A BAYESIAN MULTIFRAME SUPERRESOLUTION ITERATIVE METHOD USING SSIM-BASED CONVERGENCE SUITABLE FOR REAL CASES

Thais Pedruzzi do Nascimento*, Evandro Ottoni Teatini Salles*

*Universidade Federal do Espírito Santo
Vitória, ES, 29075-910, Brasil

Emails: thais.p.nascimento@aluno.ufes.br, evandro.salles@ufes.br

Abstract— Multiframe superresolution is a method that generates a higher resolution image (HR estimated image) from several lower resolution images (LR images). A Bayesian approach is convenient for such purpose because it is possible to incorporate pre-known information in the form of *prior*. This work implements an SAR (Simultaneous Autoregressive) *prior*, which is better for softening noise, and a $\ell_1 - norm$ and a TV (Total Variation) *prior*, both known for the maintenance of edges. Besides, TVSAR and ℓ_1SAR are also implemented in order to take advantage of both possibilities. Iterative methods, namely Majorization Minimization and Expectation Maximizations, are used to manage the non-analytical nature of the ℓ_1 *priors* and adjust the motion parameters iteratively. SSIM and PSNR-based convergence conditions are used and compared, in terms of quality and time processing evaluation. Previous works use full reference image quality metrics (FR-IQA) to choose the output image, which makes them unsuitable for real cases. This work investigates a direct form of choice of the estimated image, among the ones of all iterations. Such choice enables the application of our algorithm to real cases, where the HR original image is unavailable.

Keywords— Multiframe Superresolution, *Priors*, SSIM, Convergence condition

Resumo— Superresolução *multiframe* é um método que gera uma imagem de maior resolução (imagem HR estimada) a partir de várias imagens de resolução mais baixa (Imagens LR). Uma abordagem Bayesiana é conveniente para tal propósito porque torna possível a incorporação de informações pré-estabelecidas na forma de *priori*. Este trabalho implementa uma *priori* SAR, que é melhor para suavizar ruído, e *prioris* norma-ℓ_1 e TV, ambas conhecidas por favorecer manutenção de borda. Além disso, as combinações TVSAR e ℓ_1SAR também são implementadas para que seja possível aproveitar as vantagens das duas possibilidades. O método utilizado é iterativo, para contornar a natureza não analítica das *priori* ℓ_1 e para se ajustar os parâmetros de movimento iterativamente. Utiliza-se convergências baseada em PSNR e SSIM e compara-se as duas, em termos de avaliação de qualidade e tempo de processamento. Investiga-se uma forma direta de escolha da imagem estimada final , entre as imagens estimadas de todas as iterações. Tal escolha possibilita a aplicação direta deste algoritmo a situação reais, onde a imagem original HR não é conhecida.

Palavras-chave— Superresolução *Multiframe*, *Prioris*, SSIM, Critério de convergência

1 Introduction

The term “Superresolution” (SR) first appeared in (Irani and Peleg, 1990) and describes methods focused on enhancing images. These methods generate a higher resolution image (HR estimated image) from a set of lower resolution ones (LR images). The size of such set characterizes the Superresolution method. If we have more than one LR image, it is called Multiframe Superresolution. Whereas, Singleframe Superresolution happens when the set has only one LR image. This work focuses on Multiframe Superresolution.

Generally, Multiframe SR methods work like this: there are the original grids from each LR image and the new bigger grid, according to which the estimated HR image is supposed to be adjusted. The LR images are displaced from each other in a subpixel manner, therefore, the pixels from the HR image are actually formed by a combination of their neighbors on the grids of the LR images. However, it is necessary to align the LR images to adjust them to the new bigger grid, and this process of alignment is called registration (Figure 1). The subpixel displacement enables the Multiframe Superresolution process, since it provides new information about each pixel, i.e., area of the scene.

