VIDEO BACKGROUND MODELING UNDER IMPULSE NOISE

Paul Rodríguez[*]

Department of Electrical Engineering
Pontificia Universidad Católica del Perú
Lima, Peru

Brendt Wohlberg[†]

T-5 Applied Mathematics and Plasma Physics
Los Alamos National Laboratory
Los Alamos, NM 87545, USA

ABSTRACT

Video background modeling is an important task in many video processing applications. Most existing algorithms assume a Gaussian noise model, but digital videos are, in practice, prone to be degraded by impulse noise, due to transmission errors in wireless or high data-rate wired channels. Principal Component Pursuit (PCP), which also assumes a Gaussian noise model, is currently considered the state of the art for video background modeling. We propose a new PCP-based algorithm that fully integrates the impulse noise model and has computational performance comparable with that of current PCP implementations.

Index Terms— Principal Component Pursuit, Video Background Modeling, Impulse noise

1. INTRODUCTION

Video background modeling, which consists of segmenting the moving objects or "foreground" from the static ones or "background" is an important task in several applications. In the present paper we restrict our interest to videos acquired by a static sensor, e.g. automatic video analysis (sports, traffic, etc.), surveillance systems [1], etc, which makes this problem more tractable. Most reported algorithms for video background modeling (see [2, 3] and the many reference therein) assume a Gaussian noise model as the underlying degradation. In practice, however, video sequences are often corrupted with inter-channel correlated impulse noise (either salt & pepper or random valued) [4] during the transmission stage, as a result of external effects such as thunderstorms, electric engines, wireless phones etc. [5, 6, 7]. The removal of this type of noise in video sequences has recently attracted both academic research [8, 9] and commercial [10] developments.

A recent survey [2], including a systematic evaluation and comparative analysis of several Principal Component Pursuit (PCP) / Robust Principal Component Analysis (RPCA) [11, 12] based algorithms, has shown that this approach provides state of the art performance in the video background modeling problem. (See [3] for a survey of alternative methods.) In this context, the PCP problem is

$$\arg\min_{L,S} \|L\|_* + \lambda \|S\|_1 \ \text{ s.t. } D = L + S \tag{1}$$

where $D \in \Re^{m \times n}$ is the observed video of n frames, each of size $m = N_r \times N_c \times N_d$ (rows, columns and depth or channels respectively), $L \in \Re^{m \times n}$ is a low rank matrix representing the background, $S \in \Re^{m \times n}$ is a sparse matrix representing the foreground, $\|L\|_*$ is the nuclear norm of matrix L (i.e. $\sum_k |\sigma_k(L)|$, the sum of the singular values of L), and $\|S\|_1$ is the ℓ^1 norm of S considered as a vector. As mentioned before, this problem assumes a Gaussian noise model, which is not appropriate for real videos corrupted by impulse noise.

The focus of the present paper is the development of a computationally efficient algorithm for color video background modeling under inter-channel correlated impulse noise (either salt & pepper or random value), as frequently encountered in practice. Our method (see Section 3), which is based on a modified version of a recently proposed fast PCP algorithm [13], first constructs an estimate of the set of outliers $\mathcal{N}$ (pixels corrupted with impulse noise) which is then incorporated into a modified PCP approach, iteratively using the low rank (background) and sparse approximation (foreground) of the previous iteration to improve the current noise-free sparse approximation estimate. To the best of our knowledge this is a novel approach, the most closely related ideas being described in Section 2.2. Our experimental results and comparisons, presented in Section 4, provide computational evidence of the performance of the proposed method.

2. NOISE MODEL AND RELATED WORKS

2.1. Noise model

For the scope of this work, we will use the same noise model considered in [8, 9]. The observed video sequence is represented by the matrix $D \in \Re^{m \times n}$, where each video frame, labeled as $\mathbf{d}_k$ $k \in \{1, 2, \ldots, n\}$, is represented as a column of D. If each frame is $N_r \times N_c \times N_d$ then $\mathbf{d}_k \in \Re^{N_r \times N_c \times N_d}$,

*This research was supported by the "Fondo para la Innovación, la Ciencia y la Tecnología" (Fincyt) Program.

