UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

Ricardo Augusto Borsoi

A NEW ADAPTIVE ALGORITHM FOR VIDEO SUPER-RESOLUTION WITH IMPROVED OUTLIER HANDLING CAPABILITY

Florianópolis

2016

Ricardo Augusto Borsoi

A NEW ADAPTIVE ALGORITHM FOR VIDEO SUPER-RESOLUTION WITH IMPROVED OUTLIER HANDLING CAPABILITY

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica para a obtenção do Grau de Mestre em Engenharia Elétrica.
Orientador: Prof. José Carlos Moreira Bermudez, Ph.D.

Florianópolis

2016

Catalogação na fonte pela Biblioteca Universitária da
Universidade Federal de Santa Catarina

Borsoi, Ricardo Augusto
A New Adaptive Algorithm for Video Super-Resolution with Improved Outlier Handling Capabiliy / Ricardo Augusto Borsoi ; orientador, José Carlos Moreira Bermudez - Florianópolis, SC, 2016.
87 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia Elétrica.

Inclui referências

1. Engenharia Elétrica. 2. Processamento de imagens. 3. Super-resolução. 4. Algoritmos adaptativos. I. Bermudez, José Carlos Moreira. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

Ricardo Augusto Borsoi

A NEW ADAPTIVE ALGORITHM FOR VIDEO SUPER-RESOLUTION WITH IMPROVED OUTLIER HANDLING CAPABILITY

Esta Dissertação foi julgada aprovada para a obtenção do Título de "Mestre em Engenharia Elétrica", e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Elétrica.

Florianópolis, 30 de novembro 2016.

Prof. Dr. Marcelo Lobo Heldwein
Coordenador
Universidade Federal de Santa Catarina

Banca Examinadora:

Prof. José Carlos Moreira Bermudez, Ph.D.
Orientador

Prof. Guilherme Holsbach Costa, Dr.

Prof. Márcio Holsbach Costa, Dr.

Prof. Joceli Mayer, Ph.D.

Prof. Leonardo Koller Sacht, Dr.

Dedicated $\chi\acute{\alpha}o\varsigma$.

Knowledge $\perp$ reason.

$\zeta\epsilon\delta \subset M42$.
$C99 \subset \dagger$.
$V1489 \subset \kappa\acute{\upsilon}\kappa\nu o\varsigma$.

AGRADECIMENTOS

Ao professor José Carlos Moreira Bermudez pela orientação e empenho durante o desenvolvimento do trabalho.

Ao professor Guilherme Holsbach Costa, pela orientação, amizade e apoio ao longo dos anos.

Aos meus pais, Francisco e Rosa, pelo imenso carinho, compreensão e apoio, indispensável à cada passo deste caminho.

Aos professores do Programa de Pós-Graduação em Engenharia Elétrica pelo ambiente de aprendizado.

Aos funcionários da UFSC: Wilson Silva Costa e Marcelo Manoel Siqueira.

À CAPES pela oportunidade, incentivo e financiamento.

Aos colegas de laboratório, sem necessidade de listar razões.

To recognise untruth as a condition of life, that is certainly to impugn the traditional ideas of value in a dangerous manner, and a philosophy which ventures to do so, has thereby alone placed itself beyond good and evil.

Friedrich Nietzsche

ABSTRACT

Super resolution reconstruction (SRR) is a technique that consists basically in combining multiple low resolution images from a single scene in order to create an image with higher resolution. The main characteristics considered in the evaluation of SRR algorithms performance are the resulting image quality, its robustness to outliers and its computational cost. Among the super resolution algorithms present in the literature, the R-LMS has a very small computational cost, making it suitable for real-time operation. However, like many SRR techniques the R-LMS algorithm is also highly susceptible to outliers, which can lead the reconstructed image quality to be of lower quality than the low resolution observations. Although robust techniques have been proposed to mitigate this problem, the computational cost associated with even the simpler algorithms is not comparable to that of the R-LMS, making real-time operation impractical. It is therefore desirable to devise new algorithms that offer a better compromise between quality, robustness and computational cost. In this work, a new SRR technique based on the R-LMS algorithm is proposed. Based on the proximal-point cost function representation of the gradient descent iterative equation, an intuitive interpretation of the R-LMS algorithm behavior is obtained, both in ideal conditions and in the presence of outliers. Using a statistical model for the innovation outliers, a new regularization is then proposed to increase the algorithm robustness by allowing faster convergence on the subspace corresponding to the innovations while at the same time preserving the estimated image details. Two new algorithms are then derived. Computer simulations have shown that the new algorithms deliver a performance comparable to that of the R-LMS in the absence of outliers, and a significantly better performance in the presence of outliers, both quantitatively and visually. The computational cost of the proposed solution remained comparable to that of the R-LMS.

Keywords: image processing, super resolution, R-LMS, outliers, innovations.

RESUMO ESTENDIDO

Reconstrução com super resolução (SRR - *Super resolution reconstruction*) é uma técnica que consiste basicamente em combinar múltiplas imagens de baixa resolução a fim de formar uma única imagem com resolução superior. As principais características consideradas na avaliação de algoritmos de SRR são a qualidade da imagem reconstruída, sua robustez a *outliers* e o custo computacional associado.

Uma maior qualidade nas imagens reconstruídas implica em um maior aumento efetivo na resolução das mesmas. Uma maior robustez, por outro lado, implica que um resultado de boa qualidade é obtido mesmo quando as imagens processadas não seguem fielmente o modelo matemático adotado. O custo computacional, por sua vez, é extremamente relevante em aplicações de SRR, dado que a dimensão do problema é extremamente grande.

Uma das principais aplicações da SRR consiste na reconstrução de sequências de vídeo. De modo a facilitar o processamento em tempo real, o qual é um requisito frequente para aplicações de SRR de vídeo, algorítmos iterativos foram propostos, os quais processam apenas uma imagem a cada instante de tempo, utilizando informações presentes nas estimativas obtidas em instantes de tempo anteriores. Dentre os algoritmos de super resolução iterativos presentes na literatura, o R-LMS possui um custo computacional extremamente baixo, além de fornecer uma reconstrução com qualidade competitiva. Apesar disso, assim como grande parte das técnicas de SRR existentes o R-LMS é bastante suscetível a presença de *outliers*, os quais podem tornar a qualidade das imagens reconstruídas inferior àquela das observações de baixa resolução.

A fim de mitigar esse problema, técnicas de SRR robusta foram propostas na literatura. Não obstante, mesmo o custo computacional dos algoritmos robustos mais simples não é comparável àquele do R-LMS, tornando o processamento em tempo real infactível. Deseja-se portanto desenvolver novos algoritmos que ofereçam um melhor compromisso entre qualidade, robustez e custo computacional.

Neste trabalho uma nova técnica de SRR baseada no algoritmo R-LMS é proposta. Com base na representação da função custo do ponto proximal para a equação iterativa do método do gradiente, uma interpretação intuitiva para o comportamento do algoritmo R-LMS é obtida tanto para sua operação em condições ideais quanto na presença

de *outliers* do tipo inovação, os quais representam variações significativas na cena entre frames adjacentes de uma sequência de vídeo.

É demonstrado que o problema apresentado pelo R-LMS quanto a robustez à *outliers* de inovação se deve, principalmente, a sua baixa taxa de convergência. Além disso, um balanço direto pôde ser observado entre a rapidez da taxa de convergência e a preservação das informações estimadas em instantes de tempo anteriores. Desse modo, torna-se inviável obter, simultaneamente, uma boa qualidade no processamento de sequencias bem comportadas e uma boa robustez na presença de inovações de grande porte.

Desse modo, tem-se como objetivo projetar um algoritmo voltado à reconstrução de sequências de vídeo em tempo real que apresente uma maior robustez à outliers de grande porte, sem comprometer a preservação da informação estimada a partir da sequência de baixa resolução.

Utilizando um modelo estatístico para os *outliers* provindos de inovações, uma nova regularização é proposta a fim de aumentar a robustez do algoritmo, permitindo simultaneamente uma convergência mais rápida no subespaço da imagem correspondente às inovações e a preservação dos detalhes previamente estimados. A partir disso dois novos algoritmos são então derivados.

A nova regularização proposta penaliza variações entre estimativas adjacentes na sequência de vídeo em um subespaço aproximadamente ortogonal ao conteúdo das inovações. Verificou-se que o subespaço da imagem no qual a inovação contém menos energia é precisamente onde estão contidos os detalhes da imagem. Isso mostra que a regularização proposta, além de levar a uma maior robustez, também implica na preservação dos detalhes estimados na sequência de vídeo em instantes de tempo anteriores.

Simulações computacionais mostram que apesar da solução proposta não levar a melhorias significativas no desempenho do algoritmo sob condições próximas às ideais, quando *outliers* estão presentes na sequência de imagens o método proposto superou consideravelmente o desempenho apresentado pelo R-LMS, tanto quantitativamente quanto visualmente. O custo computacional da solução proposta manteve-se comparável àquele do algoritmo R-LMS.

Palavras-chave: processamento de imagens, super resolução, R-LMS, *outliers*, inovações

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

LIST OF SYMBOLS

CONTENTS

1 INTRODUCTION

One of the most important properties of a digital image is its quality. While the meaning of quality can be different depending on the context and application, its resolution is certainly one of the most important factors in defining quality. Image resolution can be roughly defined as the amount of detail that is contained in an image. Images with a higher resolution are desired for a multitude of applications, ranging from end user applications to forensics and machine vision. Nevertheless, physical image acquisition systems are frequently under constraints that limit the captured image resolution. For instance, satellite images used in remote sensing are acquired at a large distance from the desired scene, and the optical system faces physical limitations on the diffraction limit of the lenses. Furthermore, increasing the sensor pixel density is not a plausible solution, both due to economical constraints and because reducing the pixel size leads to a reduction of the signal to noise ratio (TIAN; MA, 2011). Besides, it might be of interest to increase the resolution of images which have been already acquired under unfavorable circumstances.

One technique that is becoming increasingly popular to overcome the limitations of physical imaging systems is super resolution reconstruction (SRR). Super resolution reconstruction consists basically in combining multiple low resolution images in order to obtain one or more images of higher resolution. This process is performed after the images are acquired, allowing for an effective increase in the resolution without the need to increase the sensor pixel density. The applications of SRR are numerous, including (PARK; PARK; KANG, 2003; COSTA, 2007):

- General purpose low cost imaging systems: applying super resolution techniques to images acquired by general end-user cameras can allow good resolution images to be obtained even when using low cost digital imaging sensors. Since the exposition time of most cameras is only in the order of a small fraction of a second, multiple images can be easily acquired to be later combined into a higher resolution image without letting the user experience a noticeable delay.

- Medical imaging systems, such as Computed Tomography (CT) and Magnetic Resonance imaging (MRI) can also benefit from resolution enhancement since the acquisition of multiple images is possible, and the availability of a greater amount of detail allows

a more reliable diagnosis to be obtained.

- In remote sensing applications, multiple satellite images of the same location are frequently available, which consequently makes it possible to apply super resolution techniques to improve the amount of details of a target being considered through the synthetic zooming of a specific region of interest.

- The reliability of video based surveillance systems and forensics is greatly dependent on the resolution of the available images, where objects such as the license plate of a car or the face of an individual frequently need to be magnified, and it is important to have a good amount of detail. Since in most surveillance systems the images are acquired and stored for future analysis, they can be processed afterwards, which allows super resolution techniques to be applied. This allows a good reliability to be obtained with less expensive systems, reducing costs both in the imaging sensor and storage equipment.

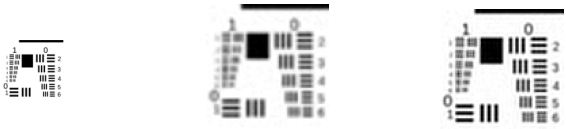
- Modern video communication systems like those based on voice over Internet Protocol (VoIP) frequently relies on a limited bandwidth. A smaller bandwidth demand can be obtained if the images resolution is increased in the end-user application and transmitted at a lower bit-rate. Furthermore, super resolution can also be applied in the conversion from standard to high definition television signals in order for those to be displayed in high definition receivers without visual artifacts.

- Pattern recognition systems are frequently employed both for surveillance and quality control, where it is desired to identify specific objects in the scene or check if a product satisfies some required specifications. The reliability of those systems can be improved through the use of super resolution techniques, which can also allow the use of less expensive equipments.

1.1 BASIC WORKING PRINCIPLE

Super resolution reconstruction manages to increase an image resolution by extracting non-redundant information from a set of low resolution (LR) images of the same object or scene. This can be performed since there are usually small differences between each of the LR images. This fact differentiates SRR from Interpolation algorithms

which, despite also increasing an image size (number of pixels), are based on the hallucination of its details, since only one single frame is employed. This difference is portrayed in Figure 1. Since SRR algorithms combine multiple images when reconstructing some specific frame, they belong to a wider class of techniques known as *frame fusion* (CAPEL, 2004), which also includes, for example, image mosaicing and denoising.



LR Bicubic Interpolation SRR

Figure 1 – Difference between super resolution and interpolation for a resolution enhancement factor of 2.

The presence of information in a set of LR images that goes beyond their individual sampling rate occurs due to the fact that each image is acquired from a position slightly different from the others, which implies that the scene or object of interest is sampled at different positions. This displacement may be originated, for example, from vibration of the camera during the acquisition of a video sequence. This way, the super resolution can also be generally seen as an inverse problem, in which we want to "reverse" the negative effects originating from the limitations of the imaging sensor.

With images acquired at different positions, their relative displacement must be known in advance before the reconstruction can take place. This fact makes SRR to be often divided in two distinct steps: the registration or alignment of the images, followed by the fusion of the low resolution images into a higher resolution one. An illustration of the alignment process can be seen in Figure 2. This work will exclusively address the image fusion step of the SRR process.

1.2 VIDEO SRR: REAL TIME SOLUTIONS AND ROBUSTNESS TO OUTLIERS

Super resolution techniques can be further divided in two major groups: the reconstruction of a single frame from various observations, and the reconstruction of an entire video sequence. While the former

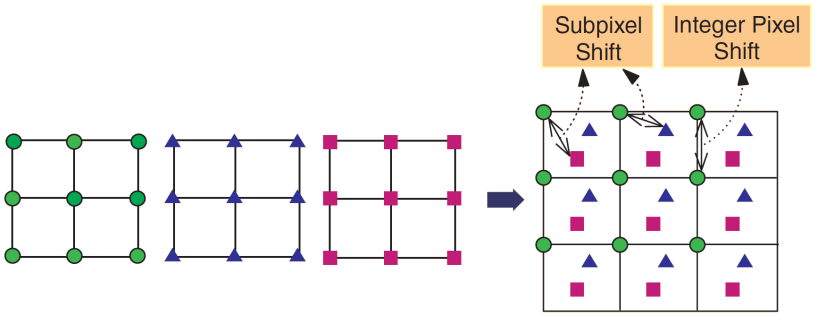


Figure 2 – Registration (alignment) of the LR images (taken from (PARK; PARK; KANG, 2003)).

is more concerned with the quality of the reconstructed images, most applications of video sequence SRR impose the need for real time processing.

Although it is possible to apply single frame reconstruction methods to super resolve a video sequence one frame at a time, methods aimed specifically at this problem have been developed (see for instance (BORMAN; STEVENSON, 1999; ZIBETTI; MAYER, 2007; BELEKOS; GALATSANOS; KATSAGGELOS, 2010)). Particularly, a class of methods based on iterative algorithms have been developed for real time processing applications, reconstructing one frame at a time and using the information contained in the previous estimations, providing a reduced computational cost, although their quality is generally inferior to that of the methods which reconstruct the entire video sequence simultaneously.

Most iterative SRR methods developed consist of approximations of the Kalman filter, many of which are based on the works of Elad et al. (ELAD; FEUER, 1999b, 1999a). Examples of such algorithms include the steepest descent (SD), and the regularized least mean squares (R-LMS) algorithm, which have one of the lowest computational cost among SRR algorithms.

1.2.1 Robustness

One of the greatest problems with SRR algorithms is their sensitivity to modeling errors. The performance of most methods when processing data that does not follow the assumed models degrades quickly,

frequently leading to reconstructed sequences of worse quality than that of the observations themselves (ZIBETTI; MAYER, 2007). This kind of data is called an *outlier*, and can be originated, for example, due to inaccurate image registration or sudden changes in the scene. The latter case, denominated innovation outliers, is a frequent problem in video sequences due to the presence of moving objects from one frame to the next or due to the motion of the camera, which changes the region of the scene acquired by the sensor.

Innovation outliers can lead to significant artifacts in the reconstructed sequence. An illustrative example of the effect of innovations on SRR results can be seen in Figure 3, which depicts the quality improvement achieved by an algorithm with a special treatment for the innovations when applied to super-resolve a LR video sequence.

(a) (b) (c) (d) (e)

Figure 3 – Example of innovations' effect on reconstruction (taken from (COSTA, 2007)). (a) HR image. (b) LR image (decimated by a factor of 2). (c) Interpolation. (d) SRR without innovations treatment. (e) SRR with innovation treatment.

