{30}$$

$$= F_j^T F_j = \mathbb{I} \tag{31}$$

$$s.t.\, \Phi \circ X = Y \tag{32}$$

$$and\, j = \{1, 2 ... J\} \tag{33}$$

The equivalent Lagrangian is

$$L(X) = ||F \circ X||_1 - \langle \lambda, \Phi \circ X - Y \rangle + \frac{\beta}{2} ||\Phi \circ X - Y||_F^2 \tag{34}$$

Video volume recovery from the measurement vectors Y is achieved by minimizing this expression.

3.3 Model 3: Background Subtraction: Sparse Signal Recovery from Compressive Measurements

Combining Model 1 and Model 2, one can formulate the background subtraction as a part of sparse signal recovery from compressive measurements. The following assumptions are made

1. The background component of the video volume in spatial basis is *low rank.*
2. The video volume is *sparse* in the framelet basis.
3. The background(low-rank) and moving objects are individually sparse in the framelet basis.

The last assumption gives us scope to recover individual elements separately. The formulation is hence the following minimization

$$\min_{X_1, X_2} ||X_1||_* + \mu_1 ||F \circ X_1||_1 + \mu_2 ||F \circ X_2||_1 \tag{35}$$

$$s.t.\, \Phi \circ (X_1 + X_2) = Y \tag{36}$$

To bring this expression to the form of suitable for applying shrinkage formula,

$$\min_{X_1, X_2} ||X_1||_* + \mu_1 ||Z_1||_1 + \mu_2 ||Z_2||_1 \tag{37}$$

$$s.t.\, \Phi \circ (X_1 + X_2) = Y \tag{38}$$

$$and\, Z_1 = F \circ X_1 \tag{39}$$

$$and\, Z_2 = F \circ X_2 \tag{40}$$

The corresponding Lagrangian takes the form

$$\begin{aligned} L(X_1, X_2, Z_1, Z_2) = ||X_1||_* + \mu_1||Z_1||_1 + \mu_2||Z_2||_1 \\ - \langle \lambda_1, F \circ X_1 - Z_1 \rangle + \frac{\beta_1}{2}||F \circ X_1 - Z_1||_F^2 \\ - \langle \lambda_2, F \circ X_2 - Z_2 \rangle + \frac{\beta_2}{2}||F \circ X_2 - Z_2||_F^2 \\ - \langle \lambda_3, \Phi \circ (X_1 + X_2) - Y \rangle + \frac{\beta_3}{2}||\Phi \circ (X_1 + X_2) - Y||_F^2 \end{aligned} \tag{41}$$

It is evident that this decomposition requires a minimization over 4 variables X_1, X_2, Z_1, Z_2 apart from setting 5 parameters μ_1, μ_2, β_1, β_2, β_3 accurately, to give optimum convergence. This can be a huge exercise, computationally and mathematically.

Hence additional conditions can be assumed

1. For large number of frames, the sparsity of background in the framelet basis can be ignored.
2. The nuclear norm of the background $||X_1||_*$ in the spatial basis, and that in the framelet basis, $||Z_1||_*$ is the same. Both X_1 and Z_1 have the same singular values. This can be easily proven.

As a result of these additional conditions, the problem statement and corresponding Lagrangian in (47) gets simplified as

$$\min_{X_1, X_2} ||Z_1||_* + \mu_2||Z_2||_1 \tag{42}$$

$$s.t.\ \Phi \circ (X_1 + X_2) = Y \tag{43}$$

$$or\ \Psi \circ (Z_1 + Z_2) = Y \tag{44}$$

$$where\ \Psi = \Phi \circ F^T \tag{45}$$

$$and\ F^T = [F_1^T\ F_2^T \dots F_J^T] \tag{46}$$

and the Lagrangian

$$\begin{aligned} L(Z_1, Z_2) = ||Z_1||_* + \mu_2||Z_2||_1 \\ - \langle \lambda, \Psi \circ (Z_1 + Z_2) - Y \rangle \\ + \frac{\beta}{2}||\Psi \circ (Z_1 + Z_2) - Y||_F^2 \end{aligned} \tag{47}$$

Analyzing the Problem:

Gradients of the Lagrangian with respect to Z_1 and Z_2

$$\begin{aligned}
\nabla_{Z_1} &= \delta||Z_1||_* - \Psi^T \circ \lambda + \beta(Z_1 + Z_2 - \Psi^T Y) && (48)\\
&= \delta||Z_1||_* + G_1(Z_1, Z_2) && (49)\\
\nabla_{Z_2} &= \delta||Z_2||_1 - \Psi^T \circ \lambda + \beta(Z_1 + Z_2 - \Psi^T Y) && (50)\\
&= \delta||Z_2||_1 + G_2(Z_1, Z_2) && (51)
\end{aligned}$$