†This research was supported by the NNSA's Laboratory Directed Research and Development Program.

with $m = N_r \times N_c \times N_d$ and $D(:,k) = \text{vec}(\mathbf{d}_k)$, where $\text{vec}(\cdot)$ is an operator that converts a matrix into a vector representation. Using this notation, the inter-channel correlated impulse noise is given by an impulse noise stage (2) follow by a correlation stage (3)

$$\mathbf{d}_k \quad = \quad \mathcal{B} \cdot (\mathbf{d}_k^*) + (\mathbf{1} - \mathcal{B}) \cdot \mathbf{r} \tag{2}$$

where arithmetic operations are to be considered element-wise, and $\mathbf{r}$ is either salt & pepper (SNP) noise or random-valued impulse (RVI) noise. $\mathcal{B}$ in (2) is a sample drawn from an i.i.d. multivariate Bernoulli distribution with success probability $1-p$, and $\mathbf{1}$ represents a vector with all elements equal to one. Since (2) can be used for SNP or RVI noise, then for the SNP noise case $r_k = \text{c}_{\min}$ or $r_k = \text{c}_{\max}$ with probability p_1 and p_2 respectively ($p = p_1 + p_2$) and for the RVI noise case r_k is drawn from a uniform distribution in $[\text{c}_{\min}, \text{c}_{\max}]$.

Finally, for the correlation stage, if pixel $d_k(\cdot,\cdot,l)$ $l \in \{1, 2, 3\}$ is corrupted, then

$$d_k(\cdot,\cdot,j) \quad = \quad \beta \cdot d_k(\cdot,\cdot,j) + (1-\beta) \cdot r \tag{3}$$

where $j \in \{1, 2, 3\} - \{l\}$ and β is a sample drawn from Bernoulli distribution with success probability $1-q$; in other words, for each pixel value and channel, if either of the other two channels has been corrupted by the impulse noise, the current channel will suffer a further corruption with a probability of $1-q$ (usually $q = 0.5$).

2.2. Previous related work

To the best of our knowledge, there are no prior methods that extend the PCP method to account for impulse noise. The two most closely related ideas are summarized below.

Wang and Trucco [14] proposed a low-rank single patch-based approach to denoise single images (not video) corrupted with non-pointwise (multi-pixel) random-valued impulse noise. This method is based on minimization of the functional $\frac{1}{2}\|W \cdot (L + S - P)\|_F^2 + \lambda\|W \cdot S\|_1 + \|L\|_*$, where P is a patch of the observed image and W is a mask indicating the location of noise-free pixels. In our case (see (12), Section 3.2) we consider a noise-free estimate of the noisy video based on previous low-rank and sparse estimates, whereas the optimization problem in [14] focuses only on the noise free entries.

Ye and Zhi [15] proposed an outlier detection method for high dimensional data (collections of similar images or frames), with application to nonlinear dimensionality reduction. In that work the inliers (from which outliers will be inferred) are detected by first constructing the nearest n data points, D_k, from the observed dataset D to image/frame $\mathbf{d}_k$, and then computing a partial SVD for each D_k from which the inliers are estimated. In our case the main goal is to estimate a sparse approximation (via (12)) from an impulse noise corrupted video and our frame-wise outlier detection is performed directly on D, whereas [15] performs the outlier detection on the low-rank approximations of D_k.

3. PROPOSED ALGORITHM

In this section we outline the PCP algorithm [13] on which our method is based before describing the proposed method for denoising of impulse noise.

3.1. Fast Principal Component Pursuit Via Alternating Minimization

A simple algorithm, amFastPCP, has recently been proposed [13] for solving the PCP problem. Instead of solving (1) directly, the approach is to solve

$$\arg\min_{L,S} \frac{1}{2}\|L + S - D\|_F^2 + \lambda\|S\|_1 \text{ s.t. } \text{rank}(L) = t \ , \tag{4}$$

via the alternating minimization

$$L_{k+1} = \arg\min_{L} \quad \|L + S_k - D\|_F^2 \text{ s.t. } \text{rank}(L) = t \tag{5}$$

$$S_{k+1} = \arg\min_{S} \quad \|L_{k+1} + S - D\|_F^2 + \lambda\|S\|_1. \tag{6}$$