Numerous algorithms have been developed in order to provide increased robustness to innovations and registration errors, which are frequently addressed uniquely under the guise of *motion errors*. Although the performance of those methods is very good, the computational cost associated with even the faster algorithms like that of Farsiu et al. (2004b) is almost prohibitive for many applications, and is not comparable to that of the sequential methods. This sets a conflicting and difficult trade-off between achieving the desired quality and robustness and meeting the processing speed demanded for a specific application.

In the context of iterative algorithms, robust methodologies have also been employed by some authors in the literature. Wang et al. (WANG; QI, 2002) proposed a modified Kalman filter that accounted for the presence of registration errors. Kim et al. (KIM et al., 2010) extended the methodology of Farsiu et al. (2004b, 2006) by detecting and removing previously estimated pixels that deviate significantly from

the presently observed image, and reinitializing the algorithm in the case of large scene changes. Su et al. (SU; WU; ZHOU, 2011) proposed a regularized version of a moving least squares algorithm which, for improved robustness, used non-quadratic norms (e.g. L_1) on the cost function, following an approach based on robust estimators which is widely employed in SRR (FARSIU et al., 2004b), besides weighing the regularization based on the estimated level of registration errors.

Although these methods provide good robustness and quality, they diverge from the objectives pertaining real time operation since the use of nonlinear methodologies or penalty functions leads to an increased computational cost, especially when compared with simpler algorithms like the R-LMS. However, the R-LMS is limited by an unfavorable trade-off between achieving good robustness or quality, which reduces its applicability in real-world situations. Therefore, it is of great interest to devise new methodologies that although more robust, still preserve a small associated computational cost and an acceptable image quality, being better suited for real time operation.

1.3 OBJECTIVES AND OUTLINE

Considering what was presented in the previous section, in this work a new super resolution reconstruction method for real-time applications is proposed. Based on the R-LMS algorithm formulation, a new cost function is proposed by simultaneously incorporating the objectives of robustness and quality. The result is a new algorithm that is less influenced by innovation outliers and, at the same time, maintains both a good reconstruction quality in their absence and a small computational cost.

In Chapter 2, the notation used in this work is presented. A mathematical formulation for the image acquisition and super resolution processes is discussed, and the reconstruction of video sequences is addressed in greater detail. The R-LMS algorithm is derived following the formulation used in (COSTA; BERMUDEZ, 2007), and robust approaches for the SRR problem are also discussed.

In Chapter 3, the R-LMS algorithm is represented through the proximal-point cost function of the gradient descent iterative equation. This allows for an intuitive interpretation of its behavior both in ideal conditions and in the presence of outliers. A new regularization is proposed in chapter 4 aiming to increase robustness by preserving the solution details during the reconstruction, while at the same time al-

lowing a subspace containing the innovation outliers to change more freely in order to achieve faster convergence. A statistical model for the innovations is proposed and two new algorithms are derived, one of which is shown to be equivalent to using a generalized version of a regularization already studied in the literature.

In Chapter 5 computer simulations are performed in order to assess the algorithm performance (both quantitatively and visually). The proposed algorithms significantly outperformed the R-LMS algorithm in the presence of outliers using both synthetic and real video sequences, while providing equivalent performance in the case of small or no outliers.

Finally, Chapter 6 concludes this work, and some suggestions for future improvements are also discussed.

2 LITERATURE REVIEW

This chapter is organized as follows. First the image acquisition and dynamics process is mathematically described. Afterwards, the video SRR process is formally defined in the form of an inverse problem. Simultaneous and sequential video SRR algorithms are then presented, with an emphasis on low cost methods and on the R-LMS algorithm, which will be employed in later chapters. Finally, the problem of robustness is discussed.

Through this work, the (R)-LMS refers to the (regularized) least mean squares algorithm applied to SRR instead of the one dimensional filtering algorithm, unless explicitly stated otherwise. Vectors are denoted by bold lower-case letters, and matrices by bold upper-case letters. The variable t is integer and is used to represent the discrete time scale. Vector norms are denoted by $\|\cdot\|$, and unless specified represent the L_2 norm.

2.1 IMAGE ACQUISITION MODEL

The signal model generally describes the distortions an image suffers while being captured by a digital sensor due to its inherent physical limitations. This process is illustrated in Figure 4.

Figure 4 – Digital image acquisition process.

In order to formally represent the acquisition model illustrated in Figure 4, it is possible to consider the high resolution (HR) images as continuous functions on the spatial domain. Nevertheless, given that in practice images can be well represented as band-limited functions, most authors define the acquisition model considering a discrete high-resolution image (supposedly sampled at a higher rate), resulting in a mathematically simpler model.

Given the matrix form representation of a digital HR image $\mathbf{X}(t)$ of size $M \times M$ and a LR image $\mathbf{Y}(t)$ of size $N \times N$, the image acquisition

process, consisting in the mapping $\mathbf{X}(t) \mapsto \mathbf{Y}(t)$, can be modeled as (PARK; PARK; KANG, 2003):

$$\mathbf{y}(t) = \mathbf{DHx}(t) + \mathbf{e}(t)\,, \tag{2.1}$$

where $\mathbf{x}(t)$ and $\mathbf{y}(t)$, of dimension $M^2 \times 1$ and $N^2 \times 1$, are the lexicographic representations of the original (HR) and degraded (LR) images, respectively, for the time instant t. The lexicographic representation consists in a mapping $\mathbb{R}^{M\times M} \rightarrow \mathbb{R}^{M^2}$ that reorders the matricial signals in the form of vectors, leading to the representation of the acquisition process in the form of matrix-vector multiplications. The reordering is performed as:

$$\mathbf{X}(t) = \begin{bmatrix} \mathbf{x}_1^{\mathrm{T}}(t) \\ \mathbf{x}_2^{\mathrm{T}}(t) \\ \vdots \\ \mathbf{x}_M^{\mathrm{T}}(t) \end{bmatrix} \quad \Rightarrow \quad \mathbf{x}(t) = \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \\ \vdots \\ \mathbf{x}_M(t) \end{bmatrix} \tag{2.2}$$

where $\mathbf{x}_i^{\mathrm{T}}(t)$ is the $i-$th line of matrix $\mathbf{X}(t)$. The matrix $\mathbf{H}$ in (2.1), of dimension $M^2 \times M^2$, models distortions due to the optical system, which usually consists in a blurring represented through a spatially invariant convolution. Matrix $\mathbf{D}$, of dimension $N^2 \times M^2$, models the downsampling that occurs in the sensor. Both optical distortions and downsampling are assumed to be time-invariant, without loss of generality, since the extension for the time-varying case with $\mathbf{D}(t)\mathbf{H}(t)$ is trivial and will therefore be omitted in favor of notational simplicity.

Vector $\mathbf{e}(t)$, of dimension $N^2 \times 1$, models additive noise generated in the imaging sensor, usually assumed to be white and zero mean, with isotropic variance given by σ_e^2 (COSTA; BERMUDEZ, 2007).

It is important to notice that the acquisition model in (2.1) only supports linear transformations over the image $\mathbf{x}(t)$. Other works in the literature consider more complex effects such as sensor saturation, which requires a nonlinear imaging model (GUNTURK; GEVREKCI, 2006). These effects, however, only become significant in specific applications (e.g. reconstruction of images with different exposition times), which leads a linear model to be general enough for the applications considered in the present work.

Given that the super-resolution process uses non-redundant information of the different images $\mathbf{y}(t)$ contained in their relative displacement, the relationship between them must also be characterized. Since this work addresses the reconstruction of video sequences, a dy-

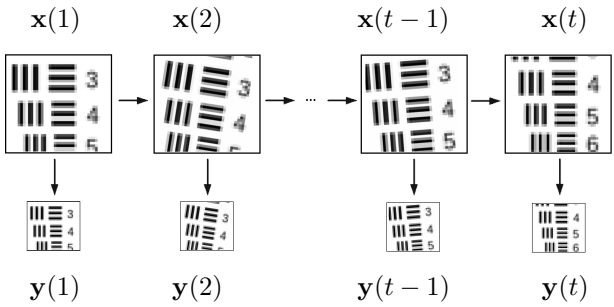


Figure 5 – Illustration of the acquisition and dynamical signal models.

namical model describing the temporal evolution of the signal $\mathbf{x}(t)$ will be employed, given by (ELAD; FEUER, 1999b):

$$\mathbf{x}(t) = \mathbf{G}(t)\mathbf{x}(t-1) + \mathbf{s}(t)\,, \tag{2.3}$$

where $\mathbf{G}(t)$, of dimension $M^2 \times M^2$, represents the displacement between the HR images from time instant $t-1$ to time instant t, and $\mathbf{s}(t)$, of dimension $M^2 \times 1$, represents the innovations that took place on the scene, which are composed by the part of $\mathbf{x}(t)$ that is statistically orthogonal to $\mathbf{x}(t-1)$, such as border effects and occlusions. When combined, the model represented by equations (2.3) and (2.1) results in the description of the HR and LR video sequences, as illustrated in Figure 5.

Given that sub-pixel displacement between the LR images $\mathbf{y}(t)$ is necessary in order to assure the presence of non-redundant information in the sequence, the motion estimation consists in an essential part of the super-resolution process, being present in the estimation of the matrix $\mathbf{G}(t)$. This matrix attributes to the $n-$th pixel of $\mathbf{x}(t)$ the intensity corresponding to a determined position (not necessarily integer) of $\mathbf{x}(t-1)$. This way, a discrete interpolation is performed to characterize the correspondence between pairs of points in the two images which are related by their relative displacement. The motion can be either unique for each pixel and defined through a vector field known as a dense optical flow, or follow global motion models, such as rotation or translation, which are defined by a reduced number of parameters (CAPEL, 2004).

The matrix $\mathbf{G}(t)$ as defined above implicitly assumes what is known as the brightness constancy hypothesis, which considers that

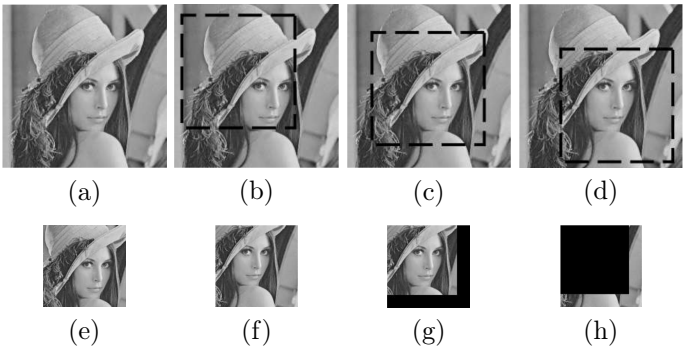


Figure 6 – Illustration of border effect innovations for global translational motion (taken from (COSTA, 2007)). (a) Original scene. (b)-(d) HR images in time instants $t-2$, $t-1$ and t (dashed) relative to the scene. (e) $\mathbf{x}(t-1)$. (f) $\mathbf{x}(t)$. (g) Registered image $\mathbf{G}(t)\mathbf{x}(t-1)$, considering Dirichlet boundary condition. (h) Innovations $\mathbf{s}(t)$.

the light intensity of the pixels in the scene does not change over time. This hypothesis is not respected in practical situations, since unpredictable changes usually happen in the scene, given either by illumination changes or by the appearance of objects in the image which were not previously visible, like borders effects or occlusions. These pixels consist in the innovations on the scene, being represented in $\mathbf{s}(t)$.

The innovations can also be divided in two main types, those due to border effects and those due to occlusions. Border innovations, which are depicted in Figure 6, come from the problem of both the image and the motion field vectors being only defined for a finite domain consisting of the ensemble of points $[1, 2, \cdots, N] \times [1, 2, \cdots, N]$. Depending on the displacement, pixels on the boundary of the image $\mathbf{x}(t)$ may not have a correspondence with any pixels in $\mathbf{x}(t-1)$, but instead with some unknown point beyond its boundary. This can be easily noticed when assuming a global translational motion model. Given that either the scene intensity or the displacement is unknown beyond the domain of $\mathbf{x}(t-1)$, determining the values of $\mathbf{x}(t)$ corresponding to these points requires the use of some boundary condition. Examples of boundary conditions frequently employed are the zero padding (Dirichlet), circular replication or mirroring (Von Neumann) (MODERSITZKI, 2003).

On the other hand, in the innovations originating due to oc-

(a) (b) (c) (d)

Figure 7 – Illustration of innovations originating from occlusions (a bird flying over a static scene). (a) Motion ($\mathbf{G}(t)$). (b)-(c) HR images on time instants $t-1$ and t. (d) Innovations ($\mathbf{s}(t)$).

clusions, pixels occluded in going from $\mathbf{x}(t-1)$ to $\mathbf{x}(t)$ have no spatial correspondence. The displacement vectors for these points are not even defined (AYVACI; RAPTIS; SOATTO, 2012). The motion in these points can be either extrapolated or hallucinated based on its boundary, or the corresponding pixels can be padded with zeros (in a condition similar to Dirichlet's). In this case, pixels of the registered image $\mathbf{G}(t)\hat{\mathbf{x}}(t-1)$ in the areas corresponding to occlusions will either become black or contain an interpolation of its neighboring pixels (which is more common due to the employment of simple image registration algorithms). An example of this type of innovation is illustrated in Figure 7.

It is also important to distinguish the innovations represented in $\mathbf{s}(t)$ from registration errors. Since the motion estimation process is a severely ill-posed inverse problem, the obtained estimate for the registration matrix $\hat{\mathbf{G}}(t)$ is imprecise, resulting in an additional source of errors in equation (2.3). Representing the error in the estimate of the registration matrix as $\Delta\mathbf{G}(t)$, the dynamical signal model considering unknown motion can be written as:

$$\begin{aligned}\mathbf{x}(t) &= \hat{\mathbf{G}}(t)\mathbf{x}(t-1) + \mathbf{s}(t) \\ &= \mathbf{G}(t)\mathbf{x}(t-1) + \Delta\mathbf{G}(t)\mathbf{x}(t-1) + \mathbf{s}(t)\,.\end{aligned}$$

Some authors treat the error signal composed by $\Delta\mathbf{G}(t)\mathbf{x}(t-1) + \mathbf{s}(t)$ uniquely in the form of *motion errors*, although acknowledging the presence of low magnitude errors ($\Delta\mathbf{G}(t)\mathbf{x}(t-1)$) and high magnitude errors, also known as outliers ($\mathbf{s}(t)$) (FARSIU et al., 2004b; ZIBETTI; MAYER, 2007). Since the characteristics of registration errors are significantly different from those of innovations, they will not be directly addressed when devising the new algorithm in this work. All points for which the video sequence does not follow the brightness

constancy hypothesis are denominated *outliers*.

2.2 THE INVERSE PROBLEM (RECONSTRUCTION)

Super-resolution reconstruction generally consists in solving an inverse problem, where given an ensemble of low resolution observations $\mathbf{y}(1), \cdots, \mathbf{y}(T)$ and the signal model in (2.1) and (2.3), it is desired to determine an estimate $\hat{\mathbf{x}}(t)$ of the high resolution image $\mathbf{x}(t)$ for all t.

Given that most SRR methods developed are aimed at the reconstruction of a single image $\hat{\mathbf{x}}$, a natural extension of these methodologies to the reconstruction of video sequences can be made by iteratively reconstructing each HR image on the sequence using a sliding observation window $\mathbf{y}(t-l), \cdots, \mathbf{y}(t+l)$ for the estimations of each $\hat{\mathbf{x}}(t)$. As this approach results in an excessively high computational cost, super resolution of video sequences has been mainly addressed in the literature using more effective approaches.

On the other hand, in algorithms specifically developed towards the reconstruction of video sequences, all the images are estimated during the super resolution process ($\hat{\mathbf{x}}(1), \cdots, \hat{\mathbf{x}}(T)$). These methods are further divided in two major classes: the *simultaneous* methods, which reconstruct the whole sequence at once, and the *sequential* methods, which reconstruct one image at a time, using the information estimated in the previous time instants (ZIBETTI; MAYER, 2007). These two classes will be treated in detail in the next section.

2.2.1 Simultaneous Methods

In this class of methods the reconstruction of the images at all time instants is performed simultaneously. This is done with the objective of attaining a better quality, although at the expense of a higher computational cost. Most approaches for the simultaneous video SRR are based in the minimization of a cost function of the form (ZIBETTI; MAYER, 2007):

$$\hat{\mathbf{x}}(1), \cdots, \hat{\mathbf{x}}(T) = \underset{\hat{\mathbf{x}}(1),\cdots,\hat{\mathbf{x}}(T)}{\arg\min} \sum_{t=1}^{T}\sum_{j=1}^{T} \underbrace{\|\mathbf{y}(t) - \mathbf{DH}\,\mathbf{M}_{t,j}\,\hat{\mathbf{x}}(j)\|_p^q}_{\text{Data}} + \alpha_{\mathfrak{R}} \underbrace{\mathfrak{R}\big(\hat{\mathbf{x}}(1), \cdots, \hat{\mathbf{x}}(T)\big)}_{\text{Regularization}} \quad (2.4)$$

where $\mathbf{M}_{t,j}$ represents the cumulative motion from frame j up to frame t, given by $\prod_{l=0}^{j+1} \mathbf{G}(t-l)$ for $j > t$, $\mathbf{I}$ for $j = t$, and $\prod_{l=0}^{j+1} \mathbf{G}^{+}(t-l)$ for $j < t$, with $(\cdot)^{+}$ being the matrix pseudoinverse. This matrix represents the geometric relationship between the different HR images.

