A New Adaptive Video SRR Algorithm with Improved Robustness to Innovations

Ricardo Augusto Borsoi*, Guilherme Holsbach Costa†, José Carlos Moreira Bermudez*

* Federal University of Santa Catarina, Florianópolis, SC - Brazil
† University of Caxias do Sul, Caxias do Sul, RS - Brazil

***Abstract*—In this paper, a new video super-resolution reconstruction (SRR) method with improved robustness to outliers is proposed. By studying the proximal point cost function representation of the R-LMS iterative equation, a better understanding of its performance is attained, which allows us to devise a new algorithm with improved robustness, while maintaining comparable quality and computational cost. Monte Carlo simulation results illustrate that the proposed method outperforms the traditional and regularized versions of the LMS algorithm.**

***Index Terms*—Super-resolution, R-LMS, outliers**

I. Introduction

Super-resolution reconstruction (SRR) is an approach to digital image quality improvement that has attracted great interest in the last decade. SRR consists basically of combining multiple low-resolution (LR) images of the same scene or object to form a higher resolution (HR) image, outperforming physical limits of image sensors. References [1], [2] review several important concepts and initial results on SRR.

The algorithms available for SRR can be generally divided in two classes: image SRR, which reconstructs a single HR image from multiple observations, and video SRR, which reconstructs an entire HR video sequence. Video SRR algorithms often improve the quality of the reconstructed sequence by employing some temporal regularization that constrains the norm of the changes in the solution between adjacent time instants [3]–[6].

One of the major issues in SRR is the computational cost of the algorithms. However, real-time video SRR applications require simple algorithms. This limitation prompted a significant interest in the development of low complexity SRR algorithms, with one notable example being the (R)-LMS algorithm [7].

Unfortunately, the performance of these simple algorithms tends to be heavily affected by the presence of outliers such as large innovations. While strategies for obtaining robust algorithms are common in the literature [5], [8], their computational cost is not comparable to that of algorithms like the R-LMS. Interpolation algorithms might seem to be a reasonable option, as their performance is not affected by outliers. However, they do not offer a quality improvement comparable to SRR methods [5]. Therefore, it is of interest to develop video SRR algorithms that combine good quality, robustness to outliers and low computational cost.

In this paper, a new adaptive SRR algorithm is proposed. Through a proximal point cost function representation of the R-LMS recursive equation, we attain a better understanding of its quality performance and robustness in different situations. This allows us to devise a new regularization that addresses the identified problems. The new resulting algorithm presents improved robustness and similar quality at a comparable computational cost.

Section 2 presents the signal model used. In Section 3, we derive the R-LMS algorithm [7] as a stochastic gradient solution to the image estimation problem. In Section 4 we study the behavior of the R-LMS algorithm under different situations. In Section 5, we propose the new regularization and derive the corresponding adaptive algorithm. In Section 6 we illustrate the performance of the proposed algorithm through simulations. Finally, in Section 7, we conclude this paper.

II. Image Acquisition Model

Given the $N \times N$ matrix representation of an LR (observed) digital image $\mathbf{Y}(t)$ and an $M \times M$ ($M > N$) matrix representation of the original HR digital image $\mathbf{X}(t)$, the acquisition process can be modeled as [1]:

$$\mathbf{y}(t) = \mathbf{D}\mathbf{H}\mathbf{x}(t) + \mathbf{e}(t)\,, \tag{1}$$

where vectors $\mathbf{y}(t)$ ($N^2 \times 1$) and $\mathbf{x}(t)$ ($M^2 \times 1$) are the lexicographic representations of the degraded and original images, respectively, at discrete time instant t. $\mathbf{D}$ is an $N^2 \times M^2$ decimation matrix and models the subsampling taking place in the sensor. $\mathbf{H}$ is an $M^2 \times M^2$ time-invariant matrix (without loss of generality) that models the blurring. Here, it is assumed to be known. The $N^2 \times 1$ vector $\mathbf{e}(t)$ models the observation (electronic) noise, whose properties are assumed to be determined from camera tests.