Note that sub-problem (5) can be solved by computing a partial (with t components) SVD of $D - S_k$, which is the only computationally demanding part of the algorithm. The solution to (6) is simple element-wise shrinkage (soft thresholding) $\text{shrink}(D - L_{k+1}, \lambda)$, where

$$\text{shrink}(x, \epsilon) = \text{sign}(x)\max\{0, |x| - \epsilon\} \ . \tag{7}$$

The solution obtained via the iterative solution of (5)-(6) is of comparable quality to the solution of the original PCP problem [13]. Furthermore, the amFastPCP algorithm delivers a useful estimate of the sparse component even after a single outer loop, being approximately an order of magnitude faster than the Inexact ALM [16] algorithm to construct a sparse component of the same quality.

3.2. Impulse Noise Removal

Define M as a mask matrix indicating locations of the outliers, i.e. pixels corrupted with impulse noise. This mask is estimated via either (i) the adaptive median filter [17] for the SNP noise case, or (ii) the directional weighted median filter [18] for the RVI noise case. Given $\mathbf{1}$, a matrix whose entries are all 1, the observed noisy video can be expressed as

$$D = M \odot D^* + (\mathbf{1} - M) \odot R, \tag{8}$$

where $\odot$ is element-wise multiplication (Hadamard product), D^* is the noise free video and R is impulse noise. Assuming that L^* and S^* are the exact low-rank and sparse approximation of D^* (i.e. $D^* = L^* + S^*$), and given that $D^* = M \odot D^* + (\mathbf{1} - M) \odot D^* = M \odot D^* + (\mathbf{1} - M) \odot (L^* + S^*)$, then

$$D^* = M \odot D^* + (\mathbf{1} - M) \odot L^* + (\mathbf{1} - M) \odot S^*, \tag{9}$$

from which it trivially follows that

$$
\begin{aligned}
L^* &= M \odot D^* + (\mathbf{1} - M) \odot L^* - M \odot S^* && (10) \\
S^* &= M \odot D^* - M \odot L^* + (\mathbf{1} - M) \odot S^*. && (11)
\end{aligned}
$$

Inspired by (10)-(11), and based on (4) we propose the following iterative optimization problem to solve the PCP problem for impulse noise corrupted videos:

$$
\arg\min_{L_k, S_k} \frac{1}{2} \| L_k + S_k - (\mathbf{1} - M) \odot L_{k-1} - (\mathbf{1} - M) \odot S_{k-1} - M \odot D \|_F^2 + \lambda \| S_k \|_1 \quad \text{s.t.} \quad \operatorname{rank}(L_k) = t \,. \quad (12)
$$

A natural approach to solving problem (12) is the alternating minimization

$$
\begin{aligned}
L_k &= \arg\min_{L} \ \| L + S_{k-1} - M \odot D - (\mathbf{1} - M) \odot L_{k-1} - (\mathbf{1} - M) \odot S_{k-1} \|_F^2 \quad \text{s.t. } \operatorname{rank}(L) = t && (13) \\
S_k &= \arg\min_{S} \ \| L_k + S - M \odot D - (\mathbf{1} - M) \odot L_k - (\mathbf{1} - M) \odot S_{k-1} \|_F^2 + \lambda \| S \|_1 && (14)
\end{aligned}
$$

Initialize

$M =$ estimated mask, $D =$ input video, $t = 1$ (initial rank)

Initial solution ($k = 0$)

$L_0 = \arg\min_L \| L - D \|_F^2 \quad \text{s.t. } \operatorname{rank}(L_k) = t$

$S_0 = \arg\min_S \| S + M \odot (L - D) \|_F^2 + \lambda \| S \|_1$

$S_0 = \text{adaptiveMedianFilter}(S_0, M, \text{nbhood})$ (optional)

for $k = 1, 2, .., \text{outerLoops}$

if $\frac{v_{\text{rank}}}{\sum_{k=1}^{\text{rank}} v_k} > \tau$ then ++t (v: singular values from $k - 1$)

solve (13) with current t

solve (14)

Algorithm 1: Proposed alternating algorithm for PCP when the input video is corrupted with impulse noise, including a simple procedure to estimate an upper bound for t in (13). From experimental results, the optional adaptive median filter step for the initial solution accelerates the convergence of the sparse approximation.