The first term of the cost function in (2.4) measures how close the estimate of the degraded image is to the actual low resolution image, given the degradation model, evaluating the *data fidelity*. The second term, on the other hand, measures how close the estimate is from *a priori* models employed for the HR image sequence (assuming a Bayesian interpretation for the estimation problem).

While the evaluation of the solution fidelity to the observed data in (2.4) is performed in the same way as is done in the super-resolution methods aimed at the reconstruction of a single frame, the regularization employed is significantly different. This is because the super-resolution of video sequences also considers the temporal relationship between different images in the sequence, besides the spatial relationship between different pixels in each image. The regularization term is of the from:

$$\mathfrak{R}\big(\hat{\mathbf{x}}(1), \cdots, \hat{\mathbf{x}}(T)\big) = \underbrace{\sum_{t=1}^{T} \|\mathbf{S}\hat{\mathbf{x}}(t)\|_p^q}_{\text{Spatial}} + \gamma_T \underbrace{\sum_{t=2}^{T} \|\hat{\mathbf{x}}(t) - \mathbf{G}(t)\hat{\mathbf{x}}(t-1)\|_l^m}_{\text{Temporal}}, \quad (2.5)$$

with $\mathbf{S}$ being a high pass filter (such as a Laplacian (GONZALEZ; WOODS, 2002, p. 182)). This first term emphasizes the spatial smoothness of the images, while the second ensures that the different images are temporally consistent (i.e. correlated). Besides the $p-$norms shown above, it is common to employ other norms with different characteristics, such as the Huber norm, for example, which is a hybrid L_1/L_2 function that besides being robust to outliers is differentiable every-

where.

The use of the temporal regularization allows the attainment of a higher quality on the reconstructed sequence with the inclusion of additional *a priori* information. This can be particularly noticed in the work of Su et al. (SU; WU; ZHOU, 2011), where employing the temporal regularization in a frame-by-frame estimation framework with a sliding observation window resulted in an increased quality if compared with the spatial regularization alone. Furthermore, the temporal regularization turns essential to define the relationship between different frames in works that considers a *diagonal* model for the data term in the cost function (i.e. not considering the inter-frame relationship). This strategy was employed in (ZIBETTI; MAYER, 2007) in order to reduce the computational cost and increase the robustness.

The temporal regularization originated with the restoration of multichannel images (i.e. images generated by multispectral sensors) through the description of the cross-correlation matrices between different image channels (GALATSANOS; CHIN, 1987). This approach was later extended to the restoration (i.e. denoising and deblurring) of video sequences by treating each image in the sequence as an individual image channel. Ozkan et al. (OZKAN et al., 1992) were the first to address video restoration problem. The resulting image sequence was obtained as the Wiener solution and the temporal correlation was present in the computation of the cross-correlation matrices between the different images, and a closed form solution was developed for the case of translational motion (avoiding the inversion of a large matrix or the estimation of the cross-correlation functions). Choi et al. (CHOI; GALATSANOS; KATSAGGELOS, 1996) later proposed a simultaneous video restoration method following a regularized least squares approach, where the temporal smoothness was enforced trough a motion compensated Laplacian operator, which accounts for the displacement in each pixel when calculating the derivative in the time axis through finite differences. The resulting optimization problem was solved iteratively using the gradient method.

Borman et al. (BORMAN; STEVENSON, 1999) were the first to address the simultaneous super-resolution of a video sequence using a Bayesian approach, where the temporal correlation between frames was again used through a motion compensated Laplacian operator, employing the Huber norm on the regularization terms to better preserve the image edges during reconstruction. The authors reported that the smooth motion trajectories enforced trough the use of the temporal prior enabled the achievement of an increased video quality. In the

work of Tian and Ma (2005, 2009) a state-space approach to super resolve video sequences was proposed. Along with the two equations describing the HR image's dynamics and observation model used on the traditional Kalman filter, a third equation describing the temporal relationship between adjacent low resolution images was also employed, providing additional information on the temporal correlation of the video sequence. Later, with a methodology based on that of Borman and Stevenson (1999), Zibetti et al. (ZIBETTI; MAYER, 2007) proposed a diagonal observation model to achieve an increased robustness, which described only the motion between the different frames through the temporal regularization. Belekos et al. (BELEKOS; GALATSANOS; KATSAGGELOS, 2010) later proposed a two-level hierarchical Bayesian methodology for the simultaneous super resolution of video sequences, using a temporal *a priori* constraint along with a spatial regularization, both employing spatially varying statistics. Richter et al. (RICHTER et al., 2011) proposed a temporal prior for video restoration where the regularization factor on each pixel was weighted inversely proportional to the temporal derivative of the motion compensated pixels in a small neighborhood, reducing the filtering strength on high temporal differences and increasing robustness. Su et al. (SU; WU; ZHOU, 2011) employed the temporal *a priori* term between adjacent images for the adaptive super resolution of video sequences. Although the single image quality did not change significantly, an improvement could be verified through a video quality assessment criteria.

2.2.2 Sequential Methods

Differently from the simultaneous SRR algorithms, sequential SRR methods perform the estimation of a single frame $\mathbf{x}(t)$ at a time, using the previous reconstruction results in the computation of the present estimate. Despite the sequences reconstructed by the sequential methods generally having inferior quality when compared with those reconstructed by the simultaneous methods, their computational cost is significantly lower, which makes them suited for both real-time processing or the reconstruction of sequences containing a large number of frames.

The sequential algorithms are based on the dynamic signal model (2.3) and (2.1), which represents a state-space model. Therefore, sequential techniques can formulate the reconstruction problem as a state estimation problem, frequently considering for this the use of methods

based on variations of the Kalman filter, leading to this class of methods to be prevalent in *adaptative* SRR algorithms.

The existing approaches are generally based on the algorithm initially proposed by Elad and Feuer (1999b, 1999a), where a Kalman filter was applied to estimate the HR image $\mathbf{x}(t)$. The same work also presents two simpler algorithms, the Steepest Descent (SD) and the Least Mean Squares (LMS), which are obtained through approximations/simplifications of the traditional Kalman filter. Later, Wang et al. (WANG; QI, 2002) presented a modified version of the Kalman filter devised to increase the robustness to registration errors. In Farsiu et al. (2004a, 2006) a simplification was proposed for the case of global translational motion, addressing the deblurring step separately, which resulted in a shift-and-add type of algorithm which was also applied to color images. This methodology was extended by Kim et al. (2010) by considering the detection and removal of previously estimated pixels that show a significant deviation when compared to the present time observation, besides the reinitialization of the algorithm in the case of high innovations with high magnitude.

Since the estimated image content is "carried" gradually through time, the image sequences reconstructed with sequential algorithms are usually temporally consistent by nature, leading the temporal regularization to be seldom employed for these methods. An exception is the work of Tian and Ma (2005, 2009), which devises an algorithm based on the Kalman filter using an additional equation to describe the temporal relationship between the LR images, introducing additional *a priori* information about the problem.

Besides having a very low computational cost, the derivation of most sequential SRR algorithms is very consistent from a theoretical viewpoint, which allows for a better understanding of their behavior (COSTA; BERMUDEZ, 2007, 2008). Furthermore, additional *a priori* information about the problem can be more easily included.

Among the sequential methods, the R-LMS algorithm stands out due to its simplicity, performance and the possibility of mathematical analysis (COSTA; BERMUDEZ, 2007). Since this algorithm will be employed in this work, it will be presented in detail in the following section.

2.3 THE R-LMS SRR ALGORITHM

Observing the reconstruction problem for the sequential algorithms, it can be noticed that the solutions are generally based on the minimization of the estimation error for a given time instant (see (PARK; PARK; KANG, 2003) and references therein)

$$\boldsymbol{\epsilon}(t) = \mathbf{y}(t) - \mathbf{D}\mathbf{H}\hat{\mathbf{x}}(t)\,, \tag{2.6}$$

where $\boldsymbol{\epsilon}(t)$ can be seen as an estimate of the observation noise in (2.1). Different methodologies can be used to perform the reconstruction, minimizing different functions of $\boldsymbol{\epsilon}(t)$, employing different models for the signals (e.g. deterministic or stochastic), as well as additional restrictions and *a priori* information, which leads to many different possibilities of algorithms with different compromises between performance and computational cost.

Considering both computational and memory costs, few among the sequential algorithms can be compared with the R-LMS, what makes it suited for real time processing (COSTA; BERMUDEZ, 2008). Initially proposed by Elad and Feuer (1999b, 1999a), the LMS algorithm performs the minimization of the mean squared value of the L_2 norm of (2.6), conditioned on the estimate $\hat{\mathbf{x}}(t)$ (ELAD; FEUER, 1999b; COSTA; BERMUDEZ, 2007), which can be translated in the cost function $\mathbf{J}_{\mathrm{MS}}(t) = \mathrm{E}\{\|\boldsymbol{\epsilon}(t)\|^2 \,|\, \hat{\mathbf{x}}(t)\}$, where $\mathrm{E}\{\cdot\}$ denotes the expectation of a random variable.

Since natural images are intrinsically smooth, the result of the estimation can be improved by incorporating this knowledge to the optimization problem in the form of a regularization for the LMS algorithm, constraining the solution that minimizes $\mathbf{J}_{\mathrm{MS}}(t)$. This way, the R-LMS algorithm (ELAD; FEUER, 1999b, 1999a) can be derived as the solution of the following constrained optimization problem:

$$\begin{aligned} \min_{\hat{\mathbf{x}}(t)} \quad & \mathbf{J}_{\mathrm{MS}}(t) = \mathrm{E}\{\|\boldsymbol{\epsilon}(t)\|^2 \,|\, \hat{\mathbf{x}}(t)\} \\ \text{subject to} \quad & \|\mathbf{S}\hat{\mathbf{x}}(t)\|^2 = 0 \end{aligned} \tag{2.7}$$

where $\mathbf{S}$ is a spatial high-pass filtering operator (e.g. a Laplacian (GONZALEZ; WOODS, 2002, p. 182)). It is important to notice that the performance surface in (2.7) is defined for a single time instant t, and the statistical expectation is computed on the ensemble, instead of on the time axis.

The Lagrangian of (2.7) is given by

$$\mathcal{L}_{\text{R-MS}}(t) = \mathrm{E}\{\|\mathbf{y}(t) - \mathbf{DH}\hat{\mathbf{x}}(t)\|^2 \,|\, \hat{\mathbf{x}}(t)\} + \alpha\|\mathbf{S}\hat{\mathbf{x}}(t)\|^2 \tag{2.8}$$

Following the gradient descent approach, the estimate of the HR image $\hat{\mathbf{x}}(t)$ should be updated in the negative direction of the gradient, which for the cost function defined in (2.8) is given by:

$$\nabla\mathcal{L}_{\text{R-MS}}(t) = \frac{\partial\mathcal{L}_{\text{R-MS}}(t)}{\partial\hat{\mathbf{x}}(t)} = -2\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\{\mathrm{E}[\mathbf{y}(t)] - \mathbf{DH}\hat{\mathbf{x}}(t)\} + 2\alpha\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}(t) \tag{2.9}$$

The recursive update equation for $\hat{\mathbf{x}}(t)$ following the gradient descent approach is therefore given by

$$\hat{\mathbf{x}}_{k+1}(t) = \hat{\mathbf{x}}_k(t) - \frac{\mu}{2}\nabla\mathcal{L}_{\text{R-MS}}(t)\,, \tag{2.10}$$

where the variable k indexes the iterative update of $\hat{\mathbf{x}}(t)$ for a single time instant t.

The R-LMS algorithm is the stochastic version of the gradient descent approach, where the gradient of the cost function in (2.9) is approximated by its instantaneous estimate (HAYKIN, 1991).

$$\begin{aligned}\nabla\mathcal{L}_{\text{R-MS}}(t) &= -2\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\{\mathrm{E}[\mathbf{y}(t)] - \mathbf{DH}\hat{\mathbf{x}}(t)\} + 2\alpha\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}(t) \\ &\simeq -2\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\{\mathbf{y}(t) - \mathbf{DH}\hat{\mathbf{x}}(t)\} + 2\alpha\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}(t)\end{aligned} \tag{2.11}$$

Using approximation (2.11) in (2.10), the update equation for the R-LMS algorithm for a single time instant t is obtained as

$$\begin{aligned}\hat{\mathbf{x}}_{k+1}(t) = \hat{\mathbf{x}}_k(t) &+ \mu\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}[\mathbf{y}(t) - \mathbf{DH}\hat{\mathbf{x}}_k(t)] \\ &- \alpha\mu\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_k(t)\,, k = 0, 1, \ldots, K-1\end{aligned} \tag{2.12}$$

The temporal update of (2.12) is based on the signal dynamics (2.3), being given by (ELAD; FEUER, 1999b)

$$\hat{\mathbf{x}}_0(t+1) = \mathbf{G}(t+1)\hat{\mathbf{x}}_K(t)\,. \tag{2.13}$$

For each time sample, (2.12) is iterated for $k = 0, \ldots, K-1$.

(a) (b) (c) (d)

Figure 8 – Reconstruction results for a video sequence containing an outlier (taken from (ZIBETTI; MAYER, 2007)). (a) Original image (HR). (b) Bicubic interpolation. (c) Single frame based SRR. (d) L_2 norm based SRR with the Borman and Stevenson (BORMAN; STEVENSON, 1999) method.

2.4 ROBUSTNESS AND REAL TIME OPERATION

Since the super resolution reconstruction problem is an inverse problem, it is well known that the quality of the solution is highly dependent on how precise are the signal models employed in (2.1) and (2.3). The presence of outliers such as registration errors and large innovations on the reconstruction process can lead to an estimated image quality that is inferior to that of the observed LR images themselves, mainly due to the presence of artifacts and ghost images. This fact makes super-resolving real video sequences unlikely to provide good quality results for most of the existing SRR algorithms (ZIBETTI; MAYER, 2007). This makes robustness a fundamental characteristic for a satisfactory performance of SRR algorithms in real world applications.

An example of this effect can be observed in Figure 8 taken from (ZIBETTI; MAYER, 2007), where images from a scene containing independent motion from a circular object are reconstructed. An algorithm that does not use information from the temporal dynamics from (2.3) is compared with that proposed by Borman and Stevenson (BORMAN; STEVENSON, 1999), where it is possible to notice that the presence of occlusions lead to significant artifacts which are visible at the borders of the black circle.

Robust super-resolution basically aims to reduce the influence of outliers in the reconstruction process, which is performed by introducing non-linear techniques in the reconstruction algorithms.

The existing methods are generally divided between those that employ a pre-processing step, where weights are attributed to pixels depending on the probability of them being outliers (eliminating them

from the reconstruction process, in a more extreme case), and those that use cost functions that provide enhanced robustness (MILANFAR, 2010).

In the cases where a pre-processing step is employed, the assignment of a reduced or zero weight for pixels identified as outliers is performed before the reconstruction, which then uses a common L_2 norm (MILANFAR, 2010). Examples of the weight computation can be seen in (ZHAO; SAWHNEY, 2002), which used the cross-correlation between the registered images, or in (LEE; KANG, 2003), where it is based on the residual estimation error $\|\mathbf{y}(t) - \mathbf{DH}\,\mathbf{M}_{t,j}\,\hat{\mathbf{x}}(j)\|$.

The second and more common approach consists on the use of a robust cost function, which can sometimes be translated into a more appropriate characterization of the signals involved. The modeling of the estimation error $\|\mathbf{y}(t) - \mathbf{DH}\,\mathbf{M}_{t,j}\,\hat{\mathbf{x}}(j)\|$ as a Gaussian noise leads to reconstruction methods based on the L_2 norm, which is sensitive to the presence of outliers. The use of a Laplacian distribution, on the other hand, leads to a more accurate model for the reconstruction error, resulting in algorithms based on the L_1 norm such as that of Farsiu et al. (FARSIU et al., 2004b). In the case of video SRR, this concept is frequently extended to the norm of the temporal regularization (BORMAN; STEVENSON, 1999; ZIBETTI; MAYER, 2007), while other works also consider an adaptive weighting of this term (SU; WU; ZHOU, 2011; RICHTER et al., 2011), as well as the employment of the motion information only in the temporal *a priori* term (ZIBETTI; MAYER, 2007), resulting in a data model that does not includes the inter-frame relationship. Robust methodologies based on adaptive algorithms were also proposed for robustness to registration errors, such as modified versions of the Kalman filter devised in (WANG; QI, 2002), or for robustness to innovations as in (KIM et al., 2010).

It is important to notice that, although these techniques achieve good reconstruction results, the use of nonlinear penalty functions significantly increases their associated computational cost. When comparing these methods with those devised for real-time operation, the computational cost associated with even the simpler robust algorithms as that of (FARSIU et al., 2004b) is not comparable with that of the R-LMS for example (COSTA; BERMUDEZ, 2008), which makes them unsuited for applications requiring real-time operation. This depicts a conflicting relationship between computational cost, quality and robustness, which motivates the development of new techniques that offer a more desirable balance between these three characteristics.