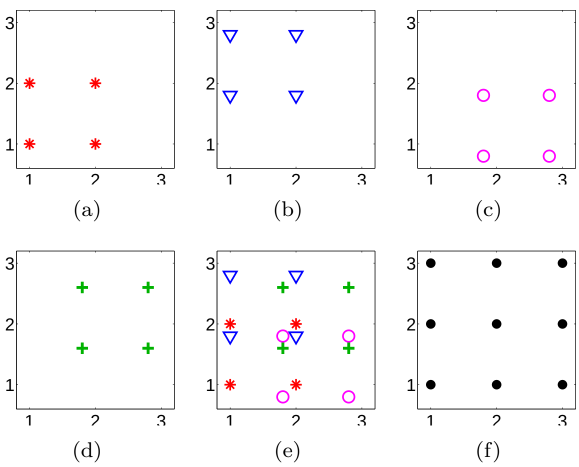


Figure 1: Scheme of pixel contribution of each LR image. (a)-(d) LR images and (e) combination after registration. Adapted from (Katsaggelos et al., 2007)

Every real imaging system changes the scene that is being registered and, therefore, they produce a deformed version of such scene, especially when the target object of the scene is far away from the camera (for example, in surveillance cameras). We can model the imaging system and, in this case, the input of the model is the HR image (representing the scene) and the output, the LR image (representing the observed image). An SR method focuses on obtaining the HR image - which is the latent variable from the model - using the LR image, the observed variable. For this reason, Superresolution is defined as a inverse problem. Moreover, the imaging system contains additive random noise, which increases the complexity of the problem.

A Bayesian approach is used in this work and others (Katsaggelos et al., 2007). Statistics frameworks, such as Bayes, are generally suitable for solving problems involving random noise. Besides, using Bayes is interesting since it brings the possibility of incorporating information in the form of prior probability density function, which is related to the HR image. The choice of the prior PDF (Probability Density Function), in our context, is associated with noise reducing and the maintenance of edges and other details. The SAR prior, in theory, reduces the noise, but also the edges. On the other hand, ℓ_1-norm and TV priors maintain more edges. The Bayesian approach sets up an optimization problem with the goal of finding the HR estimated image that maximizes the posterior PDF. The solution to this optimization problem is computed through iterative algorithms, using parameters of the imaging system model and parameters from the prior PDF itself: the so-called "hyperparameteres".

An important issue related to SR - and several other signal processing problems - is Quality Assessment. When it comes to images, the more precise way to define quality is by visual inspection, that is, just looking at the image and providing a subjective evaluation. However, this procedure is usually time-consuming and expensive (Wang et al., 2004). Hence the importance of objective metrics for image quality assessment. The concept of quality can be defined as similarity (or, conversely, error) between the tested image and a reference image. The methods based on such concept are called Full Reference Image Quality Assessment (FR-IQA). In this work, two FR-IQA methods are used: PSNR and SSIM.

The focus of this work is to investigate the effect of modifying an iterative method of SR in order to enable the use of such method in real situations. The rest of this Section explores the theoretical background used in this project and, finally, explains our motivation and objective. In Section 2, the reader finds the methodology used in the experiments. Section 3 shows the experimental results and Section 4 concludes the paper with a discussion about the results.

1.1 Imaging System Model

The imaging system model is written as

$$\mathbf{y_k} = \mathbf{A}\mathbf{H_k}\mathbf{C}(\mathbf{s_k})\mathbf{x} + \mathbf{n_k}, \tag{1}$$

where $\mathbf{y_k}$ is the k-th LR image, $\mathbf{k} = 1..L$ (L being the number of LR images), $\mathbf{x}$ is the original HR image and $\mathbf{n_k}$ is the additive noise. All of those are lexicographic representations of the respective signal. Besides, $\mathbf{A}$ is the down sampling matrix, $\mathbf{H_k}$ is the blurring matrix and $\mathbf{C}(\mathbf{s_k})$ is the warping matrix for the displacement vectors $\mathbf{s_k} = (\theta_l, c_k, d_k)^t$.

The blurring matrix describes the Point Spread Function of the camera and others deformations. The warping matrix describes the displacement among the LR images like the ones in Figure 1. And the down sampling matrix represents the scale between the LR images and the HR image, that is, this matrix defines how much the size of the image is increased by Superresolution. The random additive noise is Gaussian with zero-mean.