As for (5)-(6), sub-problems (13)-(14) can be solved via partial SVD and element-wise shrinkage, resulting in Algorithm 1. Note that while initial solution L_0 is just the low-rank approximation of the observed noisy video, the initial sparse approximation S_0 does take into account the set of corrupted pixels, represented by matrix M with elements equal to 1 for the estimated noiseless pixel set and 0 otherwise. We have observed empirically that if S_0 is denoised via an adaptive scheme that takes into account the set M (only "corrupted" elements are denoised), this greatly improves the convergence of the sparse approximation; as a trade-off between restoration quality and computational cost we have used [17] for this purpose. This denoising step is only performed once on the initial sparse approximation S_0.

4. COMPUTATIONAL RESULTS

We use a 640×480 pixel, 400-frame color traffic video sequence of 26.66 seconds at 15 fps, from the Lankershim Boulevard dataset [19, camera3] as a test video. All experiments are also performed on a scaled down 320×240 pixel version of the same video. Each frame was scaled by $\frac{1}{255}$, so that all pixel values are in the range $[0, 1]$. As "ground truth" we use the sparse video approximation, labeled S_{GT}, of the noise free video computed via the inexact ALM [16] or iALM (code downloaded from [20]) solution after 20 outer iterations.

All simulations have been carried out using Matlab code that makes heavy use of the PROPACK [21] library (which is also the case for the inexact ALM algorithm), on a 64 processor server, based on 2.1GHz AMD Opteron "6272" CPUs (L2: 2048K, RAM: 264G). The value of λ for our proposed method and as well as for [13] was empirically chosen to be initially 0.025, and then linearly decreased toward the optimal value $1/\sqrt{M}$ (see [12, Section 1.4]). Although the optimization problem is not jointly convex, we found that our algorithm converged reliably to a good solution without the need for multiple runs with different initial solutions. The experimental code is made publicly available [22] in compliance with the principle of reproducible research.

In order to assess the performance of our proposed algorithm we present two kind of results: (i) computational performance measured as the time to complete a given number of outer iterations and (ii) reconstruction quality at each outer iteration measured by $\frac{\| S_{GT} - S_k \|_1}{N}$ where S_k is the sparse video approximation of a given algorithm at the k^{th} outer iteration, and N is the number of pixels per frame (used as a normalization factor).

In Fig. 1 we compared the sparse video approximation, obtained by our proposed algorithm (labeled "Proposed") applied to the noisy observed video (corrupted with 10% of SNP noise), with those obtained by "iALM" of the noise free video, labeled "iALM noiseless" and by "amFastPCP" [13] of the (i) noise free video, labeled "amFastPCP noiseless" and (ii) frame-wise denoised observed video, labeled "amFastPCP denoised", via an adaptive ℓ^1-TV implementation (based on [23]). Note that we attempted to use [9] (considered the state of the art for impulse noise removal from color video sequences, executable Win32 code available from [24]), but the provided implementation does not properly estimate the location of the noisy pixels for our noisy test video, and thus the resulting denoised video is still too noisy (on average only 20% of the noisy pixel are denoised) to serve as a good approximation of the noise-free video. The quality of the results of the proposed method are slightly better than those obtained by "amFastPCP denoised" (as shown in Fig. 1, where 10 outer iterations are considered for each method), and comparable with (but slightly inferior to) those obtained when the the sparse video approximation is computed using the noise-free video. Similar results were obtained when the input video was corrupted with 5% and 15% of SNP noise.