3 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 (2.13) 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. This fact makes achieving super-resolution of real video sequences highly unlikely, not just for the R-LMS but for most existing algorithms (ZIBETTI; MAYER, 2007). On the other hand, super-resolution algorithms devised to be robust under the influence of innovations exhibit an increased computational cost, which turns them unsuited for real time applications (FARSIU et al., 2004b; COSTA; BERMUDEZ, 2008).

An interesting interpretation of the R-LMS algorithm is possible if we view each iteration of the gradient algorithm (2.10) (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 (BECK; TEBOULLE, 2009, Section 2.2) or (BERTSEKAS,

1999, p. 546), the gradient iteration (2.10) can be written as

$$\begin{aligned}\hat{\mathbf{x}}_{k+1}(t) =& \arg\min_{\mathbf{z}} \Big\{ \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) + \big(\mathbf{z} - \hat{\mathbf{x}}_k(t)\big)^{\mathrm{T}} \nabla \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) \\ &+ \frac{1}{\mu} \left\| \mathbf{z} - \hat{\mathbf{x}}_k(t) \right\|^2 \Big\} \quad (3.1) \\ =& \arg\min_{\mathbf{z}} \Big\{ \mathbf{z}^{\mathrm{T}} \Big(2\alpha \mathbf{S}^{\mathrm{T}} \mathbf{S} \hat{\mathbf{x}}_k(t) - 2\mathbf{H}^{\mathrm{T}} \mathbf{D}^{\mathrm{T}} \big\{ \mathrm{E}[\mathbf{y}(t)] - \mathbf{D}\mathbf{H}\hat{\mathbf{x}}_k(t) \big\} \Big) \\ &+ \frac{1}{\mu} \left\| \mathbf{z} - \hat{\mathbf{x}}_k(t) \right\|^2 \Big\} \\ =& \arg\min_{\mathbf{z}} \Big\{ 2\alpha \mathbf{z}^{\mathrm{T}} \mathbf{S}^{\mathrm{T}} \mathbf{S} \hat{\mathbf{x}}_k(t) - 2\mathbf{z}^{\mathrm{T}} \mathbf{H}^{\mathrm{T}} \mathbf{D}^{\mathrm{T}} \mathrm{E}[\boldsymbol{\epsilon}_k(t)] \\ &+ \frac{1}{\mu} \left\| \mathbf{z} - \hat{\mathbf{x}}_k(t) \right\|^2 \Big\}, \quad (3.2)\end{aligned}$$

where $\mathrm{E}[\boldsymbol{\epsilon}_k(t)]$ is the expected value of the observation error (2.6) conditioned on $\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}_k(t)$. This equivalence is verified as follows. Differentiating the expression within the external brackets in (3.1) with respect to $\mathbf{z}$ and setting it to zero yields

$$\nabla \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) + \frac{2}{\mu}(\mathbf{z} - \hat{\mathbf{x}}_k(t)) = \mathbf{0} \quad (3.3)$$

which solved for $\mathbf{z} = \hat{\mathbf{x}}_{k+1}(t)$ leads to

$$\hat{\mathbf{x}}_{k+1}(t) = \hat{\mathbf{x}}_k(t) - \frac{\mu}{2} \nabla \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) \quad (3.4)$$

which is the R-LMS gradient descent update equation (2.10).

Now, the presence of the squared norm within the external brackets in (3.1) 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.

In order to illustrate this behavior, consider for instance that the reconstructed image sequence at time instant $t-1$ is reasonably close to the real (desired) sequence, i.e. $\hat{\mathbf{x}}(t-1) \simeq \mathbf{x}(t-1)$. If we consider the video sequence to be only slightly perturbed in the next time instant such that $\|\mathbf{s}(t)\| \approx 0$, the first iteration ($k=0$ at time t) of the stochastic version of algorithm (3.2) can be written as

$$\begin{aligned}
\hat{\mathbf{x}}_1(t) &= \arg\min_{\mathbf{z}} \left\{ 2\alpha \mathbf{z}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_0(t) - 2\mathbf{z}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\boldsymbol{\epsilon}_0(t) + \frac{1}{\mu}\left\|\mathbf{z}-\hat{\mathbf{x}}_0(t)\right\|^2 \right\} \\
&\approx \arg\min_{\mathbf{z}} \left\{ 2\alpha \mathbf{z}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_0(t) - 2\mathbf{z}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\boldsymbol{\epsilon}_0(t) \right. \\
&\qquad\qquad \left. + \frac{1}{\mu}\left\|\mathbf{z}-\mathbf{G}(t)\mathbf{x}(t-1)\right\|^2 \right\} \\
&= \arg\min_{\mathbf{z}} \left\{ 2\alpha \mathbf{z}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_0(t) - 2\mathbf{z}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\boldsymbol{\epsilon}_0(t) \right. \\
&\qquad\qquad \left. + \frac{1}{\mu}\left\|\mathbf{z}-\mathbf{x}(t)+\mathbf{s}(t)\right\|^2 \right\} \\
&\approx \arg\min_{\mathbf{z}} \left\{ 2\alpha \mathbf{z}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_0(t) - 2\mathbf{z}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\boldsymbol{\epsilon}_0(t) + \frac{1}{\mu}\left\|\mathbf{z}-\mathbf{x}(t)\right\|^2 \right\}
\end{aligned} \tag{3.5}$$

where we have used (2.13) in the second line, (2.3) in the third line, and the hypothesis of $\|\mathbf{s}(t)\| \approx 0$ in the last line. Now, the norm of the observation error $\boldsymbol{\epsilon}_0(t)$ in the first iteration is given by

$$\begin{aligned}
\|\boldsymbol{\epsilon}_0(t)\| &\simeq \|\mathbf{DHx}(t) + \mathbf{e}(t) - \mathbf{DH}\big(\mathbf{x}(t)-\mathbf{s}(t)\big)\| \\
&\approx \|\mathbf{DHx}(t) + \mathbf{e}(t) - \mathbf{DHx}(t)\| \\
&= \|\mathbf{e}(t)\| .
\end{aligned}$$

Given the assumption $\hat{\mathbf{x}}(t-1) \simeq \mathbf{x}(t-1)$, this error will be small. Then, for reasonably small values of α and μ, the second term in (3.5) can be neglected (i.e. $|\mathbf{z}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\boldsymbol{\epsilon}_0(t)| \ll |2\alpha\mathbf{z}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_0(t) + \frac{1}{\mu}\left\|\mathbf{z}-\mathbf{x}(t)\right\|^2|$) and the solution $\hat{\mathbf{x}}_1(t)$ will only be slightly perturbed from the initialization $\hat{\mathbf{x}}_0(t)$ due to the first term in (3.5). Hence, the dominance of the term

$\frac{1}{\mu}\left\|\mathbf{z}-\mathbf{x}(t)\right\|^2$ will lead to a solution $\hat{\mathbf{x}}_1(t) \approx \mathbf{x}(t)$. The same reasoning can be extended to the remaining iterations for $k = 2, \ldots, K-1$, which shows that, for $\mathbf{s}(t) \approx 0$, the algorithm will lead to a reconstructed image of good quality $\hat{\mathbf{x}}_K(t) \simeq \mathbf{x}(t)$. This explains how the R-LMS algorithm preserves the reconstructed content in time and extracts information from the different observations, attaining good reconstruction results for well behaved sequences, i.e. in the absence of large innovations.

Now, let's consider the presence of a significant innovation outlier at time t, while still assuming a good reconstruction result on time $t-1$ (i.e. $\hat{\mathbf{x}}(t-1) \simeq \mathbf{x}(t-1)$). In the occurrence of an outlier at time instant t, $\mathbf{s}(t)$ in (2.3) will have a significant energy. Then, repeating (3.5) without the assumption $\|\mathbf{s}(t)\| \approx 0$, we have for the first iteration that

$$\begin{aligned}
\hat{\mathbf{x}}_1(t) &= \arg\min_{\mathbf{z}} \left\{ 2\alpha \mathbf{z}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_0(t) - 2\mathbf{z}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\boldsymbol{\epsilon}_0(t) + \frac{1}{\mu}\left\|\mathbf{z}-\hat{\mathbf{x}}_0(t)\right\|^2 \right\} \\
&\approx \arg\min_{\mathbf{z}} \left\{ 2\alpha \mathbf{z}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_0(t) - 2\mathbf{z}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\boldsymbol{\epsilon}_0(t) \right. \\
&\qquad \left. + \frac{1}{\mu}\left\|\mathbf{z}-\mathbf{G}(t)\mathbf{x}(t-1)\right\|^2 \right\} \\
&= \arg\min_{\mathbf{z}} \left\{ 2\alpha \mathbf{z}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_0(t) - 2\mathbf{z}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\boldsymbol{\epsilon}_0(t) \right. \\
&\qquad \left. + \frac{1}{\mu}\left\|\mathbf{z}-\big(\mathbf{x}(t)-\mathbf{s}(t)\big)\right\|^2 \right\}
\end{aligned} \tag{3.6}$$

where the observation error is given by

$$\begin{aligned}
\|\boldsymbol{\epsilon}_0(t)\| &\simeq \|\mathbf{DHx}(t) + \mathbf{e}(t) - \mathbf{DH}\big(\mathbf{x}(t)-\mathbf{s}(t)\big)\| \\
&= \|\mathbf{e}(t) + \mathbf{DHs}(t)\| .
\end{aligned} \tag{3.7}$$

For a fast convergence of the algorithm for a fixed value of t and $k = 1, \ldots, K$, meaning that one could choose K small, the cost function should allow for a considerable change of the estimate towards $\mathbf{x}(t)$ when going from $\hat{\mathbf{x}}_l(t)$ to $\hat{\mathbf{x}}_{l+1}(t)$, $l = 1, \ldots, K-1$. We discuss the case $l = 0$ (first iteration) and then extend the conclusions to other values of l. Now, for large values of $\mathbf{s}(t)$, $\|\boldsymbol{\epsilon}_0(t)\|$ in (3.7) will be large and dominated by the term $\mathbf{DHs}(t)$. Moreover, for values of α and μ typically chosen for an outlier-free situation, the term $\frac{1}{\mu}\|\mathbf{z} - \big(\mathbf{x}(t) - \mathbf{s}(t)\big)\|^2$ will be very large if $\mathbf{z} \simeq \mathbf{x}(t)$. Hence, the estimation $\hat{\mathbf{x}}_1(t)$, while being still close to the initialization (which does not contain the outlier

$\mathbf{s}(t)$), will be far from the desired solution $\mathbf{x}(t)$ (which should contain $\mathbf{s}(t)$). This explains the poor transient performance of the algorithm in the presence of outliers.

Performance improvement in the presence of outliers could be sought by increasing the value of μ to reduce the influence of the term $\frac{1}{\mu}\|\mathbf{z} - \big(\mathbf{x}(t) - \mathbf{s}(t)\big)\|^2$ in (3.6). However, μ cannot be made arbitrarily large for stability reasons. Hence, improvement would have to come from increasing the importance of the first term in (3.6). Expression (3.6) can be written as

$$\begin{aligned}
\hat{\mathbf{x}}_1(t) \approx \quad & \arg\min_{\mathbf{z}} \left\{ 2\alpha \mathbf{z}^\top \mathbf{S}^\top \mathbf{S} \hat{\mathbf{x}}_0(t) - 2\mathbf{z}^\top \mathbf{H}^\top \mathbf{D}^\top \boldsymbol{\epsilon}_0(t) + \frac{1}{\mu} \|\mathbf{z} - \hat{\mathbf{x}}_0(t)\|^2 \right\} \\
= 2 \ & \arg\min_{\mathbf{z}} \left\{ \alpha \left(\mathbf{S}\mathbf{z}\right)^\top \left(\mathbf{S}\hat{\mathbf{x}}_0(t)\right) - \left(\mathbf{D}\mathbf{H}\mathbf{z}\right)^\top \boldsymbol{\epsilon}_0(t) \right. \\
& \qquad \left. + \frac{1}{2\mu} \|\mathbf{z} - \hat{\mathbf{x}}_0(t)\|^2 \right\} \\
\approx 2 \ & \arg\min_{\mathbf{z}} \left\{ \alpha \left(\mathbf{S}\mathbf{z}\right)^\top \left(\mathbf{S}\hat{\mathbf{x}}_0(t)\right) - \left(\mathbf{D}\mathbf{H}\mathbf{z}\right)^\top \left(\mathbf{D}\mathbf{H}\mathbf{s}(t)\right) \right. \\
& \qquad \left. + \frac{1}{2\mu} \|\mathbf{z} - \hat{\mathbf{x}}_0(t)\|^2 \right\}
\end{aligned}$$

First, note that $\hat{\mathbf{x}}_0(t) = \mathbf{G}(t)\hat{\mathbf{x}}_K(t-1)$ does not include information on the outlier $\mathbf{s}(t)$, as it has been introduced in $\mathbf{x}(t)$. Hence, any compensation for the effect of a large $\mathbf{s}(t)$ in the last term would have to come from the second term. Now, the first term in (3.6) has an effect that opposes that of the last term by favoring a solution that is orthogonal and smaller to that of the previous iteration. Therefore, by increasing the contribution of this term, it is possible to allow larger changes in the solution, leading to a faster convergence in k. However, since $\mathbf{S}$ is a high pass filter and the second term affects only the projection of $\mathbf{z}$ in the row space of $\mathbf{DH}$, this also leads the solution to lose the estimated details. Thus, increasing the value of α in an attempt to speed up convergence in the presence of large innovations by reducing the influence of the last term in (3.6) will also reduce the temporal correlation of the estimated image sequence, resulting in an overly blurred solution with lower quality in the absence of outliers. The same reasoning can clearly be extended to the remaining iterations for $l = 1, \ldots, K-1$.

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 $\frac{1}{\mu}\|\mathbf{z} - \big(\mathbf{x}(t) - \mathbf{s}(t)\big)\|^2$. 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 defying the purpose of the algorithm.

3.1 ILLUSTRATIVE EXAMPLE

The behavior of the R-LMS algorithm is illustrated in the following example, where we consider the reconstruction of a synthetic video sequence generated through small translational displacements of a 32×32 window over a larger natural image. At a specific time instant during the video (the 32nd frame), an outlier is introduced by adding a black square of size 16×16 to the scene. The square remain in the scene for 3 frames, before disappearing again. The HR sequence is convolved with a 3×3 uniform blurring mask, down-sampled by a factor of 2. Finally, a white Gaussian noise with variance 10 is added to generate the low resolution video.

The R-LMS algorithm is applied to super-resolve the synthetic LR videos generated, and the mean square error (MSE) is measured between the original and reconstructed sequences by averaging the results from 50 realizations. To illustrate the weighting effect between the step size and regularization parameter in the cost function, we reconstruct the sequence with both $\alpha = 2 \times 10^{-4}$ and $\alpha = 100 \times 10^{-4}$, for $\mu = 4$. For the evaluation of the effect of the number of iterations per time interval, we run the algorithm with $K = 2$ and $K = 100$. The MSE is depicted in Figure 9. From the two curves for $K = 2$ one can verify that a large value for α (red curve) reduces the MSE in the presence of the outlier, while the greater temporal correlation induced by a small value of α (black curve) tends to reduce the error for small innovations and to increase it in the presence of an outlier. Comparing the blue ($K = 100$) and the black ($K = 2$) curves, both for $\alpha = 2 \times 10^{-4}$ and $\mu = 4$, one verifies that the MSE can be substantially decreased by employing the R-LMS algorithm with a large K. The MSE is smaller than that obtained for $K = 2$ both for small and for large innovations. This performance improvement is because the algorithm is allowed to converge slowly for each time interval. Figure 10 shows the MSE as a function of k for time instant $t = 32$, when the outlier is present. These

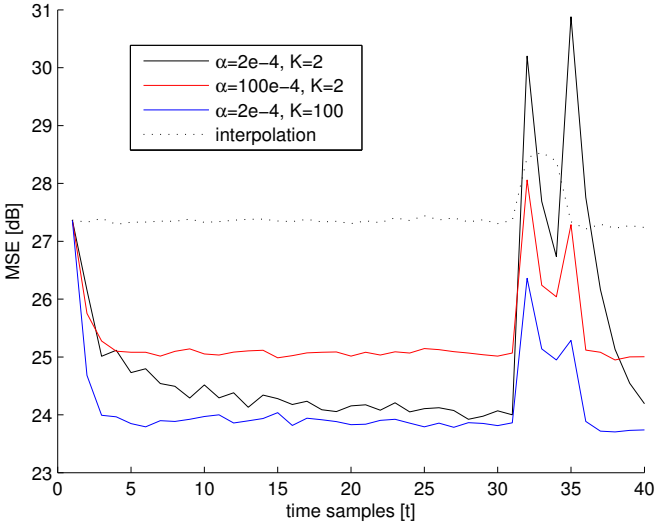


Figure 9 – MSE results for the R-LMS algorithm with different values of α and K.

results illustrate the property that a large value of K is necessary to achieve a significant MSE reduction for a fixed value of t.