The dynamics of the input signal is modeled by [7]

$$\mathbf{x}(t) = \mathbf{G}(t)\mathbf{x}(t-1) + \mathbf{s}(t)\,, \tag{2}$$

where $\mathbf{G}(t)$ is the warp matrix that describes the relative displacement from $\mathbf{x}(t-1)$ to $\mathbf{x}(t)$. Vector $\mathbf{s}(t)$ models the innovations in $\mathbf{x}(t)$.

III. The R-LMS-SRR algorithm

Several SRR solutions are based on the minimization of the estimation error (see [1] and references therein)

$$\boldsymbol{\epsilon}(t) = \mathbf{y}(t) - \mathbf{D}\mathbf{H}\hat{\mathbf{x}}(t) \tag{3}$$

ISBN 978-0-9928626-7-1

where $\hat{\mathbf{x}}(t)$ is the estimated HR image, and $\boldsymbol{\epsilon}(t)$ can be interpreted as the estimate of $\mathbf{e}(t)$ in (1). The LMS-SRR algorithm attempts to minimize the mean-square value of the L_2 norm of (3) conditioned on the estimate $\hat{\mathbf{x}}(t)$ [7], [9]. Thus, it minimizes the cost function $\mathbf{J}_{\text{MS}}(t) = \mathrm{E}\{\|\boldsymbol{\epsilon}(t)\|^2 \,|\, \hat{\mathbf{x}}(t)\}$.

Since natural images are known to be intrinsically smooth, this *a priori* knowledge can be added to the estimation problem in the form of a regularization to the LMS algorithm by constraining the solution that minimizes $\mathbf{J}_{\text{MS}}(t)$. The R-LMS algorithm [10], [11] then arises as the solution to the following constrained optimization problem

$$\mathcal{L}_{\text{R-MS}}(t) = \mathrm{E}\{\|\mathbf{y}(t) - \mathbf{D}\mathbf{H}\hat{\mathbf{x}}(t)\|^2 \,|\, \hat{\mathbf{x}}(t)\} + \alpha\|\mathbf{S}\hat{\mathbf{x}}(t)\|^2\,, \quad (4)$$

where $\mathbf{S}$ is the Laplacian operator [12, p. 182]. Note that the performance surface in (4) is defined for each time instant t, and the expectation is taken over the ensemble.

Following the steepest descent method, the HR image estimate is updated in the negative direction of the gradient

$$\nabla\mathcal{L}_{\text{R-MS}}(t) = -2\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\{\mathrm{E}[\mathbf{y}(t)] - \mathbf{D}\mathbf{H}\hat{\mathbf{x}}(t)\} + 2\alpha\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}(t) \quad (5)$$

and thus the iterative update of $\hat{\mathbf{x}}(t)$ for a fixed value of t is given by

$$\hat{\mathbf{x}}_{k+1}(t) = \hat{\mathbf{x}}_k(t) - \frac{\mu}{2}\nabla\mathcal{L}_{\text{R-MS}}(t), \;\; k = 0, 1, \ldots, K-1 \quad (6)$$

where $K \in \mathbb{Z}_+$ is the number of iterations of the algorithm, and μ is the step size used to control the convergence speed. The factor 1/2 is just a convenient scaling.

The R-LMS algorithm is the stochastic version of the steepest descent algorithm. Using the instantaneous estimate of (5) in (6) yields

$$\hat{\mathbf{x}}_{k+1}(t) = \hat{\mathbf{x}}_k(t) + \mu\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}[\mathbf{y}(t) - \mathbf{D}\mathbf{H}\hat{\mathbf{x}}_k(t)] - \alpha\mu\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_k(t), \quad k = 0, 1, \ldots, K-1\,, \quad (7)$$

which is the R-LMS update equation for a fixed value of t. The time update of (7) is based on the signal dynamics (2), and is performed by the following expression [7]:

$$\hat{\mathbf{x}}_0(t+1) = \mathbf{G}(t+1)\hat{\mathbf{x}}_K(t)\,. \quad (8)$$

Between two time updates, (7) is iterated for $k = 0, \ldots, K-1$. The estimate $\hat{\mathbf{x}}(t)$ at a given time instant t is then given by $\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}_K(t)$.