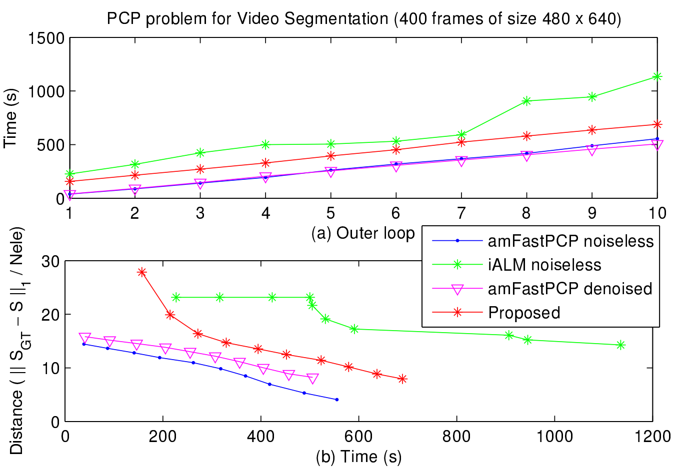

Fig. 1. Computational performance (a) and sparse representation quality (b) comparisons for a noisy video (corrupted with 10% of SNP noise) between our propose algorithm (labeled "Proposed", applied to the noisy video), the amFastPCP [13] (labeled "amFastPCP noiseless") and iALM [16] (labeled "iALM noiseless") algorithm, applied to the original noise-free video. We also include results for the amFastPCP when the noisy video is first frame-wise denoised (labeled "amFastPCP denoised") via an adaptive ℓ^1-TV implementation (based on [23]). We do not include the pre-processing time (denoising, 9010 sec.) for "amFastPCP noiseless" nor for "Proposed" (785 sec. for outlier estimation, see (8)).

Adaptive ℓ^1-TV denoising

Noise level	average SNR		average PSNR	
	320×240	640×480	320×240	640×480
5%	27.12	24.38	39.27	36.01
10%	24.41	22.55	36.56	34.41
15%	22.31	20.86	34.46	32.48

Table 1. Average SNR/PSNR frame reconstruction quality obtained via adaptive ℓ^1-TV [23] (pre-processing associated with "amFastPCP denoised") applied to the noisy video.

Moreover, our proposed method is one order of magnitude faster than "amFastPCP denoised": For the "amFastPCP denoised' case, the time required to denoise (via adaptive ℓ^1-TV [23]) the 640×480 pixel, 400-frame test video (corrupted with 10% of SNP noise) was, 9010 seconds, with an average SNR/PSNR frame's reconstruction quality of 22.55 dB and 34.41 dB respectively. This time performance is comparable to that of other methods for impulse noise denoising of color video sequences (for instance [9]). For our proposed algorithm, the pre-processing (mask estimation, see (8)) time was 785 seconds. For other noise levels or video resolutions, in Tables 1 and 2 we provide the average SNR/PSNR frame reconstruction quality (denoising stage related to "amFastPCP denoised") and pre-processing time for the "amFastPCP denoised" and "Proposed" methods respectively. From Table 2 we conclude that the pre-processing time for our proposed algorithm is one order of magnitude faster than that of "amFastPCP denoised". Furthermore, taking into account the actual time needed to compute the sparse approximation (which is comparable for both cases), overall our proposed method is still one order of magnitude faster than "amFastPCP noiseless".

Computational cost: Pre-processing time (s.)

Noise level	Adaptive ℓ^1-TV denoising		Mask estimation (see (8))	
	320×240	640×480	320×240	640×480
5%	1906	10259	217	758
10%	1926	9010	220	785
15%	1996	9446	224	790

Table 2. Pre-processing computational performance to denoise (via adaptive ℓ^1-TV [23]), task associated with "amFastPCP denoised" and for the mask estimation (see (8)), task associated with our proposed algorithm. The latter is one order of magnitude faster than the former.

Due to space constraints, we do not present results for the random value impulse (RVI) noise case (these can be obtained via [22]), but we mention that while the computational performance is about the same as for the SNP noise case and that the quality of the sparse approximation is also slightly better than "amFastPCP denoised", the quality of the sparse approximation obtained by our proposed method is not as good as for the SNP case when compared to that of the "amFastPCP noiseless" or "iALM noiseless". This is due to the fact that the mask estimation for the RVI noise is not as precise as in the case of SNP noise.

5. CONCLUSIONS

Digital videos are, in practice, prone to be degraded by impulse noise, nevertheless most video background modeling methods assume a Gaussian noise model. We have presented a computationally efficient method that integrates the impulse noise model into the PCP methodology, considered the state of the art for video background modeling applications. Our computational results show that our proposed method is able to provide a good sparse representation whose quality is comparable with that obtained from a noise free video, while being slightly better (and one order of magnitude faster) than that obtained by first denoising the noisy video prior to the PCP decomposition. Future work will focus on providing a theoretical analysis of our proposed algorithm.