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.

3.2 RELATED APPROACHES FOR VIDEO SRR

In the context of adaptive algorithms such as the R-LMS, 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 (CHOI; GALATSANOS; KATSAGGELOS, 1996; BORMAN; STEVENSON, 1999; ZIBETTI; MAYER, 2007), can be interpreted as the application of the well known least perturbation or minimum disturbance principle. This principle states that *the parameters of an adaptive system should only be disturbed in a minimal fashion in the light of new data* (HAYKIN, 1991, p. 355). Using this principle, the one-dimensional LMS algorithm can be shown to correspond not to an approximate solution of a gradient-based optimiza-

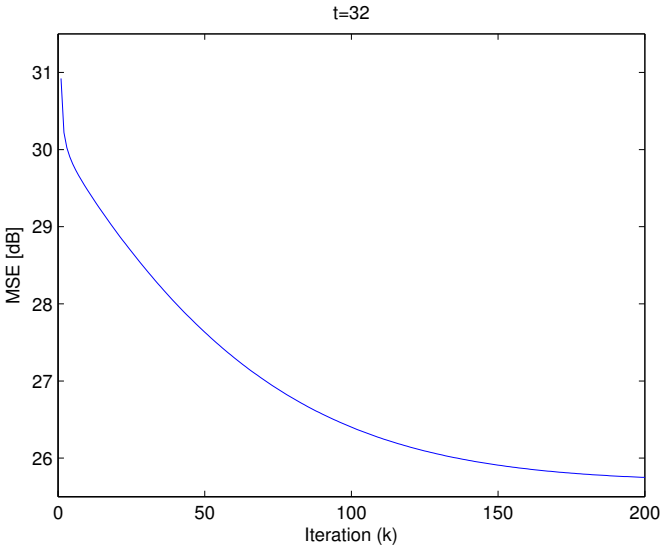


Figure 10 – MSE evolution per iteration during a single time instant $t = 32$.

tion problem, but to the exact solution of a constrained optimization problem (SAYED, 2003, p. 216).

Differently from simultaneous video SRR methods, the R-LMS algorithm's cost function (2.7) is defined for a single time instant. Thus, the proximal regularization described in Section 3 only guarantees consistence between consecutive iterations in k. Therefore, since the previous time instant's solution $\hat{\mathbf{x}}(t-1)$ is only introduced during the initialization in (2.13), consistence between consecutive time instants is only achieved if the solution is not disturbed during all iterations $k = 1, \ldots, K$ (i.e. $\hat{\mathbf{x}}_K(t) \simeq \hat{\mathbf{x}}_0(t)$).

To alleviate this limitation, one might be tempted to modify the optimization problem (3.2) by introducing an additional temporal regularization term as follows:

$$\hat{\mathbf{x}}_{k+1}(t) = \arg\min_{\mathbf{z}} \Big\{ \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) + \big(\mathbf{z} - \hat{\mathbf{x}}_k(t)\big)^{\mathrm{T}} \nabla \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) + \frac{1}{\mu} \left\| \mathbf{z} - \hat{\mathbf{x}}_k(t) \right\|^2 + \frac{1}{\alpha_{\mathrm{T}} \mu} \left\| \mathbf{z} - \mathbf{G}(t) \hat{\mathbf{x}}(t-1) \right\|^2 \Big\} \tag{3.8}$$

where α_{T} is a weighting factor controlling the temporal disturbance. Albeit removing the dependence of its solution on the time initialization (2.13), the algorithm in (3.8) fails to achieve good results. Instead, this new regularization term 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). This is clearly perceived by assuming again that $\|\mathbf{s}(t)\|$ is large and $\hat{\mathbf{x}}(t-1) \simeq \mathbf{x}(t-1)$, and examining the norm of the last term in (3.8) for $\mathbf{z} = \hat{\mathbf{x}}_{k+1}(t)$

$$\|\hat{\mathbf{x}}_{k+1}(t) - \mathbf{G}(t)\hat{\mathbf{x}}(t-1)\| \approx \|\hat{\mathbf{x}}_{k+1}(t) - \big(\mathbf{x}(t) - \mathbf{s}(t)\big)\|$$

which will be large if $\hat{\mathbf{x}}_{k+1}(t) \simeq \mathbf{x}(t)$ not only for $k = 1$, but for all iterations. Furthermore, when the innovations are small, this term shows to be unnecessary since in this case the R-LMS can retain the temporal consistency even for a large number of iterations (K) as demonstrated in the example of section 3.1 with $K = 100$.

Since the temporal regularization introduced in (3.8) is not effective in increasing algorithm robustness or quality, these issues must be addressed using other approaches. Most works in the literature regarding both single-frame and video SRR seek robustness by considering cost functions which result in non-linear algorithms, using techniques such as non-quadratic (e.g. L_1) error norms (FARSIU et al., 2004b; BORMAN; STEVENSON, 1999; ZIBETTI; MAYER, 2007), or signal dependent regularization weighting (SU; WU; ZHOU, 2011; RICHTER et al., 2011). 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, since a good result is achieved for large K. A different approach is therefore required to adequately handle the innovations in the R-LMS algorithm.

In the following, we propose to use meaningful *a priori* information about the statistical nature of the innovations in deriving a new stochastic SRR method using the least perturbation principle. The proposed approach can improve the robustness of the R-LMS algorithm while retaining a reduced computational cost. By employing statistical information about $\mathbf{s}(t)$, which has been overlooked in the design of simple SRR algorithms, it becomes possible to provide robustness to the innovations while maintaining a good reconstruction quality.

4 CONSTRUCTING AN INNOVATION-ROBUST REGULARIZATION

In order to achieve the desired effects, we propose to modify the norm being minimized in the last term of (3.8) through the inclusion of a weighting matrix $\mathbf{Q}$ properly designed to emphasize the image details in the regularization term. This will allow the resulting algorithm to attain a faster speed of convergence with a good quality, while at the same time decreasing the influence of the innovations from the optimization process.

The new constraint is then given by

$$\|\mathbf{Q}\big(\hat{\mathbf{x}}_{k+1}(t) - \mathbf{G}(t)\hat{\mathbf{x}}(t-1)\big)\| \tag{4.1}$$

and $\mathbf{Q}$ must be designed to preserve the details of the estimated images, and so that the presence of innovations will have a minimal effect upon the regularization term. Thus, it is desired that

$$\begin{aligned} \mathbf{Q}\ \mathbf{x}(t) &\sim \text{details} \\ \mathbf{Q}\ \mathbf{s}(t) &\sim \mathbf{0} \end{aligned} \tag{4.2}$$

which means that the image details must lie in the column space of $\mathbf{Q}$, while the innovations lie in its nullspace. Therefore, if we assume the reconstructed image in time instant $t-1$ to be reasonably close to the real (desired) image, (i.e. $\hat{\mathbf{x}}_K(t-1) \simeq \mathbf{x}(t-1)$), we can write the modified restriction as

$$\|\mathbf{Q}\hat{\mathbf{x}}_{k+1}(t) - \mathbf{Q}\mathbf{x}(t) + \mathbf{Q}\mathbf{s}(t)\|\,. \tag{4.3}$$

If $\mathbf{Q}$ is selected such that (4.2) applies, we will have that $\|\mathbf{Q}\mathbf{s}(t)\| \approx 0$ even in the presence of an outlier, which allows us to approximate (4.3) as follows:

$$\|\mathbf{Q}\hat{\mathbf{x}}_{k+1}(t) - \mathbf{Q}\mathbf{x}(t) + \mathbf{Q}\mathbf{s}(t)\| \ \approx\ \|\mathbf{Q}\hat{\mathbf{x}}_{k+1}(t) - \mathbf{Q}\mathbf{x}(t)\|\,. \tag{4.4}$$

Therefore, such a restriction will enable the preservation of the image details even in the presence of large innovations. The question that arises is how to design the transformation matrix $\mathbf{Q}$ to achieve the required properties. We propose to base the design of matrix $\mathbf{Q}$ on a stochastic model for the innovations.

4.1 STATISTICAL PROPERTIES OF INNOVATION IN NATURAL IMAGE SEQUENCES

In this section we study the behavior of innovations in natural image sequences with the objective of determining a meaningful definition for the transformation matrix $\mathbf{Q}$.

The statistical properties of natural images have been thoroughly studied in the literature. A largely employed probabilistic model for natural images is characterized by a zero-mean and highly leptokurtotic, fat-tailed distribution, with its power spectral density remarkably close to a $1/f_r^{\rho}$ function, where f_r is the absolute spatial frequency and ρ is close to 2 (SCHAAF; HATEREN, 1996). This behavior led to the development, for example, of sparse derivative prior models for natural images (TAPPEN; RUSSELL; FREEMAN, 2003), which have been widely used in image processing algorithms.

When it comes to representing video sequences, however, obtaining accurate probabilistic models for the signals in the dynamic evolution of the sequence, particularly the innovations, is a more challenging task. This is due to the dependence of the signal statistics on the generally unknown movement in the scene. With the motion estimated from the observed and frequently low-resolution video sequence, the employed model must distinguish between errors originating from the image registration and those caused by true changes in the scene, which are often labeled as outliers in the community.

The modeling of large magnitude changes in the scene has already been considered for the image matching problem. Hasler et. al. (HASLER et al., 2003) proposed to consider the error patterns generated by non-coinciding regions of an aligned image pair to be similar to the error generated by comparing two random regions of the underlying scene. This relationship clearly arises in a dynamical model for a video sequence when the motion model fails to account for unpredictable changes between two adjacent images, generating an error signal that will consist of the difference between the new image and a misaligned part of the previous image. Considering the case of one dimensional signals for simplicity, the auto-correlation function of the difference between two patches of an image separated by Δ samples can be computed as:

$$\begin{aligned}\mathbf{r}_{\Delta}(l) &= \mathrm{E}[\{I(p) - I(p-\Delta)\}\{I(p-l) - I(p-\Delta-l)\}] \\ &= 2\mathbf{r}_I(l) - \mathbf{r}_I(l-\Delta) - \mathbf{r}_I(l+\Delta) \end{aligned} \tag{4.5}$$

where $\mathrm{E}[\,\cdot\,]$ denotes statistical expectation, $I(p)$ is a point in the one dimensional image, $\mathbf{r}_\Delta(l)$ is the auto-correlation of the simulated outlier, and $\mathbf{r}_I(l)$ is the image auto-correlation function. If the covariance between the image pixels diminishes with their distance, for a sufficiently large value of Δ the terms $\mathbf{r}_I(l \pm \Delta)$ will become approximately equal to the square of the mean image value. Therefore, the auto-correlation function of the simulated outlier will be similar to that of a natural image.

This interpretation can be more intuitively achieved by considering a different approach and modeling the innovations considering a scene model composed by the interactions of objects in an occlusive environment (LEE; MUMFORD; HUANG, 2001). Innovations in a video sequence can be broadly described as pixels in $\mathbf{x}(t)$ that cannot be described as a linear combination of the pixels in $\mathbf{x}(t-1)$ (i.e. are statistically orthogonal). These pixels will be here divided as

$$\mathbf{s}(t) = \mathbf{d}(t) + \boldsymbol{\eta}(t) \tag{4.6}$$

where $\boldsymbol{\eta}(t)$ consists of small changes on the scene originating, for example, from specular surfaces. $\boldsymbol{\eta}(t)$ can be modeled as a low power high frequency noise. $\mathbf{d}(t)$ represents large magnitude changes (outliers) arising due to occlusions or to objects suddenly appearing on the scene (such as image borders), and is usually sparse and compact (BAKER et al., 2011) [1].

A region of the scene corresponding to a dis-occluded area typically reveals part of a background or object at a different depth from the camera. Hence, the nonzero pixels in $\mathbf{d}(t)$ will consist of highly correlated compact regions. Furthermore, the joint pixel statistics should actually be similar to that of natural scenes in these locations. This conclusion becomes straightforward if we consider, for instance, the Dead Leaves image formation model (LEE; MUMFORD; HUANG, 2001), which characterizes a natural scene by a superposition of opaque objects of random sizes and positions occluding each other. Here, a dis-occluded area would correspond to the removal of an object (or a "leaf") at random from the topmost part of the z-axis. The corresponding region in the new image will therefore be composed of the next objects present on the z-axis. Since the area behind the view plane is completely filled with objects (superimposed "leaves") in this model, there is no difference between the statistical properties of a region in the foremost-top

[1] Note that $\mathbf{s}(t)$ is not to be confused with registration errors due to the ill-posed nature of the motion estimation process. The latter can be shown to originate from a random linear combination of the pixels in $\mathbf{x}(t-1)$ (COSTA; BERMUDEZ, 2007).

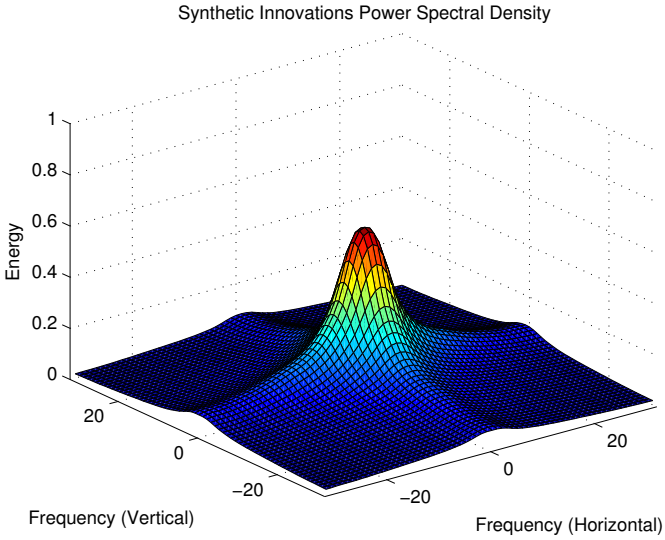


Figure 11 – Power spectral density of synthetically generated innovations.

image and those of a region behind an object. This reinforces the notion of correlation obtained by considering the more generic outlier model of Hasler et al. (HASLER et al., 2003).

To verify the proposed innovations model, we have determined the power spectral density (PSD) of synthetic images representing the innovations. These images were generated by pasting small pieces of the difference between two independent natural images with sizes ranging from 5×5 to 15×15 in random positions of a 64×64 background. We have extracted the small pieces from 20 different natural images, so that they emulate small regions appearing in the occluded regions of a video sequence. The PSD is computed by averaging 200 realizations of a Monte Carlo (MC) simulation. Figure 11 shows the obtained result. It can be clearly seen that the energy is concentrated in the lower frequencies of the spectrum, resulting in a highly correlated signal.

4.2 CHOOSING THE OPERATOR $\mathbf{Q}$

Natural scene innovations tend to be highly correlated in space. Thus, their energy tends to be primarily concentrated in the low spatial frequencies. Hence, the operator $\mathbf{Q}$ should in general emphasize the high frequency components to accomplish the design objectives in (4.2). Unfortunately, the specific scenes to be processed are not known in advance, what hinders the determination of the statistical properties of the innovations, and thus of the optimal operator $\mathbf{Q}$. A simple solution with reduced computational complexity is to use a simple high-pass filter with small support, such as a differentiator or a Laplacian. For simplicity, the Laplacian filter mask will be employed during the remaining of this work. Thus, we shall use $\mathbf{Q} = \mathbf{S}$, leaving the search of an optimal operator for a future work.

4.3 A FAST CONVERGENCE SRR ALGORITHM WITH ROBUSTNESS TO INNOVATIONS

To derive the new algorithm, we propose a new cost function that minimizes the perturbation only on the details of the reconstructed image, while at the same time observing the objectives of the R-LMS algorithm. Differently from (3.8), the new cost function allows for more flexibility for the component of the solution in the subspace corresponding to the outlier while retaining its quality. Such strategy leads to an increased algorithm robustness.