Equating $G_1(Z_1, Z_2)$ and $G_2(Z_1, Z_2)$ to zero, obtain the updates

$$\begin{aligned}
Z_1^{k+1} &= U^k \max(\Sigma^k - \frac{1}{\beta}\mathbb{I}, 0)(V^T)^k && (52)\\
where\, \frac{1}{\beta}\Psi^T \circ \lambda^k + \Psi^T Y - Z_2^k &= U^k \Sigma^k (V^T)^k && (53)
\end{aligned}$$

Similarly,

$$\begin{aligned}
Z_2^{k+1} &= \text{sign}(\bar{Z}_2).(\max(|Z_2| - \frac{\mu}{\beta}) && (54)\\
where\, \bar{Z}_2 &= \frac{1}{\beta}\Psi^T \circ \lambda^k + \Psi^T Y - Z_1^{k+1} && (55)
\end{aligned}$$

Finally the update for the Lagrangian multiplier is

$$\begin{aligned}
\lambda^{k+1} &= \lambda^k - \gamma\beta(\Psi \circ (Z_1^{k+1} + Z_2^{k+1}) - Y) && (56)\\
where\, \beta &= \frac{0.25}{||Y||_1} && (57)\\
and\, 0 \leq &\gamma \leq \frac{\sqrt{5}+1}{2} && (58)
\end{aligned}$$

Finally, we recover the low rank and sparse components in the spatial basis as

$$\begin{aligned}
X_1^* &= W^T \circ Z_1^* && (59)\\
X_2^* &= W^T \circ Z_2^* && (60)
\end{aligned}$$

Algorithm: Method 3: Low Rank & Sparse Decomposition Using Compressed Measurements

```
Initialize $\Phi, Y, \beta, tolerance, maximum_iterations, \lambda^0, Z_1^0, Z_2^0$
while stopping criteria not met do
    update $Z_1$: $Z_1^{k+1} = U^k \max(\Sigma^k - \frac{1}{\beta}\mathbb{I}, 0)(V^T)^k$
    update $Z_2$: $Z_2^{k+1} = \text{sign}(\bar{Z}_2).(\max(|Z_2| - \frac{\mu}{\beta})$
    update $\lambda$: $\lambda^{k+1} = \lambda^k - \gamma\beta(\Psi \circ (Z_1^{k+1} + Z_2^{k+1}) - Y)$
$X_1^* = W^T \circ Z_1^*$
$X_2^* = W^T \circ Z_2^*$
```

where $Z_1^*, Z_2^*, X_1^*, X_2^*$ correspond to the final solution which minimizes the Lagrangian.

4 Simulation and Results

The Model 1, Model 2 and Model 3 are implemented, on two sets of video volumes. The first consists of a simplified fabricated video frame and the second consists of a re-sized surveillance video frame sequence, both of size 32 × 32 pixels. The algorithms have been implemented using MATLAB R2011a.

4.1 Sample 1

Actual video frame sequence:
Displaying 8 equally spaced frames of the total 15 frames taken:

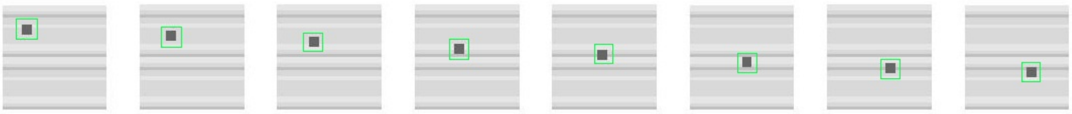

Figure 1: Video Frame Sequence

Actual low rank:

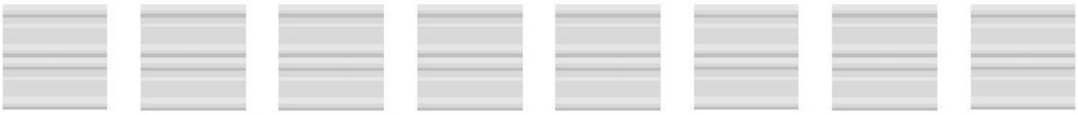

Figure 2: Background:Low Rank

Actual sparse:

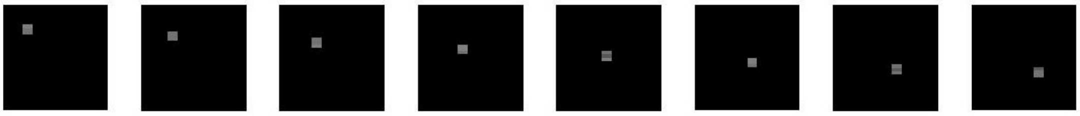

Figure 3: Moving Object:Sparse

4.2 Low Rank & Sparse Decomposition Using Model 1

Recovered sparse and low rank:
The following was recovered by setting $\mu = 0.1$ in as low as 10 iterations.