1.2 Quality Metrics

1.2.1 Peak signal-to-noise ratio (PSNR)

The PSNR unit is decibels (dB) and it measures the ratio between the maximum power of a signal and the noise that corrupts it. As a quality metrics, it measures the ratio between the maximum power of the estimated image and the mean squared error (MSE) between the two images being compared. It is calculated as shown below,

$$PSNR = 10\log_{10}(\frac{MAX_x^2}{MSE}), \tag{2a}$$

where

$$MSE = \frac{1}{MM}\sum_{i=0}^{M-1}\sum_{j=0}^{M-1}[x(i,j) - \hat{x}(i,j)]^2, \tag{2b}$$

and MAX_x^2 is the maximum possible pixel value.

1.2.2 Structural Similarity Index Method

SSIM is an absolute measurement and its value is between -1 and 1. However, it typically lies in the range [0,1], being 1 if two identical images are compared, and 0 if two completely different images are compared. When calculating this value, the system separates the task into three comparisons: luminance, contrast and structure, as discussed in (Wang et al., 2004).

$$SSIM(x,\hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + c_1)(2\sigma_{x\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)}, \tag{3}$$

where μ_x and $\mu_{\hat{x}}$ are the average of and x and $\hat{x}$, respectively. Besides, σ_x and $\sigma_{\hat{x}}$ are the variances

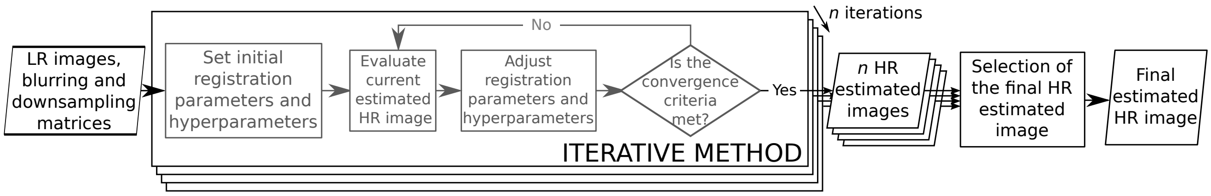


Figure 2: General operation of the algorithm

of and x and $\hat{x}$, respectively, $\sigma_{x\hat{x}}$ is the covariance of both signals and c_1 and c_2 are variables that stabilize the division with a weak denominator.

1.3 Bayesian Framework

The Bayesian framework we use (Villena et al., 2010), (Villena et al., 2013) is based on,

$$p(\mathbf{x}|\mathbf{y_k}) = \frac{p(\mathbf{y_k}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y_k})} \propto p(\mathbf{y_k}|\mathbf{x})p(\mathbf{x}), \qquad (4)$$

where $p(\mathbf{x})$ is the prior and $p(\mathbf{x}|\mathbf{y_k})$ is the posterior. The likelihood $p(\mathbf{y_k}|\mathbf{x})$ has gaussian PDF with mean equals to $\mathbf{AH_kC(s_k)x}$.

The prior may be defined as one of the equations below, depending on its type. For the SAR prior, the equation is,

$$p_1(\mathbf{x}\,|\alpha_1) \propto \alpha_1^{\frac{M^2}{2}} \exp -\frac{\alpha_1}{2}||C\,\mathbf{x}\,||^2 \qquad (5)$$

where C is the Laplacian operator, α_1 is the model hyperparameter and M^2 is the HR lexicographic vector dimension.

The ℓ_1-norm prior is describes as,

$$p_2(\mathbf{x}\,|\alpha_2)$$
$$\propto (\alpha_2^h\alpha_2^v)^{\frac{M^2}{4}} \exp - \sum_{i=1}^{M^2}[\alpha_2^h||\Delta_i^h(\mathbf{x})||_1 + \alpha_2^v||\Delta_i^v(\mathbf{x})||_1], \qquad (6)$$

where $\Delta_i^h(\mathbf{x})$ and $\Delta_i^v(\mathbf{x})$ are the horizontal and vertical first order differences, respectively, for the pixel i, and α_2^h and α_2^v are the model hyperparameters, each for one of the directions.