We propose to solve the following optimization problem:

$$\hat{\mathbf{x}}_{k+1}(t) = \arg\min_{\mathbf{z}} \Big\{ \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) + \big(\mathbf{z} - \hat{\mathbf{x}}_k(t)\big)^{\mathrm{T}} \nabla \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) + \frac{1}{\mu} \left\|\mathbf{z} - \hat{\mathbf{x}}_k(t)\right\|^2 + \frac{1}{\alpha_{\mathrm{T}}\mu} \left\|\mathbf{Q}\mathbf{z} - \mathbf{Q}\mathbf{G}(t)\hat{\mathbf{x}}(t-1)\right\|^2 \Big\} \tag{4.7}$$

Calculating the gradient of the cost function with respect to $\mathbf{z}$ and setting it equal to $\mathbf{0}$ we have:

$$\mathbf{0} = \nabla \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) + \frac{2}{\mu}\big(\mathbf{z} - \hat{\mathbf{x}}_k(t)\big) + \frac{2}{\alpha_{\mathrm{T}}\mu}\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\big(\mathbf{z} - \mathbf{G}(t)\hat{\mathbf{x}}(t-1)\big)$$

$$\left(\frac{2}{\mu}\mathbf{I} + \frac{2}{\alpha_{\mathrm{T}}\mu}\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\right)\mathbf{z} = -\nabla \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) + \frac{2}{\mu}\hat{\mathbf{x}}_k(t) + \frac{2}{\alpha_{\mathrm{T}}\mu}\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\mathbf{G}(t)\hat{\mathbf{x}}(t-1)$$

Solving for $\mathbf{z} = \hat{\mathbf{x}}_{k+1}(t)$ and approximating the statistical expectations by their instantaneous values yields the iterative equation for the new algorithm:

$$\hat{\mathbf{x}}_{k+1}(t) = \left(\mathbf{I} + \frac{1}{\alpha_{\mathrm{T}}}\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\right)^{-1}\left\{\hat{\mathbf{x}}_k(t) + \frac{1}{\alpha_{\mathrm{T}}}\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\mathbf{G}(t)\hat{\mathbf{x}}(t-1) - \mu\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\big[\mathbf{D}\mathbf{H}\hat{\mathbf{x}}_k(t) - \mathbf{y}(t)\big] - \mu\alpha\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_k(t)\right\}, \quad (4.8)$$

where the time update is based on the signal dynamics (2.3) and performed by $\hat{\mathbf{x}}_0(t+1) = \mathbf{G}(t+1)\hat{\mathbf{x}}_K(t)$ (ELAD; FEUER, 1999b).

It is clear that the proposed algorithm is a generalization of the R-LMS algorithm and of the classic least perturbation approach (3.8), as it collapses to these solutions if $\alpha_{\mathrm{T}} \to \infty$ or $\mathbf{Q} = \mathbf{I}$, respectively.

Algorithm (4.8) should have a good performance both with and without the presence of the outliers, at the cost of little additional computational effort. A rough comparison with the R-LMS reveals the need of 3 additional matrix-vector multiplications per iteration, along with the storage of a filtered and registered version of the previous estimation, since the matrix inversion can be computed *a priori* for a fixed α_{T}. Although this matrix is usually sparse, its storage is rather costly. If the operator $\mathbf{Q}$ is chosen to be a block circulant (BC) matrix (such as a Laplacian), then $\left(\mathbf{I} + \frac{1}{\alpha_{\mathrm{T}}}\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\right)^{-1}$ is known to be block circulant as well (MAZANCOURT; GERLIC, 1983), and can therefore be computed as a convolution, leading to important memory savings.

One should note that although (4.8) may resemble the Gradient Projection Method (GPM) (BERTSEKAS, 1976), this is not generally true, as $\left(\mathbf{I} + \frac{1}{\alpha_{\mathrm{T}}}\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\right)^{-1}$ is not necessarily a projection matrix (i.e. $\mathbf{M}^2 \neq \mathbf{M}$). As a consequence, the convergence and stability results obtained for the GPM algorithm cannot be directly applied to (4.8), making the performance analysis and parameter design more difficult.

4.4 A SIMPLIFIED ALGORITHM

Whereas algorithm (4.8) should present a good cost-benefit ratio, the aforementioned limitations motivates the pursuit of another algorithm that trades a small performance loss for both a reduction in memory cost and a more predictable performance. This section describes one possible modification.

Since the details of the solution are minimally disturbed between

iterations, we can safely assume that $\mathbf{Q}\hat{\mathbf{x}}_{k+1}(t) \approx \mathbf{Q}\hat{\mathbf{x}}_k(t)$. Therefore, we can employ a linear approximation for the quadratic regularization introduced in the last term of (4.7). Using a first-order Taylor series expansion of this norm with respect to the transformed variable $\mathbf{Qz}$ about the point $\mathbf{Qz} = \mathbf{Q}\hat{\mathbf{x}}_k(t)$. The resulting cost function can be written as:

$$\begin{aligned}\hat{\mathbf{x}}_{k+1}(t) =& \arg\min_{\mathbf{z}} \Bigg\{ \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) + \big(\mathbf{z} - \hat{\mathbf{x}}_k(t)\big)^{\mathrm{T}} \nabla \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) \\ &+ \frac{1}{\mu}\left\|\mathbf{z} - \hat{\mathbf{x}}_k(t)\right\|^2 + \frac{1}{2\tilde{\alpha_{\mathrm{T}}}\mu}\left\|\mathbf{Q}\hat{\mathbf{x}}_k(t) - \mathbf{QG}(t)\hat{\mathbf{x}}(t-1)\right\|^2 \\ &+ \frac{1}{2\tilde{\alpha_{\mathrm{T}}}\mu} 2\big\{\mathbf{Q}\hat{\mathbf{x}}_k(t) - \mathbf{QG}(t)\hat{\mathbf{x}}(t-1)\big\}^{\mathrm{T}}\big\{\mathbf{Qz} - \mathbf{Q}\hat{\mathbf{x}}_k(t)\big\} \Bigg\} \end{aligned} \tag{4.9}$$

$$\begin{aligned}=& \arg\min_{\mathbf{z}} \Bigg\{ \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) + \big(\mathbf{z} - \hat{\mathbf{x}}_k(t)\big)^{\mathrm{T}} \nabla \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) \\ &+ \frac{1}{\mu}\left\|\mathbf{z} - \hat{\mathbf{x}}_k(t)\right\|^2 + \frac{1}{\tilde{\alpha_{\mathrm{T}}}\mu}\big\{\mathbf{Q}\big(\mathbf{z} - \mathbf{G}(t)\hat{\mathbf{x}}(t-1)\big)\big\}^{\mathrm{T}} \\ &\qquad \cdot \big\{\mathbf{Q}\big(\hat{\mathbf{x}}_k(t) - \mathbf{G}(t)\hat{\mathbf{x}}(t-1)\big)\big\} \Bigg\}. \end{aligned} \tag{4.10}$$

It can be seen that the last term in (4.10) can be written as:

$$\begin{aligned}&\big\{\mathbf{Q}\big(\hat{\mathbf{x}}_{k+1}(t) - \mathbf{G}(t)\hat{\mathbf{x}}(t-1)\big)\big\}^{\mathrm{T}}\big\{\mathbf{Q}\big(\hat{\mathbf{x}}_k(t) - \mathbf{G}(t)\hat{\mathbf{x}}(t-1)\big)\big\} \\ &\quad = \left\|\mathbf{Q}\hat{\mathbf{x}}_{k+1}(t) - \mathbf{QG}(t)\hat{\mathbf{x}}(t-1)\right\|^2 \\ &\quad - \big\{\mathbf{Q}\big(\hat{\mathbf{x}}_{k+1}(t) - \mathbf{G}(t)\hat{\mathbf{x}}(t-1)\big)\big\}^{\mathrm{T}}\big\{\mathbf{Q}\big(\hat{\mathbf{x}}_{k+1}(t) - \hat{\mathbf{x}}_k(t)\big)\big\}, \end{aligned} \tag{4.11}$$

where it becomes apparent that the perturbation of the details in time $\mathbf{Q}(\hat{\mathbf{x}}_{k+1}(t) - \mathbf{G}(t)\hat{\mathbf{x}}(t-1))$ is being minimized, while at the same time changes between two consecutive iterations $\mathbf{Q}\big(\hat{\mathbf{x}}_{k+1}(t) - \hat{\mathbf{x}}_k(t)\big)$ are allowed as long as they are correlated with the changes in time. Nevertheless, as changes in the details are small (i.e. $\mathbf{Q}\hat{\mathbf{x}}_{k+1}(t) \approx \mathbf{Q}\hat{\mathbf{x}}_k(t)$) this term will not have a significant influence on the solution.

Note that if the algorithm initialization is selected as $\hat{\mathbf{x}}_0(t) = \mathbf{G}(t)\hat{\mathbf{x}}_K(t-1)$ (ELAD; FEUER, 1999b), the linearized regularization introduced in the last term of (4.9) is equal to zero for the first iteration ($k = 1$). Therefore, $K \geqslant 2$ iterations per time instant are necessary in

order to have an improvement over the R-LMS algorithm. This is not the case for the algorithm proposed in (4.8), where an improvement can be obtained even for $K = 1$. It can be seen that optimization problem (4.9) is equivalent to:

$$\hat{\mathbf{x}}_{k+1}(t) = \arg\min_{\mathbf{z}} \mathcal{L}(\hat{\mathbf{x}}_k(t)) + \big(\mathbf{z} - \hat{\mathbf{x}}_k(t)\big)^{\mathrm{T}} \Big\{ \nabla \mathcal{L}_{\text{R-MS}}(\hat{\mathbf{x}}_k(t)) + 2\alpha_{\mathrm{T}} \mathbf{Q}^{\mathrm{T}} \mathbf{Q} \big(\hat{\mathbf{x}}_k(t) - \mathbf{G}(t)\hat{\mathbf{x}}(t-1)\big) \Big\} + \frac{1}{\mu} \|\mathbf{z} - \hat{\mathbf{x}}_k(t)\|^2 \tag{4.12}$$

where $\alpha_{\mathrm{T}} = \frac{1}{2\alpha_{\mathrm{T}}\mu}$. It is clear that equation (4.12) assumed the form of a standard gradient descent algorithm, where the Lagrangian being optimized at a single time instant t in this case is given by:

$$\mathcal{L}(\hat{\mathbf{x}}(t)) = \mathrm{E}\Big\{ \|\mathbf{D}\mathbf{H}\hat{\mathbf{x}}(t) - \mathbf{y}(t)\|_2^2 + \alpha_{\mathrm{T}} \|\mathbf{Q}\hat{\mathbf{x}}(t) - \mathbf{Q}\mathbf{G}(t)\hat{\mathbf{x}}(t-1)\|_2^2 + \alpha \|\mathbf{S}\hat{\mathbf{x}}(t)\|_2^2 \Big| \hat{\mathbf{x}}(t) \Big\}. \tag{4.13}$$

Similarly to (3.8), by using $\mathbf{Q} = \mathbf{I}$ on (4.13), the algorithm particularizes to the classical Temporal Regularization case, which is not expected to be robust since the outliers are not accounted for. In this case, it can also be seen that the proposed algorithm becomes equivalent to the well known temporal regularization, commonly employed in simultaneous video SRR in order to achieve temporal consistency (BARNER; SARHAN; HARDIE, 1999; ZIBETTI; MAYER, 2007).

Differentiating the Lagrangian in (4.13) with respect to $\hat{\mathbf{x}}(t)$, we obtain

$$\nabla \mathcal{L}(\hat{\mathbf{x}}(t)) = 2\alpha_{\mathrm{T}} \mathbf{Q}^{\mathrm{T}} \big[\mathbf{Q}\hat{\mathbf{x}}(t) - \mathbf{Q}\mathbf{G}(t)\hat{\mathbf{x}}(t-1)\big] + 2\alpha \mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}(t) + 2\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}} \big\{ \mathbf{D}\mathbf{H}\hat{\mathbf{x}}(t) - \mathrm{E}[\mathbf{y}(t)] \big\}. \tag{4.14}$$

The steepest descent algorithm updates $\hat{\mathbf{x}}(t)$ in the negative direction of the gradient, resulting in

$$\hat{\mathbf{x}}_{k+1}(t) = \hat{\mathbf{x}}_k(t) - \frac{\mu}{2} \nabla \mathcal{L}(\hat{\mathbf{x}}(t)). \tag{4.15}$$

By applying the stochastic approximation for the statistical expecta-

tions by its instantaneous values, we obtain the new algorithm based on the linearized version of the proposed regularization:

$$\hat{\mathbf{x}}_{k+1}(t) = \hat{\mathbf{x}}_k(t) - \mu\alpha_{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\big[\mathbf{Q}\hat{\mathbf{x}}_k(t) - \mathbf{Q}\mathbf{G}(t)\hat{\mathbf{x}}(t-1)\big] - \mu\alpha\mathbf{S}^{\mathrm{T}}\mathbf{S}\hat{\mathbf{x}}_k(t) - \mu\mathbf{H}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\big[\mathbf{D}\mathbf{H}\hat{\mathbf{x}}_k(t) - \mathbf{y}(t)\big], \quad (4.16)$$

which is the iterative update equation for a fixed t and for $k = 1, \ldots, K$. Like the traditional R-LMS, the time update of (4.16) is based on the signal dynamics (2.3), and performed by $\hat{\mathbf{x}}_0(t+1) = \mathbf{G}(t+1)\hat{\mathbf{x}}_K(t)$ (ELAD; FEUER, 1999b).

4.5 COMPUTATIONAL COST OF THE PROPOSED SOLUTION

The computational and memory costs of the proposed solutions are still comparable to those of the R-LMS algorithm. An important property of the problem that allows a fast implementation of both the (R)-LMS and the proposed methods is the spatial invariance assumption of the operators $\mathbf{M} = (\mathbf{I} + \frac{1}{\alpha_{\mathrm{T}}}\mathbf{Q}^{\mathrm{T}}\mathbf{Q})^{-1}$, $\mathbf{H}$, $\mathbf{S}$ and $\mathbf{Q}$, which results in them being block-circulant matrices. In this case, the corresponding matrix-vector products can be computed in the form of a bidimensional convolution of the image by the respective convolution masks $\mathbf{m}$, $\mathbf{h}$, $\mathbf{s}$, and $\mathbf{q}$.

The computational and memory costs for the algorithms considered in this work can be seen in Tables 1 and 2. Since the cardinality of the convolution masks (denoted by $|\cdot|$) is usually much smaller than that of the HR image, the convolutions can be efficiently computed in the spatial domain. Therefore, the number of operations performed by the algorithms scale linearly with the number of HR image pixels M^2. The motion between the frames can be estimated using fast image registration algorithms such as that proposed by Caner et al. (CANER et al., 2006), which have an approximate computational cost of $\kappa^2 M^2 + g_{\max}^2 M^2 + M^2$ (with κ being the size of a small image block and $g_{\max}$ being the maximum displacement value). Using this registration algorithm, the computational cost in floating point operations per second of super resolving a video sequence at 30 frames per second for the algorithm in (4.16) is approximately given by

$$\approx 30\big(\kappa^2 + g_{\max}^2 + 1 + K(3|\mathbf{h}| + 2|\mathbf{s}| + 2|\mathbf{q}| + 3)\big)M^2\,.$$

As an example, supposing $\kappa = 3$, $g_{\max} = 10$, $|\mathbf{h}| = 30$, $|\mathbf{s}| =$

$|\mathbf{q}| = 10$ and $K = 2$, the aggregate computational cost becomes approximately $10^4 M^2$ operations per second. Considering the reconstruction of images of size $M = 1000$ pixels, this results in $\approx 10^{10}$ operations per second, which lies within the range of present-day video board devices. This illustrates the suitability of the proposed solution for real time processing applications (PHILLIPS, 2009).

	Memory
LMS	$M^2 + \|\mathbf{h}\|$
R-LMS	$M^2 + \|\mathbf{h}\| + \|\mathbf{s}\|$
Proposed 1 (4.8)	$2M^2 + \|\mathbf{h}\| + \|\mathbf{s}\| + \|\mathbf{m}\| + \|\mathbf{q}\|$
Proposed 2 (4.16)	$2M^2 + \|\mathbf{h}\| + \|\mathbf{s}\| + \|\mathbf{q}\|$

Table 1 – Memory cost of the algorithms.

	Operations
LMS	$3\|\mathbf{h}\|M^2 + 2M^2$
R-LMS	$3\|\mathbf{h}\|M^2 + 2M^2 + 2\|\mathbf{s}\|M^2$
Proposed 1 (4.8)	$3\|\mathbf{h}\|M^2 + 2M^2 + 2\|\mathbf{s}\|M^2 + 2\|\mathbf{q}\|M^2 + \|\mathbf{m}\|M^2$
Proposed 2 (4.16)	$3\|\mathbf{h}\|M^2 + 2\|\mathbf{s}\|M^2 + 2\|\mathbf{q}\|M^2 + 3M^2$

Table 2 – Computational cost per iteration of the algorithms (additions and multiplications, surplus additions, and re-samplings were considered).

5 RESULTS

In this chapter the performance of the proposed methods is evaluated through three examples. The objective of the first example is to evaluate the algorithm average performance without outliers, in a close-to-ideal environment. Therefore, synthetically generated video sequences with small translational motion are used to enable the execution of Monte Carlo simulations and the control of the occurrence of modeling errors. The motion between the frames is also assumed to be known *a priori*. Moreover, the computation of the mean squared reconstruction error is also made possible as we have access to the desired HR images. Since the new regularization introduced in algorithms (4.8) and (4.16) depends on a registered version of the estimated image on the previous time instant, a decline in the algorithm performance is expected in the presence of inaccurate motion estimation, as already reported for the case of the classical temporal regularization algorithms in (CHOI; GALATSANOS; KATSAGGELOS, 1996). Therefore, to evaluate the influence of motion estimation, this simulation is also performed using a typical registration algorithm to estimate the relative position of the frames.

In the second example, the performance of the proposed algorithms under the presence of innovation outliers is evaluated. A synthetic simulation emulates the case of a *flying bird* when an object suddenly appears in a frame or moves independently of the background, generating occlusions and leading to a high level of innovations in some specific frames of the video sequence.