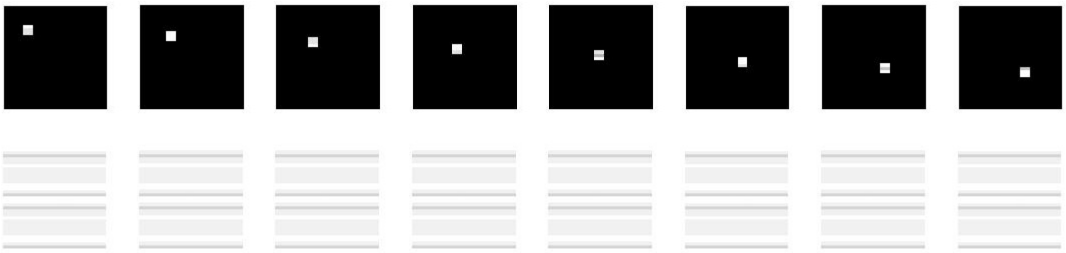

Figure 4: Sparse and Low-rank Decomposition

4.3 Sparse Signal Recovery From Compressive Measurements Using Model 2

Compressive measurement of video volume, taking 50% of samples:

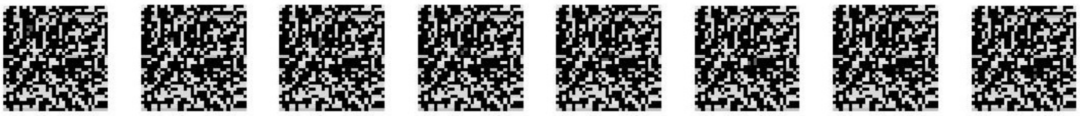

Figure 5: Compressive Sensing

Recovered video volume:

Figure 6: Sparse Recovery

4.4 Background Subtraction Using Model 3

Recovered low rank, after taking 50% samples, at $\mu = 0.12$:

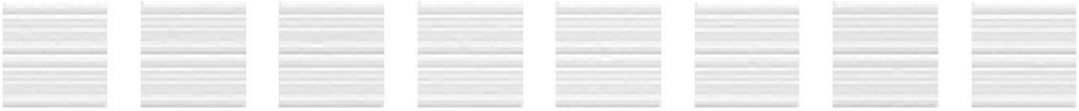

Figure 7: Low Rank: Recovered Background

Recovered sparse:

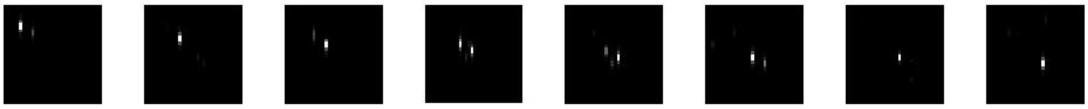

Figure 8: Sparse

4.5 Sample 2

Actual video frame sequence:

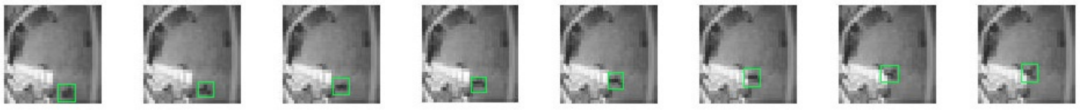

Figure 9: Video Frame Sequence

4.6 Low Rank & Sparse Decomposition Using Model 1

Recovered Sparse and Low Rank decomposition with $\mu = 0.3$:

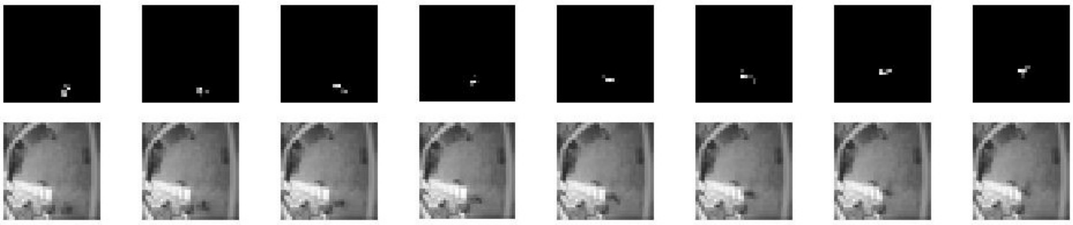

Figure 10: Sparse and Low Rank Decomposition

4.7 Sparse Signal Recovery From Compressive Measurements Using Model 2

Actual video frame sequence and corresponding compressive measurements using 60% samples:

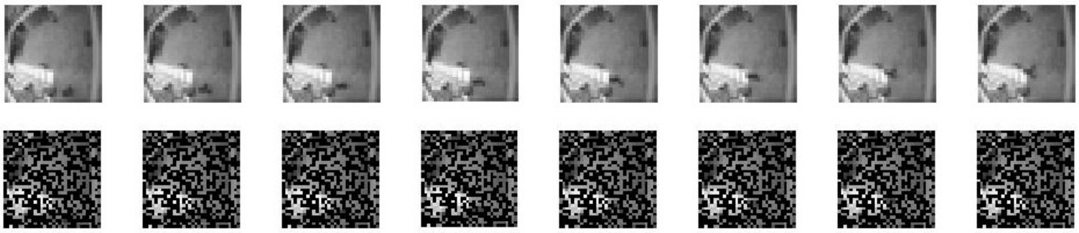

Figure 11: Video Volume and Compressive Measurements

Recovered video volume:

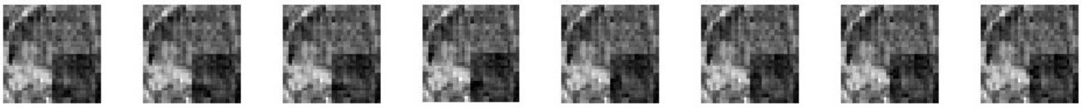

Figure 12: Sparse Recovery

4.8 Background Subtraction Using Model 3

Recovered sparse and recovered low rank at $\mu = 0.2$:

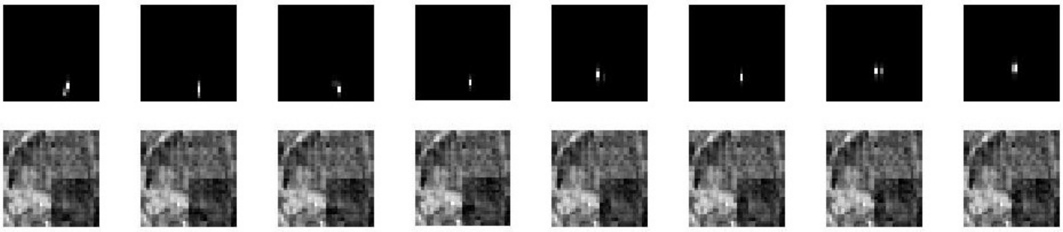

Figure 13: Sparse & Low Rank Recovery From Compressive Measurements

5 Future Work

The method described in subsection 3.3 works best when the number of frames used are large. To obtain better results, one can consider a slight modification, which assumes sparsity of the background in the framelet basis as well. Apart from this, a real-time spatio-temporal component can be added to this modelling, which uses characteristics of the continuous trajectories that are usually found in videos. Instead of using randomly permuted rows of the Walsh-Hadamard matrix, one can use an optimized measurement matrix for performing measurements.

References

[SA04] Ivan W Selesnick and A Farras Abdelnour. "Symmetric wavelet tight frames with two generators". In: *Applied and Computational Harmonic Analysis* 17.2 (2004), pp. 211–225.

[CRT06] Emmanuel J Candes, Justin K Romberg, and Terence Tao. "Stable signal recovery from incomplete and inaccurate measurements". In: *Communications on pure and applied mathematics* 59.8 (2006), pp. 1207–1223.

[Cev+08] Volkan Cevher et al. "Compressive sensing for background subtraction". In: *Computer Vision–ECCV 2008.* Springer, 2008, pp. 155–168.

[YY09] Xiaoming Yuan and Junfeng Yang. *Sparse and low-rank matrix decomposition via alternating direction methods.* Tech. rep. 2009.

[Pen+10] Yigang Peng et al. "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images". In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on.* June 2010, pp. 763–770. DOI: 10.1109/CVPR.2010.5540138.

[Can+11] Emmanuel J Candès et al. “Robust principal component analysis?” In: *Journal of the ACM (JACM)* 58.3 (2011), p. 11.

[JDS13] Hong Jiang, Wei Deng, and Zuowei Shen. “Surveillance Video Processing Using Compressive Sensing”. In: *CoRR* abs/1302.1942 (2013). URL: http://arxiv.org/abs/1302.1942.

[Mah+14] Amol Mahurkar et al. “Selective Visualization of Anomalies in Fundus Images via Sparse and Low Rank Decomposition”. In: *ACM SIGGRAPH 2014 Posters*. SIGGRAPH ’14. Vancouver, Canada: ACM, 2014, 109:1–109:1. ISBN: 978-1-4503-2958-3. DOI: 10.1145/2614217.2630595. URL: http://doi.acm.org/10.1145/2614217.2630595.