And the Total Variation (TV) prior is defined as,

$$p_3(\mathbf{x}\,|\alpha_3) \propto \alpha_3^{\frac{M^2}{2}} \exp -\frac{\alpha_3}{2}\sum_{i=1}^{M^2}\sqrt{[\Delta_i^h(\mathbf{x})^2 + \Delta_i^v(\mathbf{x})^2]}, \qquad (7)$$

where $\Delta_i^h(\mathbf{x})$ and $\Delta_i^v(\mathbf{x})$ are the horizontal and vertical differences, and α_3 is the model hyperparameter.

The HR estimated image is calculated as the result of an optimization problem that is summarized on Figure 2. The set of LR images $(\mathbf{y_k})$ are, of course, known by the method. So are the down sampling $\mathbf{A}$ and blurring matrices $\mathbf{H_k}$, the latter because we consider that the modeling of the camera is already established. These three arguments are the inputs of the algorithm. The main calculations are performed by the iterative part of the process: we set the initial registration parameters and hyperparameters, then we evaluate the current estimated HR image (as explained in detail in Villena et al. (2010)), adjust all parameters to be used in the next iteration and finally, check the convergence criteria. After the iterative method, n HR estimated images are available and, therefore, the next step is to select the final HR estimated image, which is the result of the algorithm.

1.4 Motivation and Objective

The algorithms used in this work are iterative, thus, they all have convergence conditions. Such conditions are based on the two images estimated in consecutive iterations. This work uses an SSIM-based convergence criteria and another one based on MSE. As a result of these iterative methods, we have more than one estimated image. In fact, we have one image for each iteration. Some works, such as Villena et al. (2013) choose the best one as the result of the SR algorithms. In other words, they choose the one with higher quality metrics value. However, both PSNR and MSE are full-reference quality metrics, that is, they compare the estimated image with the HR original one. Although this approach works well on simulated problems, it can not be applied on real ones, where the HR original image - the reference - isn't available.

Therefore, we have changed the parameter to choose the final result of the algorithm. Instead of electing the best image, the one computed in the last iteration is defined as the final estimated one. We investigate such results using SSIM and PSNR as quality metrics. Good results indicate that it is possible to use the algorithm on real situations.

2 Methods

The results presented in this paper were obtained using a modified version of the software developed according to the algorithms in Babacan et al. (2011) Villena et al. (2013), Nascimento and Salles (2015), and Villena et al. (2010). The original version is available at

<http://decsai.ugr.es/pi/superresolution> and it was released on July 7th, 2015.

Our experiments are synthetic, i.e. the LR images are generated from a given HR original image, in order to use the quality metrics PSNR and SSIM. The six images in Figure 3 were used as original HR images. We generate the LR images using the model in Equation 1, as the following parameters. The blurring mask is a mean filter, that is, h_k is a 3×3 matrix with all elements equal to 1/9. The displacement vectors are $\theta_l = [0°, 3°, -3°, 5°, -5°]$, $c_k = [0, 0, 0.5, 1, 0]$ and $d_k = [0, 0.5, 0, 0, 1]$. Five LR images were generated for each case, $L = 5$, and the down sampling scale is $P = 2$.

Different noises, SR methods and convergence conditions were tested. Regarding the noise, it is always gaussian with zero mean, but its variance changes with the LR images' SNR (signal-to-noise ratio), whose values vary among 1, 5, 10, 20 and 25 db. As for the SR methods, we tested TV, ℓ_1-norm, SAR, ℓ_1SAR and TVSAR. Finally, the convergence conditions are MSE-based and SSIM-based, respectively, $\text{MSE}(\mathbf{\hat{x}_{j-1}}, \mathbf{\hat{x}_j})/||\mathbf{\hat{x}_{j-1}}||^2 < 10^{-5}$ and $\text{SSIM}(\mathbf{\hat{x}_{j-1}}, \mathbf{\hat{x}_j}) > 0.9999$. For both inequations $\mathbf{\hat{x}_{j-1}}$ is the estimated image of the previous iteration and $\mathbf{\hat{x}_j}$ is the estimated image of the current iteration.