Finally, the third example is devised to illustrate the algorithm performance when super-resolving real video sequences. The algorithm is tested in the presence of complex motion patterns and frames with large levels of innovations and registration errors.

For both algorithms proposed in this work, the matrix $\mathbf{Q}$ employed was chosen to be a Laplacian filter. For the case of the identity matrix $\mathbf{Q} = \mathbf{I}$, the method of (4.16) particularizes to the classical temporal regularization largely employed on the literature (CHOI; GALATSANOS; KATSAGGELOS, 1996; ZIBETTI; MAYER, 2007). No improvement (quantitative or perceptual) could be obtained in this case when compared to the R-LMS algorithm (with $\mathbf{Q} = 0$). Therefore, these cases are not reported here.

The boundary condition for the convolution matrices was chosen to be circulant, which allows for simplicity of implementation and

makes the inverse computed in (4.8) a circulant matrix as well (MAZANCOURT; GERLIC, 1983) (therefore also implementable as a convolution). The boundary condition for the motion matrix $\mathbf{G}(t)$ for the global translational case was selected to be circulant as well in order to allow a simple implementation, whereas for the case of a dense motion field these regions were linearly interpolated based on their neighboring pixels.

5.1 EXAMPLE 1

In this example, the performance of the proposed method is evaluated through a Monte Carlo (MC) simulation with 50 realizations. In order to enable the evaluation of the mean square reconstruction error (MSE), each HR video sequence is created based on the translation of an 256×256 window over a static image, resulting in whole-image translational movements. The window displacements consisted in a random walk process (i.i.d. unitary steps) on both horizontal and vertical directions. The still images considered for generating each video sequence consisted of natural scenes such as *Lena*, *Cameraman*, *Baboon* and others, and were totally distinct from each other. The resulting sequence was then blurred with an uniform unitary gain 3×3 mask and decimated 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 decimated images.

Since the behaviors of both the proposed methods and the R-LMS are highly dependent on the selected parameters (as demonstrated in the examples of Chapter 3.1), the parameters must be carefully selected to yield an honest comparison. The parameters for each method were selected to achieve the minimum MSE in steady-state operation (i.e. for large t). The steady-state MSE for each set of parameters was estimated by running an exhaustive search over a small, independent set of images and averaging the MSE in the last 5 frames. The parameter values that resulted in the best performance are presented in Table 3.

We applied both standard and regularized versions of the LMS as well as the proposed methods to super resolve the synthetic sequences, all initialized with $\hat{\mathbf{x}}(1)$ as a bicubic interpolation of the first LR image. The motion vectors were first considered to be known *a priori* in order to remove the influence of a registration algorithm, and $K = 2$ iterations per time instant were used (notice that due to the time update

	LMS	**R-LMS**	**Proposed 1** (4.8)	**Proposed 2** (4.16)
μ	2	2.75	1.15	3
α	–	5×10^{-4}	1.5×10^{-4}	1×10^{-4}
α_{T}	–	–	82	0.02

Table 3 – Parameter values used on the simulations with the with outlier-free sequences.

in (2.13), when $K = 1$ the method proposed in (4.16) particularizes to the R-LMS algorithm in (2.12)). The super-resolved sequences were compared to the original HR one and the mean squared error (MSE) was computed across all realizations.

The MSE performance is depicted on Figure 12. It can be seen that the proposed methods outperform the traditional LMS versions as is shown through a reduction of the MSE. Furthermore, both algorithms proposed in this work achieve the same mean square error given enough time instants. Besides, the algorithm using the approximated version of the temporal disturbance in (4.16) reaches the minimum MSE faster than that using the original regularization in (4.8) in the absence of outliers.

In order to evaluate the algorithms in a more realistic scenario, the MC simulation is repeated considering the influence of registration errors. To accomplish this, the *Horn & Schunck* registration algorithm as provided by (SUN; ROTH; BLACK, 2010) is employed[1], with the velocity fields being averaged across the entire image in order to compute the global displacements. The algorithm parameters were the same used in the previous simulation. The resulting MSE is depicted in Figure 13, and an example of a reconstructed image of a resolution test chart can be seen in Figure 14.

It can be seen that although the differences between the performances of the different algorithms decreased in the presence of registration errors, the proposed methods still outperform the conventional (R)-LMS algorithms. Nevertheless, as we are basically preserving information (details) from previous frames which must be registered, large levels of registration errors reduce the effectiveness of the proposed methods. Furthermore, the performance of the method presented in (4.8) showed greater sensitivity to unknown registration as its per-

[1]The parameters were set as: `lambda=1×10^3`, `pyramid_levels=4`, `pyramid_spacing=2`.

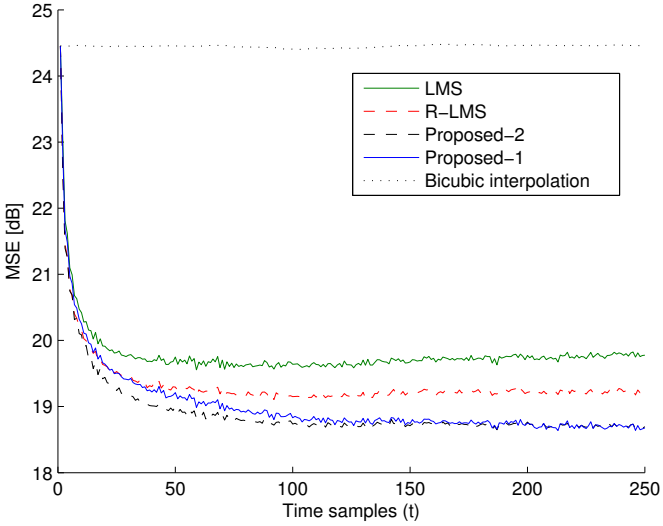


Figure 12 – Average MSE per pixel for the known motion case.

formance degraded more when compared to (4.16). The reconstructed images were found to be perceptually similar among the four evaluated algorithms, although a careful inspection reveals a slight improvement in the reconstruction result using the method given in (4.16). Nevertheless, as it will be illustrated in the following examples, when evaluated in the presence of outliers the proposed methods perform considerably better than the remaining algorithms.

5.2 EXAMPLE 2

This example evaluates the algorithm robustness to innovation outliers by means of super-resolving synthetic video sequences containing a suddenly appearing object, which is independent from the background. Therefore, the first MC simulation of *Example 1* will be repeated, this time with the inclusion of an $N \times N$ black square appearing in the middle of the 32nd frame of every sequence and disappearing in the 35th frame, emulating the behaviour of a *flying bird* outlier on the scene.

The simulation is first performed for the set of parameters shown

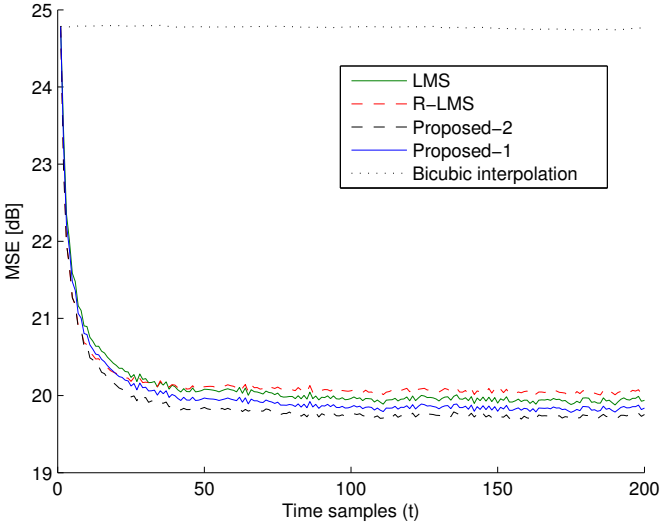


Figure 13 – Average MSE per pixel in the presence of registration errors.

in Table 3, which have been used in the previous example. The MSE evolution is depicted in Figure 15-(a) and 15-(c). It can be seen that the estimated solutions deviate significantly from the desired image in the presence of an outlier. Comparing the different algorithms, it can be noticed that the method proposed in (4.8) offer a slight improvement in comparison with the LMS and R-LMS algorithms, whereas the improvement achieved by the method presented in (4.16) is much more significant. These results suggest that the algorithm (4.16) is the most robust to outliers among all tested algorithms.

A performance improvement may be achieved at the cost of some loss in steady-state performance. To illustrate this point, we have determined an alternative set of parameters to obtain the minimum MSE averaged between frames 30 and 40. We did this by performing an exhaustive search for reconstructing a small independent set of images. The resulting parameters are presented in Table 4. The MSE evolution is depicted in Figures 15-(b) and 15-(d).

It can be noticed that both proposed methods provide a significant performance gain when compared to the remaining algorithms in the presence of outliers in frames 32 and 35, where the method of (4.8) performed slightly better than that of (4.16). Although the

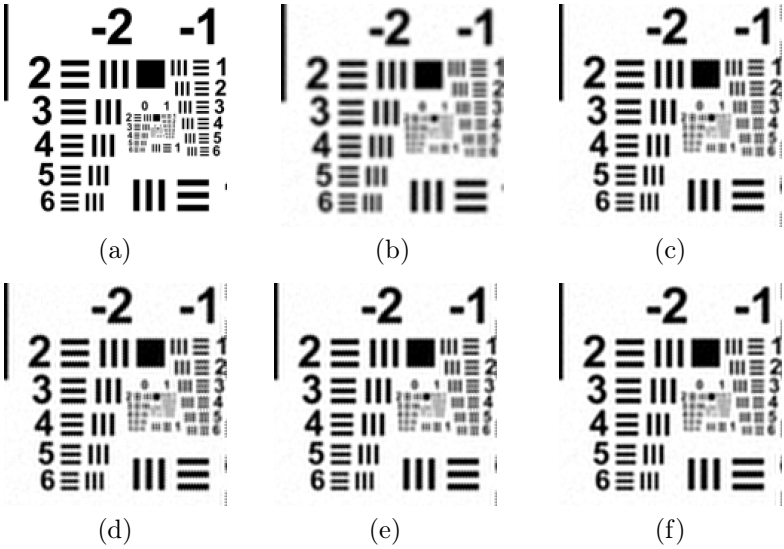


Figure 14 – Sample of the 200th reconstructed frame. (a) Original image. (b) Bicubic interpolation (MSE=30.13dB). (c) LMS (MSE=25.77dB). (d) R-LMS algorithm (MSE=25.54dB). (e) Proposed Method 1 (4.8) (MSE=25.65dB). (f) Proposed Method 2 (4.16) (MSE=25.47dB).

R-LMS MSE was similar to that achieved by the proposed methods for frames 33 and 34 (when the black square remained on the sequence), its steady-state performance decreased considerably (i.e. for large t). While the steady-state performances of the proposed algorithms also decreased when compared with the simulation employing the parameters in Table 3, the difference between them and the LMS and R-LMS remained significant. Moreover, algorithm (4.8) was slightly more affected when compared with (4.16), which showed to be less sensitive to the parameter selection, performing reasonably well in both simulations.

A visual inspection of the reconstructed images on frame 32 portrayed in Figure 16 endorse the quantitative result. The black square introduced in the sequence is significantly better represented for the proposed methods (when it is indeed present in the HR image), and a slight improvement can be noticed in the result of (4.8) when compared to (4.16).

	LMS	**R-LMS**	**Proposed 1** (4.8)	**Proposed 2** (4.16)
μ	4.7	4.2	2.2	3.4
α	–	40×10^{-4}	18×10^{-4}	1×10^{-4}
α_{T}	–	–	16	0.017

Table 4 – Parameter values used on the simulations considering the presence of outliers.

5.3 EXAMPLE 3

The objective of this example is to illustrate the effectiveness of the proposed methods when super-resolving real video sequences. We considered the *Foreman* sequence, which displays a closeup of a talking man before changing to a view of a building under construction. In this case, the true motions of the objects and camera are unknown, demanding the estimation of a dense velocity field, for which the *Horn & Shunck* algorithm will be adopted again (with the same parameters shown in Table 4 but now considering the displacement to be unique for each image pixel).

To allow for a quantitative evaluation of the reconstruction, the original video was used as an available HR image sequence. For simplicity only the 256×256 upper-right region of the original sequence was considered so that the resulting images were square. Like in *Example 1*, the HR sequence was blurred with an uniform unitary gain 3×3 mask, decimated by a factor of 2 and white Gaussian noise with variance $\sigma^2 = 10$ was then added to form the LR images. The standard LMS versions and the proposed methods were used to super-resolve the LR sequence, with $K = 2$ iterations per time sample and the parameters set at the values in Table 4. It is important to notice that, since this example contains a significant level of innovations, the parameters were chosen based on the context of *Example 2*. Nevertheless, they were not guaranteed to be optimal, as we were working with a single video sequence, with unknown motion and in the presence of registration errors.

The resulting mean square error evolution for the sequence is depicted in Figure 17, where it can be seen that the proposed methods performs quantitatively better than the traditional algorithms. Note that although the plots illustrate the behavior of the algorithms, it should not be used as a means to evaluate their average performance,

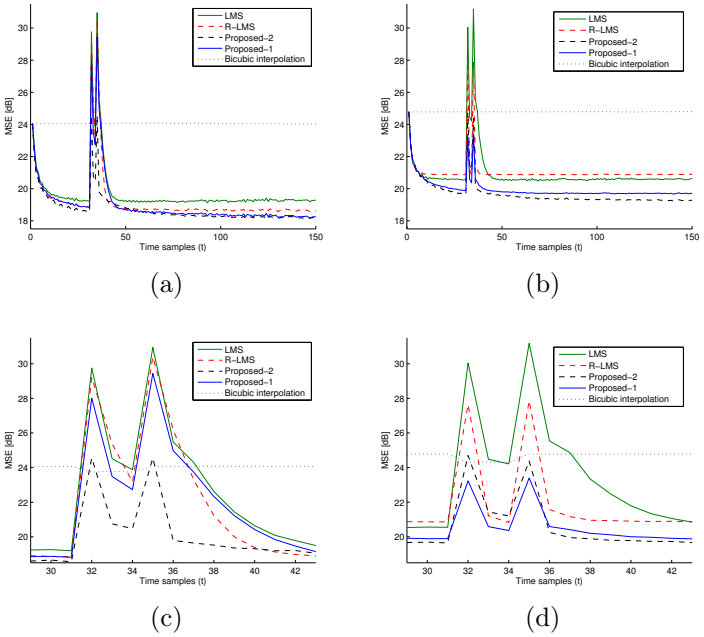


Figure 15 – MSE per pixel with an outlier at frames 32-35. (a) and (c): Full sequence and zoom with for reconstruction with parameters of Table 3. (b) and (d): Full sequence and zoom with for reconstruction with parameters of Table 4.

as it portrays a single realization and therefore is not statistically relevant. Furthermore, due to the significant level of innovations, the algorithm in (4.8) performs better than its version with the approximated regularization in (4.16), as already noted in *Example 2*.

It is also when a high degree of innovations is present in the scene that the improvement offered by the proposed methods can be most clearly observed (as for example in frames 180-200). Their performance is significantly better on these situations, with the reconstruction error exhibiting a more regular characteristic across the entire sequence, and not being considerably influenced by the outliers. To illustrate this scenario, the 93rd super-resolved frame is depicted on Figure 18, where the advantage of the proposed methods becomes apparent through a more clear reconstruction result, as opposed to a vast amount of artifacts found on the images reconstructed by the traditional algorithms,

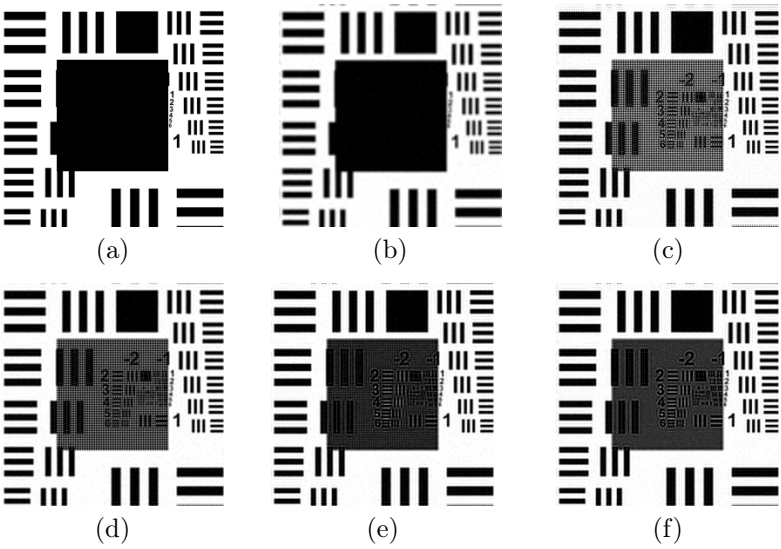


Figure 16 – Sample of 32th frame of a reconstructed sequence. (a) Original image (the black square is present in the desired image). (b) Bicubic interpolation (MSE=30.21dB). (c) LMS (MSE=34.82dB). (d) R-LMS algorithm (MSE=32.36dB). (e) Proposed Method 1 (4.8) (MSE=29.72dB). (f) Proposed Method 2 (4.16) (MSE=28.21dB).

which mainly covers the regions where innovations are present.