IV. R-LMS Performance in the Presence of Outliers

The R-LMS algorithm is computationally efficient when implemented with few stochastic gradient iterations (small K) per time instant t. Nevertheless, one important issue that plagues most low-complexity super-resolution algorithms is the occurrence of outliers. Take for instance the R-LMS algorithm, which is derived under the assumption that the solution $\mathbf{x}(t)$ is only slightly perturbed between time instants. When the estimate $\hat{\mathbf{x}}(t)$ has already achieved a reasonable quality (i.e. $\hat{\mathbf{x}}(t) \simeq \mathbf{x}(t)$), the initialization for the next time instant performed according to (8) will already be relatively close to the optimal solution, what explains the good steady-state performance of the algorithm. However, due to the slow convergence of the R-LMS, the presence of innovation outliers is known to negatively affect the quality of the super-resolved images, often creating visible artifacts that can result in reconstructed images of quality inferior to that of the observed LR images themselves.

An interesting interpretation of the R-LMS algorithm is possible if we view each iteration of the gradient algorithm (6) (for a fixed value of t) as a proximal regularization of the cost function $\mathcal{L}_{\text{R-MS}}(t)$ linearized about the estimation of the previous iteration $\hat{\mathbf{x}}_k(t)$. Proceeding as in [13, Section 2.2] or [14, p. 546], the gradient iteration (6) can be written as

$$\hat{\mathbf{x}}_{k+1}(t) = \arg\min_{\mathbf{z}} \left\{ 2\alpha\mathbf{z}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_k(t) - 2\mathbf{z}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\,\mathrm{E}[\boldsymbol{\epsilon}_k(t)] + \frac{1}{\mu}\,\|\mathbf{z} - \hat{\mathbf{x}}_k(t)\|^2 \right\}, \quad (9)$$

where $\mathrm{E}[\boldsymbol{\epsilon}_k(t)]$ is the expected value of the observation error (3) conditioned on $\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}_k(t)$.

Now, the presence of the squared norm within the external brackets in (9) means that the optimization algorithm seeks $\hat{\mathbf{x}}_{k+1}(t)$ that minimizes the perturbation $\hat{\mathbf{x}}_{k+1}(t) - \hat{\mathbf{x}}_k(t)$ at each iteration. Evidencing this property leads to a more detailed understanding of the dynamical behavior of the algorithm, its robustness properties and the reconstruction quality it provides.

For instance, this constraint on the perturbation of the solution explains how the algorithm tends to preserve in $\hat{\mathbf{x}}(t)$ details estimated during the previous time instants and that were present in $\hat{\mathbf{x}}(t-1)$. However, the presence of this term also opposes changes from $\hat{\mathbf{x}}_k(t)$ to $\hat{\mathbf{x}}_{k+1}(t)$, and thus tends to slow down the reduction of the observation error from $\boldsymbol{\epsilon}_k(t)$ to $\boldsymbol{\epsilon}_{k+1}(t)$, as changes in $\boldsymbol{\epsilon}_k(t)$ require changes in $\hat{\mathbf{x}}_k(t)$. Therefore, this algorithm cannot simultaneously achieve a fast convergence rate and preserve the super resolved details. Then, for the interesting practical case of a small number of iterations per time instant (small K), the time sequence of reconstructed images will either converge fast but presenting low temporal correlation between time estimations (therefore leading to a solution that approaches an interpolation of $\mathbf{y}(t)$), or will converge slowly and yield a highly correlated image sequence that generally presents better quality but is susceptible to innovation outliers, thus showing a significant deviation from the desired signal in their presence.