6. REFERENCES

[1] T. Bouwmans, F. Baf, and B. Vachon, *Handbook of Pattern Recognition and Computer Vision*, chapter Statistical Background Modeling for Foreground Detection: A Survey, pp. 181–199, World Scientific Publishing, 2010.

[2] C. Guyon, T. Bouwmans, and E. Zahzah, "Robust principal component analysis for background subtraction: Systematic evaluation and comparative analysis," in *Principal Component Analysis*, P. Sanguansat, Ed., chapter 12. InTech, 2012.

[3] M. Shah, J. Deng, and B. Woodford, "Video background modeling: recent approaches, issues and our proposed techniques," *Machine Vision and Applications*, pp. 1–15, 2013.

[4] Z. Ma, H. Wu, and B. Qiu, "A robust structure-adaptive hybrid vector filter for color image restoration," *Image Processing, IEEE Transactions on*, vol. 14, no. 12, pp. 1990–2001, 2005.

[5] P. Torio and M. Sanchez, "Mitigation of impulsive noise in digital video broadcasting terrestrial using orthogonal polarization reception," *Consumer Electronics, IEEE Transactions on*, vol. 55, no. 4, pp. 1798–1804, 2009.

[6] S. Ghadimi, J. Hussian, T. Sidhu, and S. Primak, "Effect of impulse noise on wireless relay channel," *Wireless Sensor Network*, vol. 4, no. 6, pp. 167–172, 2012.

[7] R. Fantacci, A. Tani, and D. Tarchi, "Impulse noise mitigation techniques for xDSL systems in a real environment," *Consumer Electronics, IEEE Transactions on*, vol. 56, no. 4, pp. 2106–2114, 2010.

[8] T. Melange, M. Nachtegael, and E. Kerre, "Fuzzy random impulse noise removal from color image sequences," *Image Processing, IEEE Transactions on*, vol. 20, no. 4, pp. 959–970, 2011.

[9] L. Jin, H. Liu, X. Xu, and E. Song, "Quaternion-based impulse noise removal from color video sequences," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 23, no. 5, pp. 741–755, 2013.

[10] Broadcom Corporation, "Broadcom to demonstrate innovative impulse noise protection technology at broadband world forum," `http://investor.broadcom.com/releasedetail.cfm?ReleaseID=351462`, 2007.

[11] 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.

[12] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, May 2011.

[13] P. Rodríguez and B. Wohlberg, "Fast principal component pursuit via alternating minimization," in *Proc. of IEEE Int. Conf. on Image Processing (ICIP)*, Melbourne, Australia, Sep. 2013, pp. 69–73.

[14] R. Wang and E. Trucco, "Single-patch low-rank prior for non-pointwise impulse noise removal," in *International Conference on Computer Vision 2013 (ICCV'2013)*, 2013.

[15] Q. Ye and W. Zhi, "Outlier detection in the framework of dimensionality reduction," `http://www.ms.uky.edu/~qye/reports/ODA.pdf`, 2013.

[16] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," arXiv:1009.5055v2, 2011.

[17] H. Hwang and Richard A. Haddad, "Adaptive median filters: New algorithms and results," *IEEE Trans. Image Process.*, pp. 499–502, 1995.

[18] Yiqiu Dong and Shufang Xu, "A new directional weighted median filter for removal of random-valued impulse noise," *IEEE Signal Processing Letters*, vol. 14, pp. 193–196, 2007.

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

[20] "Low-rank matrix recovery and completion via convex optimization (sample code)," `http://perception.csl.illinois.edu/matrix-rank/sample_code.html`.

[21] R. M. Larsen, "PROPACK," Functions for computing the singular value decomposition of large and sparse or structured matrices, available from `http://sun.stanford.edu/~rmunk/PROPACK`.

[22] B. Wohlberg P. Rodriguez, "Fast PCP simulation," `http://sites.google.com/a/istec.net/prodrig/Home/en/pubs`.

[23] R. Rojas and P. Rodríguez, "Spatially adaptive total variation image denoising under salt and pepper noise," in *European Signal Processing Conference (EUSIPCO)*, Barcelona, Spain, 2011, pp. 314–318.

[24] L. Jin, H. Liu, X. Xu, and E. Song, "Quaternion-based impulsive noise removal from color video sequences," `http://www.cbib.cn/lianghaijin.html#_VideoDenoising%281%29`.