For each one of the simulations, we had chosen the image estimated at the last iteration. Regarding such estimated image, we registered its PSNR and SSIM values. Besides, the execution time and number of iterations of each simulation were also registered. Therefore, we compare the algorithm with both convergence conditions in terms of quality and speed.

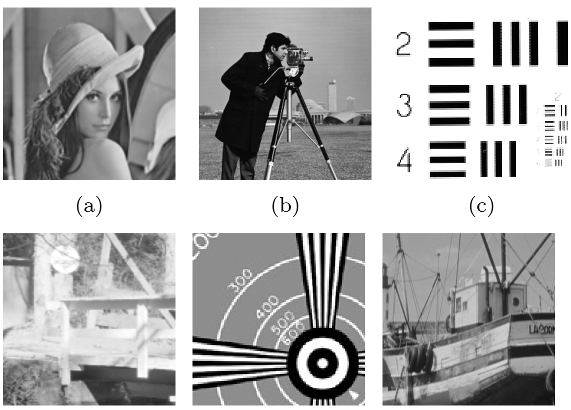


Figure 3: Original HR Images. (a) *lena*, 128×128 pixels; (b) *cameraman*, 256×256 pixels; (c) *card*, 120×120 pixels; (d) *cr5210*, 128×128 pixels; (e) *EIAcen120*, 120×120 and (f) *boats*, 256×256 pixels.

3 Experimental Results

Each type of experiment, according to the noise, SR methods and convergence condition, was executed six times: one for each HR image. The average SSIM values were calculated to show the differences between the five SR methods. We can remark that, for both convergence conditions, the results are similar. In Figures 4 (SSIM-based convergence) and 5 (MSE-based convergence) we can see that the presence of SAR increases the quality, especially for lower SNR values (10 db and below), i.e, when the noise is more intense. We can also notice that the combination of TV and SAR (85% TV + 15% SAR) provides bad results, in contrast to the results given by the ℓ_1SAR combination.

The average execution time is available at Table 1. Such values are separated per original HR image ,due to the strong relationship between execution time and dimension. The images *boats* and *cameraman* present the highest values of execution time and they both have 256×256 dimension. On the other hand, the images named *card* and *EIAcen120* are the smallest ones with 120×120 pixels and present, on average, the fastest results.

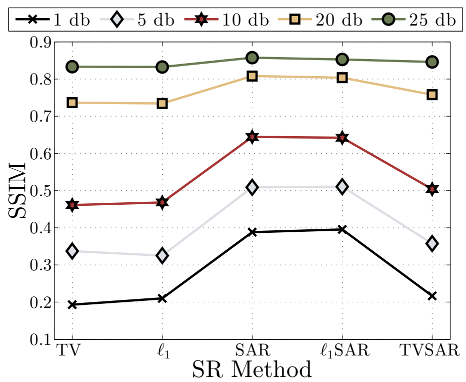


Figure 4: SSIM average values according to each SNR and SR methods, using SSIM-based convergence