One should note that, although the solution $\hat{\mathbf{x}}(t)$ can hardly approach the desired solution $\mathbf{x}(t)$ in few iterations, if the total number of iterations K during a single time interval t is sufficiently large, the solution can adapt to track the innovations even with a large weighting for the term $\frac{1}{\mu}\|\mathbf{z} - \mathbf{x}_k(t)\|^2$. This way it becomes possible for the algorithm to achieve and maintain a good reconstruction quality both during normal operation and in the presence of an outlier, although at a prohibitive computational cost, thus defeating the purpose of the algorithm.

This behavior is illustrated in Figure 1, where the R-LMS

ISBN 978-0-9928626-7-1 © EURASIP 2017

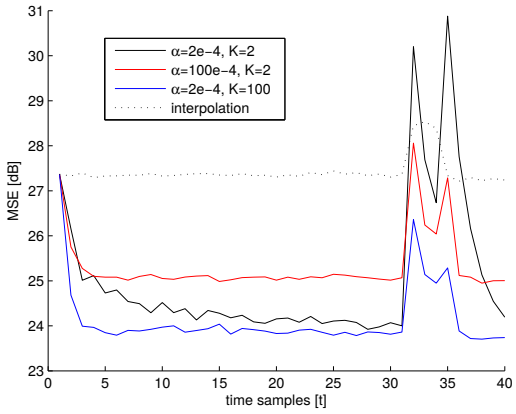

Fig. 1. MSE results for the R-LMS algorithm with different values of α and K.

algorithm was applied to reconstruct synthetically generated video sequences containing an outlier at $t = 32$, with both $\alpha = 2 \times 10^{-4}$ and $\alpha = 100 \times 10^{-4}$, for $\mu = 4$, and for $K = 2$ and $K = 100$. It can be seen that a larger regularization parameter, while increasing the reconstruction error for small innovations, substantially decreases it in the presence of outliers. On the other hand, allowing the algorithm to converge slowly in each time interval leads to a better reconstruction result in both situations. In the light of the aforementioned limitations of the R-LMS algorithm, it is desired to devise an algorithm that performs better both in terms of robustness, quality and computational cost.

V. A Fast Video SRR Algorithm Robust to Innovations

In the context of the R-LMS algorithm, the temporal regularization, which consists in constraining the value of $\|\hat{\mathbf{x}}(t) - \mathbf{G}(t)\hat{\mathbf{x}}(t-1)\|$ in the SRR cost function [3], [5], can be interpreted as the application of the well known least perturbation or minimum disturbance principle.

However, differently from simultaneous video SRR methods, the proximal regularization described in Section IV guarantees consistence between consecutive iterations in k. Therefore, since the previous time instant's solution $\hat{\mathbf{x}}(t-1)$ is introduced during the initialization in (8), consistence between consecutive time instants is achieved if the solution is not significantly disturbed during all iterations $k = 1, \ldots, K$ (i.e. $\hat{\mathbf{x}}_K(t) \simeq \hat{\mathbf{x}}_0(t)$),

Albeit adding the temporal regularization to the R-LMS algorithm removes its dependence on the time initialization (8), it makes the algorithm less robust since it prevents convergence to the desired solution $\mathbf{x}(t)$ in the presence of large innovations even for a large number of iterations (large K). Furthermore, as already discussed, temporal consistency can already be retained due to the proximal regularization (even for large K).

In the case of simultaneous SRR, robust algorithms are obtained as the result of non-linear cost functions [3], [5], [8]. Although these techniques achieve good reconstruction results, their increased computational cost turns real-time operation unfeasible even for the faster algorithms.

Differently from the simultaneous methods, the robustness problem of the R-LMS is related with its slow convergence. Therefore, in order to devise an algorithm that achieves a faster convergence without experiencing the loss of the estimated details like the R-LMS does, we will consider a modified version of the temporal regularization. We will only constrain the changes in the details of the estimated image between time instants, therefore preserving the desired content while allowing the algorithm to converge faster to adapt to the new information observed in $\mathbf{y}(t)$.