At the time intervals where the amount of innovations is not so significant, all four compared methods perform similarly under the MSE criterion. Nevertheless, it is still possible to have a noticeable difference in the perceptual quality of the reconstructed images. An example of such a situation is illustrated with the reconstruction results for the 33rd frame depicted in Figure 19. Although for this frame the quantitative difference between the results of the four algorithms is only slight, the images super-resolved by the proposed methods still offers a better perceptual quality with reduced artifacts at places where small localized motion occur (particularly close to the man's mouth).

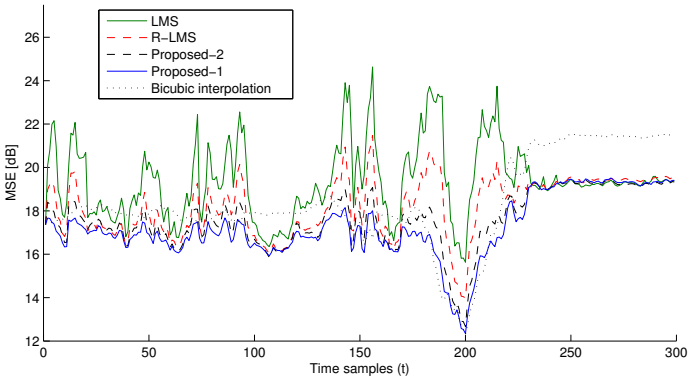


Figure 17 – Average MSE per pixel for the *Foreman* video sequence.

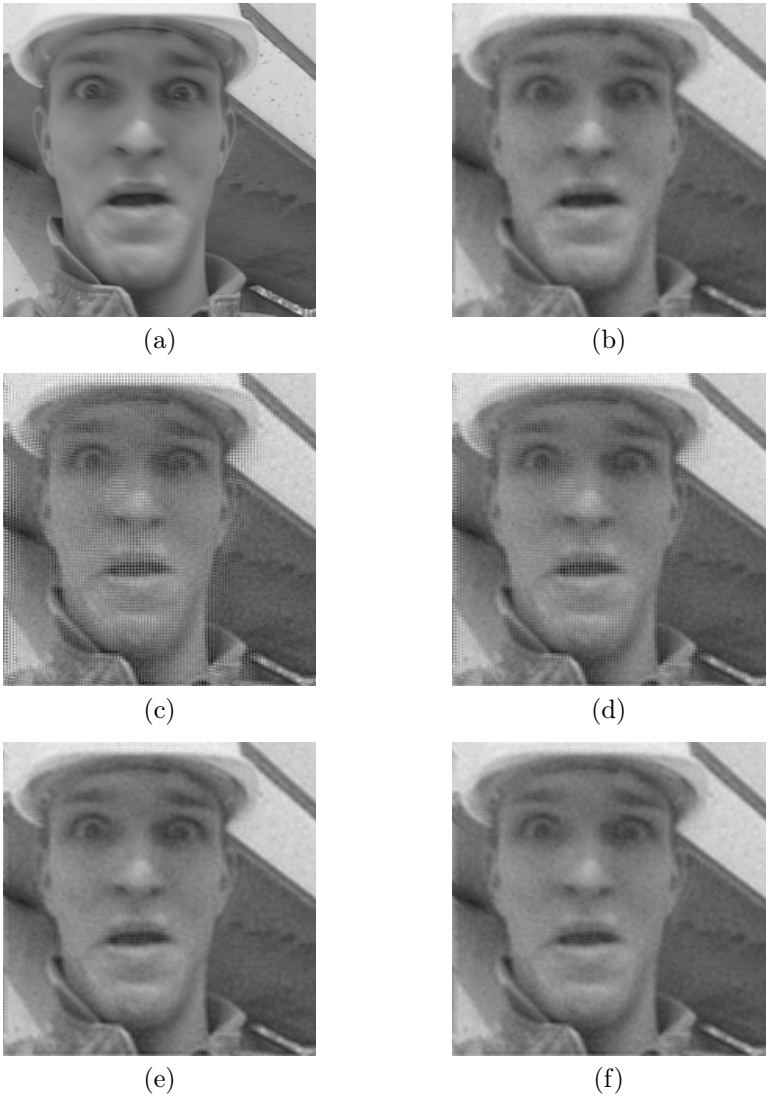
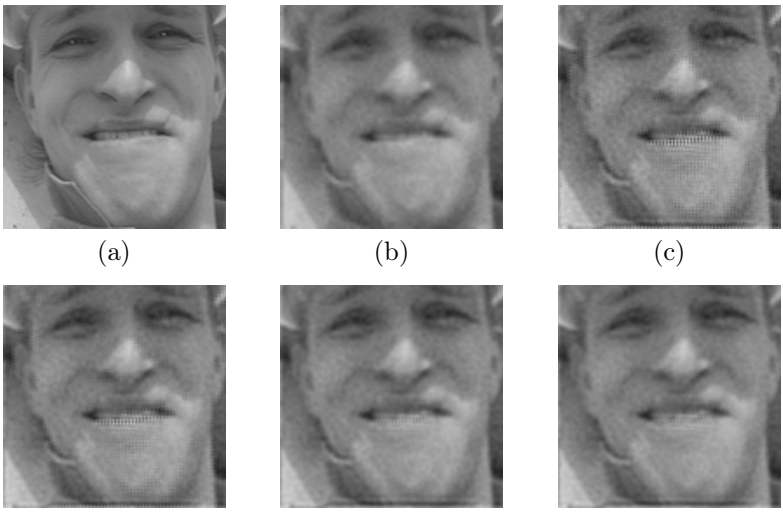


Figure 18 – Sample of the 93th reconstructed frame (with large innovation's level). (a) Original image. (b) Bicubic interpolation (MSE=17.47dB). (c) LMS (MSE=22.56dB). (d) R-LMS (MSE=20.14dB). (e) Proposed Method 1 (4.8) (MSE=17.50dB). (f) Proposed Method 2 (4.16) (MSE=18.38dB).



(a) (b) (c)

(d) (e) (f)

Figure 19 – Sample of the 33th reconstructed frame (with small innovation level). (a) Original image. (b) Bicubic interpolation (MSE=17.90dB). (c) LMS (MSE=17.49dB). (d) R-LMS (MSE=16.93dB). (e) Proposed Method 1 (4.8) (MSE=16.78dB). (f) Proposed Method 2 (4.16) (MSE=17.02dB).

6 CONCLUSIONS

This work proposed a new iterative super resolution reconstruction method aimed at an increased robustness to innovation outliers at real-time operation. Using the proximal point regularization formulation of the gradient descent algorithm employed in the Regularized Least Mean Squares methods (R-LMS) (ELAD; FEUER, 1999b), it was possible to attain a better understanding of the R-LMS algorithm behavior, specifically concerning the conflicting trade-off between the preservation of the image content estimated during previous time instants and the robustness to large innovations (outliers), which was shown to be present for the interesting case of a small number of gradient iterations per time instant K, since it was related to the slow speed of convergence of the gradient method.

Following this interpretation, a new regularization was proposed for the proximal optimization problem. By employing statistical information about the innovations in natural image sequences, it was possible to devise a regularization that allowed the estimated solution to converge faster in the subspace of the image related to the innovations, while at the same time preserving the image details which had been previously estimated. Hence, the proposed regularization increases the algorithm robustness to outliers without significantly decreasing the image quality in their absence, with only a modest increase in the resulting computational cost. A quadratic form of the proposed regularization, as well as a linear approximation were considered, resulting in two new methods. The latter could also be written as a gradient descent algorithm with an additional regularization in the form of a modified version of the temporal regularization already employed in the literature for simultaneous video restoration and SRR (CHOI; GALATSANOS; KATSAGGELOS, 1996; BORMAN; STEVENSON, 1999; ZIBETTI; MAYER, 2007).

It was shown that the proposed methods performed similarly to the traditional algorithms (LMS and R-LMS) in the absence of outliers, both with the motion known *a priori* or estimated with a registration algorithm. When large innovations were present in the video sequence, the proposed methods performed significantly better, both with synthetic and real video sequences, showing an increased robustness to outliers without a significant loss of performance during their absence.

6.1 SUGGESTIONS FOR FUTURE WORKS

Some topics are now suggested as possible extensions of this work:

- **An improved model for the innovations**: The choice of the matrix $\mathbf{Q}$ as a Laplacian filter that was employed in this work was selected based exclusively on an intuitive model for the innovation characteristics as a low frequency signal. A more detailed model for the outliers might provide the necessary information for the selection of a matrix $\mathbf{Q}$ that provides an increased, or even the best possible performance.

- **Analysis of the algorithm performance**: It is well known that the performance of super resolution algorithms in general is very sensitive to changes in the parameters (such as step size and regularization parameters for the R-LMS and for the proposed methods). Therefore, a theoretical evaluation of the algorithm performance is of great interest before their employment in a practical scenario, since it may allow for the adjustment of their parameters to achieve the best performance.

REFERENCES

AYVACI, A.; RAPTIS, M.; SOATTO, S. Sparse occlusion detection with optical flow. *International Journal of Computer Vision*, Springer, v. 97, n. 3, p. 322–338, 2012.

BAKER, S. et al. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, Springer, v. 92, n. 1, p. 1–31, 2011.

BARNER, K. E.; SARHAN, A. M.; HARDIE, R. C. Partition-based weighted sum filters for image restoration. *Image Processing, IEEE Transactions on*, IEEE, v. 8, n. 5, p. 740–745, 1999.

BECK, A.; TEBOULLE, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, SIAM, v. 2, n. 1, p. 183–202, 2009.

BELEKOS, S. P.; GALATSANOS, N. P.; KATSAGGELOS, A. K. Maximum a posteriori video super-resolution using a new multichannel image prior. *Image Processing, IEEE Transactions on*, IEEE, v. 19, n. 6, p. 1451–1464, 2010.

BERTSEKAS, D. P. On the goldstein-levitin-polyak gradient projection method. *Automatic Control, IEEE Transactions on*, IEEE, v. 21, n. 2, p. 174–184, 1976.

BERTSEKAS, D. P. Nonlinear programming. Athena scientific, 1999.

BORMAN, S.; STEVENSON, R. L. Simultaneous multi-frame map super-resolution video enhancement using spatio-temporal priors. In: IEEE. *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*. [S.l.], 1999. v. 3, p. 469–473.

CANER, G. et al. Local image registration by adaptive filtering. *IEEE Transactions on Image Processing*, IEEE, v. 15, n. 10, p. 3053–3065, 2006.

CAPEL, D. *Image Mosaicing and Super-Resolution (Cphc/Bcs Distinguished Dissertations.)*. [S.l.]: Springer, 2004. ISBN 1852337710.

CHOI, M. G.; GALATSANOS, N. P.; KATSAGGELOS, A. K. Multichannel regularized iterative restoration of motion compensated

image sequences. *Journal of Visual Communication and Image Representation*, Elsevier, v. 7, n. 3, p. 244–258, 1996.

COSTA, G. H. Estudo do comportamento do algoritmo lms aplicado à reconstrução de vídeo com super-resolução. Florianópolis, SC, 2007.

COSTA, G. H.; BERMUDEZ, J. C. M. Statistical analysis of the lms algorithm applied to super-resolution image reconstruction. *Signal Processing, IEEE Transactions on*, IEEE, v. 55, n. 5, p. 2084–2095, 2007.

COSTA, G. H.; BERMUDEZ, J. C. M. Informed choice of the lms parameters in super-resolution video reconstruction applications. *Signal Processing, IEEE Transactions on*, IEEE, v. 56, n. 2, p. 555–564, 2008.

ELAD, M.; FEUER, A. Super-resolution reconstruction of image sequences. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 21, n. 9, p. 817–834, 1999.

ELAD, M.; FEUER, A. Superresolution restoration of an image sequence: Adaptive filtering approach. *Trans. Img. Proc., IEEE*, IEEE Press, Piscataway, NJ, USA, v. 8, n. 3, p. 387–395, mar. 1999. ISSN 1057-7149. <http://dx.doi.org/10.1109/83.748893>.

FARSIU, S.; ELAD, M.; MILANFAR, P. A practical approach to superresolution. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Electronic Imaging 2006*. [S.l.], 2006. p. 607–703.

FARSIU, S. et al. Dynamic demosaicing and color superresolution of video sequences. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Optical Science and Technology, the SPIE 49th Annual Meeting*. [S.l.], 2004. p. 169–178.

FARSIU, S. et al. Fast and robust multiframe super resolution. *Image processing, IEEE Transactions on*, IEEE, v. 13, n. 10, p. 1327–1344, 2004.

GALATSANOS, N.; CHIN, R. Digital restoration of multi-channel images. In: IEEE. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'87*. [S.l.], 1987. v. 12, p. 1244–1247.

GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. 1. ed. [S.l.]: Prentice hall, 2002.

GUNTURK, B. K.; GEVREKCI, M. High-resolution image reconstruction from multiple differently exposed images. *Signal Processing Letters, IEEE*, IEEE, v. 13, n. 4, p. 197–200, 2006.

HASLER, D. et al. Outlier modeling in image matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 25, n. 3, p. 301–315, 2003.

HAYKIN, S. *Adaptive Filter Theory*. 2. ed. [S.l.]: Prentice Hall, 1991.

KIM, M. et al. Robust dynamic super resolution under inaccurate motion estimation. In: IEEE. *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*. [S.l.], 2010. p. 323–328.

LEE, A. B.; MUMFORD, D.; HUANG, J. Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *International Journal of Computer Vision*, Springer, v. 41, n. 1-2, p. 35–59, 2001.

LEE, E. S.; KANG, M. G. Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration. *Image Processing, IEEE Transactions on*, IEEE, v. 12, n. 7, p. 826–837, 2003.

MAZANCOURT, T. D.; GERLIC, D. The inverse of a block-circulant matrix. *Antennas and Propagation, IEEE Transactions on*, IEEE, v. 31, n. 5, p. 808–810, 1983.

MILANFAR, P. *Super-resolution imaging*. Boca Raton: CRC Press, 2010.

MODERSITZKI, J. *Numerical methods for image registration*. [S.l.]: Oxford university press, 2003.

OZKAN, M. K. et al. Efficient multiframe wiener restoration of blurred and noisy image sequences. *Image Processing, IEEE Transactions on*, IEEE, v. 1, n. 4, p. 453–476, 1992.

PARK, S. C.; PARK, M. K.; KANG, M. G. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE*, IEEE, v. 20, n. 3, p. 21–36, 2003.

PHILLIPS, E. CUDA accelerated linpack on clusters. In: *ACM/IEEE Int. Conf. on Supercomp., SC'09 (presentation)*. [S.l.: s.n.], 2009.

RICHTER, M. et al. Spatio-temporal regularization featuring novel temporal priors and multiple reference motion estimation. In: IEEE. *Broadband Multimedia Systems and Broadcasting (BMSB), 2011 IEEE International Symposium on*. [S.l.], 2011. p. 1–6.

SAYED, A. H. *Fundamentals of adaptive filtering*. [S.l.]: John Wiley & Sons, 2003.

SCHAAF, v. A. Van der; HATEREN, J. v. van. Modelling the power spectra of natural images: statistics and information. *Vision research*, Elsevier, v. 36, n. 17, p. 2759–2770, 1996.

SU, H.; WU, Y.; ZHOU, J. Adaptive incremental video super-resolution with temporal consistency. In: IEEE. *Image Processing (ICIP), 2011 18th IEEE International Conference on*. [S.l.], 2011. p. 1149–1152.

SUN, D.; ROTH, S.; BLACK, M. J. Secrets of optical flow estimation and their principles. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. [S.l.], 2010. p. 2432–2439.

TAPPEN, M. F.; RUSSELL, B. C.; FREEMAN, W. T. Exploiting the sparse derivative prior for super-resolution and image demosaicing. In: CITESEER. *In IEEE Workshop on Statistical and Computational Theories of Vision*. [S.l.], 2003.

TIAN, J.; MA, K.-K. A new state-space approach for super-resolution image sequence reconstruction. In: IEEE. *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. [S.l.], 2005. v. 1, p. I–881.

TIAN, J.; MA, K.-K. A state-space super-resolution approach for video reconstruction. *Signal, image and video processing*, Springer, v. 3, n. 3, p. 217–240, 2009.

TIAN, J.; MA, K.-K. A survey on super-resolution imaging. *Signal, Image and Video Processing*, Springer, v. 5, n. 3, p. 329–342, 2011.

WANG, Z.; QI, F. Super-resolution video restoration with model uncertainties. In: IEEE. *Image Processing. 2002. Proceedings. 2002 International Conference on*. [S.l.], 2002. v. 2, p. 852–856.

ZHAO, W.; SAWHNEY, H. S. Is super-resolution with optical flow feasible? In: *Computer Vision–ECCV 2002*. [S.l.]: Springer, 2002. p. 599–613.

ZIBETTI, M. V. W.; MAYER, J. A robust and computationally efficient simultaneous super-resolution scheme for image sequences. *Circuits and Systems for Video Technology, IEEE Transactions on*, IEEE, v. 17, n. 10, p. 1288–1300, 2007.