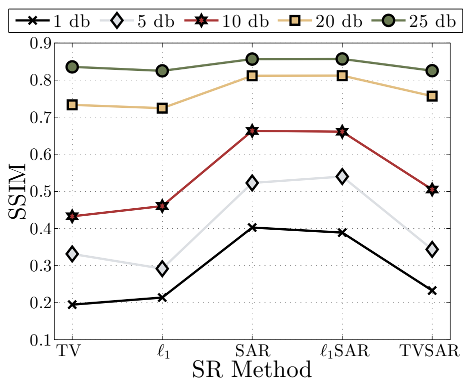
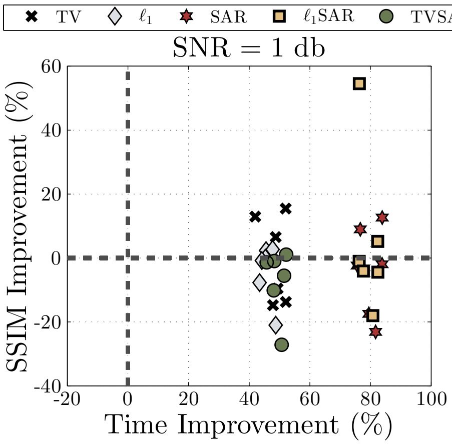
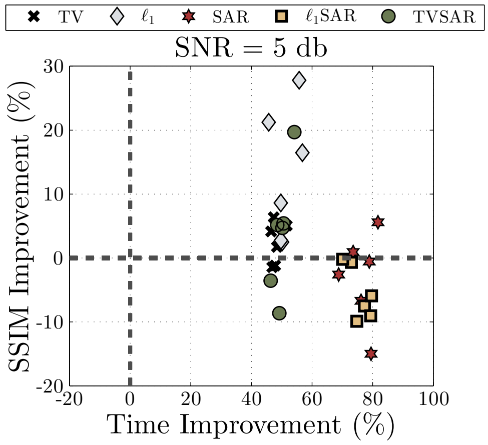


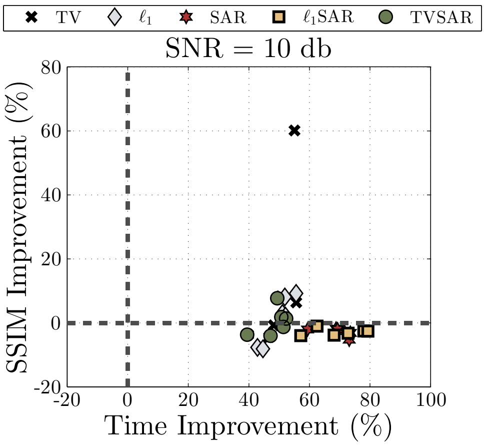
Figure 5: SSIM average values according to each SNR and SR methods, using MSE-based convergence



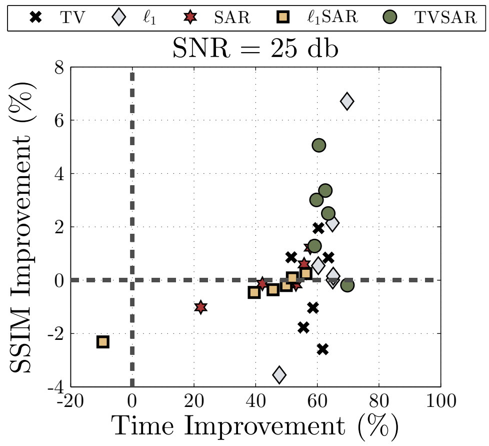
(a)



(b)



(c)



(d)

Figure 6: Comparison between SSIM and time improvement (both in percentage), when changing the convergence condition to SSIM, for each one of the signal-to-noise ratios. (a) SNR = 1 db, (b) 5 dB, (c) 10 db and (d) 25 db

Table 1: Average execution time (in seconds), per image and convergence condition

Image	$T_{\mathbf{MSE}}$	$T_{\mathbf{SSIM}}$
card	85.82	33.34
cr5210	118.10	37.09
lena128	104.46	36.86
boats	514.34	164.22
cameraman	525.65	169.72
EIAcen120	84.98	33.44

Figure 6 shows the comparison between SSIM and time improvement. The SSIM-based convergence, opposed to the MSE-based one, provides faster results in 99.33% of the cases. This means that, from 150 experiments we had run, only one of them was slower than the respective one with MSE-based convergence. Such loss of time is evaluated in 9.52% of the MSE execution time (3.73 seconds). The yellow square on the bottom-left quadrant in Figure 6(d) represents this experiment, which was executed with image *EIAcen120*. Moreover, the average time improvement, concerning all 150 experiments, is 57.19%.

In Figures 6(a) and 6(b) it is clear that changing the convergence to SSIM promotes a bigger time improvement when using SAR and ℓ_1SAR, in these cases, with more intense noise. The same behavior is noticed in Figure 6(c), but not so clearly.