The new algorithm is obtained by minimizing the following cost function:

$$\mathcal{L}(t) = \mathrm{E}\{\|\mathbf{y}(t) - \mathbf{DH}\hat{\mathbf{x}}(t)\|^2 \,|\, \hat{\mathbf{x}}(t)\} + \alpha\|\mathbf{S}\hat{\mathbf{x}}(t)\|^2 + \alpha_{\mathrm{T}}\|\mathbf{S}\big(\hat{\mathbf{x}}(t) - \mathbf{G}(t)\hat{\mathbf{x}}(t-1)\big)\|^2 . \quad (10)$$

Calculating the gradient of the cost function with respect to $\hat{\mathbf{x}}(t)$, setting it equal to $\mathbf{0}$, and approximating the statistical expectations by their instantaneous values yields the iterative equation for the new algorithm:

$$\hat{\mathbf{x}}_{k+1}(t) = \hat{\mathbf{x}}_k(t) - \mu\alpha_{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\big(\mathbf{S}\hat{\mathbf{x}}_k(t) - \mathbf{SG}(t)\hat{\mathbf{x}}(t-1)\big) - \mu\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\big(\mathbf{DH}\hat{\mathbf{x}}_k(t) - \mathbf{y}(t)\big) - \mu\alpha\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_k(t), \quad (11)$$

where the time update is based on the signal dynamics (2) and performed by $\hat{\mathbf{x}}_0(t+1) = \mathbf{G}(t+1)\hat{\mathbf{x}}_K(t)$ [7].

VI. Results

The performance of the proposed algorithm is illustrated through a Monte Carlo (MC) simulation with 50 realizations. The HR video sequences were generated using images of different scenes of size $M = 256$, with global translational motion consisting of $\{-1, 1\}$ i.i.d. displacements. The resulting sequence was then blurred with an uniform 3×3 mask and downsampled to a factor of 2, resulting in LR images of dimension $N = 128$. Finally, white Gaussian noise with variance $\sigma^2 = 10$ was added to the downsampled images. Circulant boundary conditions were used for simplicity.

We applied both standard and regularized versions of the LMS and the proposed algorithm (11), all initialized with a bicubic interpolation of the first LR image, and used $K = 2$ iterations per time instant. The super resolved sequences were compared to the original one in HR and the mean squared error (MSE) was computed across all realizations. The traditional temporal regularization [3] did not present any improvement over the R-LMS algorithm and was therefore omitted from the comparison. The parameters were selected by exhaustive search over a small set of images for the algorithms to achieve their best performance in steady state (for large t) considering known motion, and are shown in Table I.

We first evaluate the algorithms without the influence of innovation outliers in order to assess the quality of the reconstructed images. For a more realist evaluation, the motion between different frames was estimated using the *Horn & Schunk*

ISBN 978-0-9928626-7-1 © EURASIP 2017

	LMS	R-LMS	Algorithm (11)
μ	2	2.75	3
α	–	5×10^{-4}	1×10^{-4}
α_T	–	–	0.02

TABLE I
PARAMETER VALUES USED IN THE SIMULATIONS.

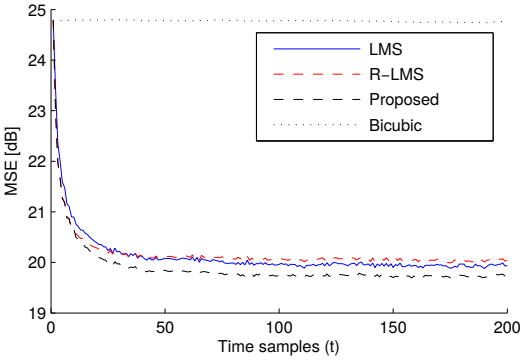

Fig. 2. Mean square error evolution with estimated motion.

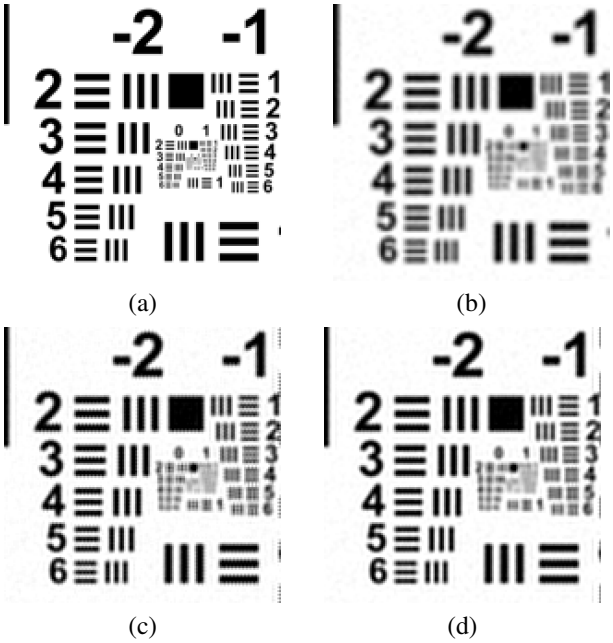

Fig. 3. Sample of the 200th reconstructed frame (without outliers). (a) Original image. (b) Bicubic interpolation (MSE=30.13dB). (c) R-LMS algorithm (MSE=25.54dB). (d) proposed method (MSE=25.47dB).

registration algorithm provided in [15], [16], with the motion vectors averaged over the entire image to result in global translational displacements. The MSE performance is depicted on Figure 2. It can be seen that the algorithms performed similarly, although a slight improvement was noticed for the proposed method. A sample of the reconstructed images can be seen in Figure 3. Although the perceptual quality is similar for the algorithms being compared, a careful evaluation reveals a slight improvement for the proposed method (the LMS result was similar to that of the R-LMS and was thus omitted due to space limitation).

In order to evaluate the performance of the proposed algorithm under the influence of outliers, the previous simulation was repeated considering the inclusion of a suddenly appearing object, independent from the background. This was performed by including an $N \times N$ black square to the middle of the 32nd frame of every sequence. The black square remained on the scene until the 35th frame, when it disappears, emulating the behavior of a *flying bird* outlier. The MSE evolution is depicted in Figure 4. It can be seen that the proposed method offered a significant performance gain when compared to the remaining algorithms in the presence of outliers, between frames 32 and 35. A visual inspection of the reconstructed images as portrayed in Figure 5 supports the quantitative result, since the black square introduced in the sequence is significantly better represented for the proposed method (when it is indeed present in the HR image). This indicates that the proposed method is more robust to outliers.

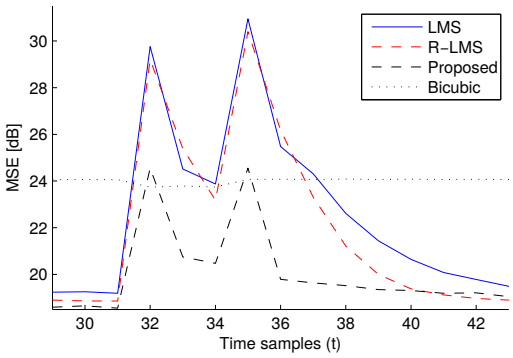

Fig. 4. Mean square error evolution with outlier.

VII. CONCLUSIONS

In this paper, a new super-resolution reconstruction algorithm with improved robustness to innovation outliers was proposed. Trough the proximal point cost function associated with the R-LMS update equation, its dynamical behavior was studied, revealing a conflict between tracking the innovations and preserving estimated content. To address this problem, a new regularization was then proposed by constraining changes in the details of the estimated image. Simulation results showed a significant improvement in the robustness of the algorithm to large innovations, while a similar image quality was observed otherwise.