The quality improvement - measured by SSIM - happens in 56.67% of the cases. The correlation between the two sets of data (SSIM value of MSE-based convergence and SSIM value of SSIM-based convergence) is 0.9913. This result is shown, graphically, in Figure 7.

4 Conclusions

We have investigated a blind method to choose the resulting estimated image from an iterative superresolution approach with SSIM-based convergence. Previous works choose the best estimated image according to full reference image quality assessment (FR-IQA) metrics, PSNR or SSIM, thus the original image is needed in such cases. Therefore, those algorithms are not suitable for real situ-

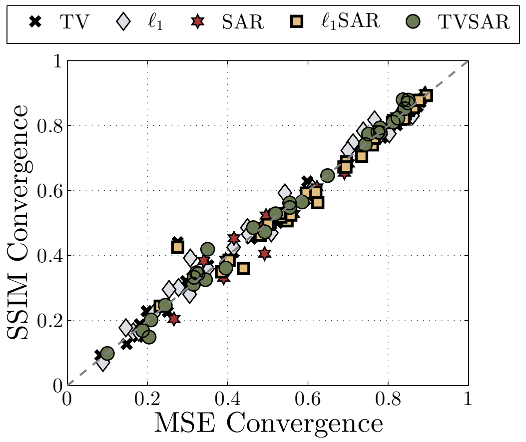


Figure 7: Correlation of the experiments with MSE-based convergence and the experiments with SSIM-based convergence

ations. Our algorithm is a modification of an iterative method that uses an SSIM-based convergence criterion and we compare it to a modification of a method that uses MSE-based convergence criterion. On both cases we modify the method of choice to our blind proposition, which is, choosing the estimated image from the last iteration.

SSIM as FR-IQA metrics is used here only to compare the results in a objective manner. Besides convergence, we have made some comparisons regarding priors too. ℓ_1SAR and SAR provide better results, with higher SSIM values. They also provide the biggest time improvement, when having more intense noise. The experiments we have carried out show that our method is more effective in cases with intense noises (10 db and bellow) and in the presence of SAR.

The SSIM-based convergence method presents faster results in 99.33% of the cases, with an average gain of 57.19% in execution time. The convergence conditions are highly correlated, in terms of quality. That is, even if quality is not improved, the SSIM-based convergence method is worth using due to the improvement in execution time. Moreover, our method is not only faster as it is directly applicable to real cases.

We finally conclude that the SSIM convergence condition is a good choice, when choosing the last image estimated on the last iteration as the output of the Superresolution algorithm. Such condition provides faster execution with satisfactory results.

Acknowledgment

Thais Pedruzzi do Nascimento would like to acknowledge CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior for the scholarship granted.

References

Babacan, S. D., Molina, R. and Katsaggelos, A. K. (2011). Variational Bayesian Super Resolution., *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society* **20**(4): 984–999.

Irani, M. and Peleg, S. (1990). Super resolution from image sequences, *Pattern Recognition, 1990. Proceedings., 10th International Conference on* **2**: 115–120.

Katsaggelos, A. K., Molina, R. and Mateos, J. (2007). *Super Resolution of Images and Video*, Vol. 3.

Nascimento, T. P. and Salles, E. O. T. (2015). Using SSIM as Convergence Criteria in a Variational Superresolution Bayesian Approach, *XI Workshop de Visão Computacional*, pp. 44–49.

Villena, S., Vega, M., Babacan, S. D., Molina, R. and Katsaggelos, A. K. (2010). Using the Kullback-Leibler Divergence to Combine Image Priors in Super-Resolution Image Reconstruction, *Proceedings - International Conference on Image Processing, ICIP* **2**(4): 893–896.

Villena, S., Vega, M., Babacan, S. D., Molina, R. and Katsaggelos, A. K. (2013). Bayesian combination of sparse and non-sparse priors in image super resolution, *Digital Signal Processing: A Review Journal* **23**(2): 530–541.

Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity, *IEEE Transactions on Image Processing* **13**(4): 600–612.