REFERENCES

[1] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang, "Super-resolution image reconstruction: a technical overview," *Signal Processing Magazine, IEEE*, vol. 20, no. 3, pp. 21–36, 2003.
[2] Kamal Nasrollahi and Thomas B. Moeslund, "Super-resolution: A comprehensive survey," *Machine Vision & Applications*, vol. 25, no. 6, pp. 1423–1474, 2014.
[3] Sean Borman and Robert L Stevenson, "Simultaneous multi-frame map super-resolution video enhancement using spatio-temporal priors,"

ISBN 978-0-9928626-7-1 © EURASIP 2017

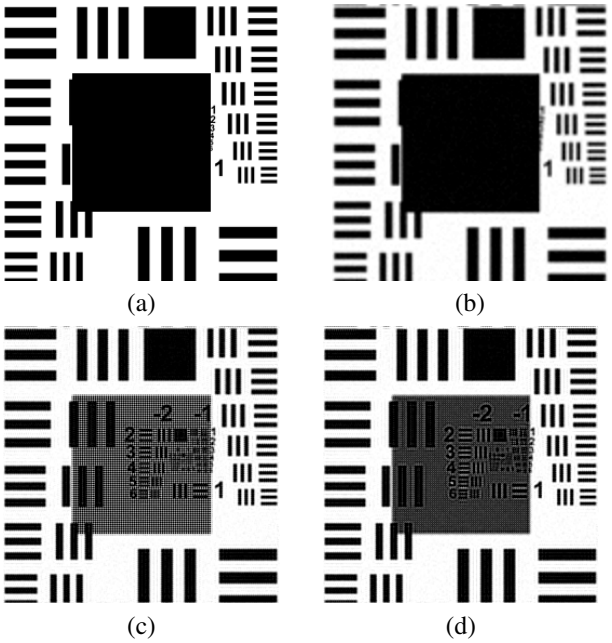

Fig. 5. Sample of the 34th reconstructed frame (with an outlier). (a) Original image. (b) Bicubic interpolation (MSE=30.32dB). (c) R-LMS algorithm (MSE=34.20dB). (d) proposed method (MSE=30.02dB).

in *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*. IEEE, 1999, vol. 3, pp. 469–473.

[4] Jing Tian and Kai-Kuang Ma, "A state-space super-resolution approach for video reconstruction," *Signal, image and video processing*, vol. 3, no. 3, pp. 217–240, 2009.

[5] Marcelo Victor Wüst Zibetti and Joceli Mayer, "A robust and computationally efficient simultaneous super-resolution scheme for image sequences," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 10, pp. 1288–1300, 2007.

[6] Stefanos P Belekos, Nikolas P Galatsanos, and Aggelos K Katsaggelos, "Maximum a posteriori video super-resolution using a new multichannel image prior," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1451–1464, 2010.

[7] M. Elad and A. Feuer, "Superresolution restoration of an image sequence: Adaptive filtering approach," *Trans. Img. Proc., IEEE*, vol. 8, no. 3, pp. 387–395, Mar. 1999.

[8] Sina Farsiu, M Dirk Robinson, Michael Elad, and Peyman Milanfar, "Fast and robust multiframe super resolution," *Image processing, IEEE Transactions on*, vol. 13, no. 10, pp. 1327–1344, 2004.

[9] Guilherme Holsbach Costa and José Carlos M Bermudez, "Statistical analysis of the lms algorithm applied to super-resolution image reconstruction," *Signal Processing, IEEE Transactions on*, vol. 55, no. 5, pp. 2084–2095, 2007.

[10] Guilherme Holsbach Costa and José Carlos M Bermudez, "Registration errors: Are they always bad for super-resolution?," *Signal Processing, IEEE Transactions on*, vol. 57, no. 10, pp. 3815–3826, 2009.

[11] Michael Elad and Arie Feuer, "Super-resolution reconstruction of image sequences," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 9, pp. 817–834, 1999.

[12] Rafael C Gonzalez and Richard E Woods, *Digital image processing*, Prentice hall, 1 edition, 2002.

[13] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[14] Dimitri P Bertsekas, *Nonlinear programming*, Athena scientific Belmont, 1999.

[15] Deqing Sun, Stefan Roth, and Michael J Black, "Secrets of optical flow estimation and their principles," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2432–2439.

[16] Berthold KP Horn and Brian G Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

ISBN 978-0-9928626-7-1 © EURASIP 2017