Patch-based Statistical Performance Analysis of Upsampling for Precise Super–Resolution

Djamila Aouada, Kassem Al Ismaeil and Björn Ottersten
Interdisciplinary Centre for Security, Reliability, and Trust, University of Luxembourg, Luxembourg, Luxembourg

Keywords: Super-resolution, Affine Bias Model, UP-SR, MSE, Depth Camera, Patch-based.

Abstract: All existent methods for the statistical analysis of super–resolution approaches have stopped at the variance term, not accounting for the bias in the mean square error. In this paper we give an original derivation of the bias term. We propose to use a patch-based method inspired by the work of (Chatterjee and Milanfar, 2009). Our approach, however, is completely new as we derive a new affine bias model dedicated for the multi-frame super resolution framework. We apply the proposed statistical performance analysis to the *Upsampling for Precise Super–Resolution* (UP-SR) algorithm. This algorithm was shown experimentally to be a good solution for enhancing the resolution of depth sequences in both cases of global and local motions. Its performance is herein analyzed theoretically in terms of its approximated mean square error, using the proposed derivation of the bias. This analysis is validated experimentally on simulated static and dynamic depth sequences with a known ground truth. This provides an insightful understanding of the effects of noise variance, number of observed low resolution frames, and super–resolution factor on the final and intermediate performance of UP–SR. Our conclusion is that increasing the number of frames should improve the performance while the error is increased due to local motions, and to the upsampling which is part of UP-SR.

1 INTRODUCTION

Multi–frame super–resolution (SR) is an inverse image reconstruction problem. It consists in estimating a high resolution (HR) reference image from multiple observed low resolution (LR) frames (Milanfar, 2010), where the ratio between HR and LR is known as the SR factor. Depth sensors of limited resolutions, such as the 3D MLI by IEE S.A. of resolution (56×64) (3d MLI, 2014) and the PMD camboard nano of resolution (120×160) (pmd CamBoard nano, 2014), are good examples of current technologies that could benefit from the multi-frame SR framework.

There have been some attempts to derive the asymptotic limits of SR (Rajagopalan and Kiran, 2003; Robinson and Milanfar, 2006). Those, however, do not consider the bias of an SR estimator despite it being always part of an image reconstruction solution (Chatterjee and Milanfar, 2009). Moreover, they assume a Gaussian noise model while UP-SR exploits an additive Laplace noise model.

Recently, Al Ismaeil et al. (K. Al Ismaeil, 2013a) proposed a new multi-frame SR approach for the enhancement of static depth scenes captured with these cameras. In (K. Al Ismaeil, 2013b), the authors have extended this work to dynamic depth scenes subject to local motions, i.e., scenes containing one or more moving objects. This algorithm is referred to as *Upsampling for Precise Super-Resolution* (UP-SR). It is based on upsampling the observed LR frames prior to their registration. This has led to rewriting the general SR data model to a simplified image denoising problem from multiple noisy and blurred observations. The denoising is then achieved using a Maximum Likelihood (ML) approach. In both (K. Al Ismaeil, 2013a) and (K. Al Ismaeil, 2013b) the performance of UP-SR was characterized experimentally.

In this paper, in order to reach a better understanding of this algorithm, and to separate the effect of the number of frames and the effect of the SR factor, we derive its performance in terms of mean square error (MSE) at a given noise level. The MSE is composed of a variance and a bias term. We propose to adapt to the considered problem the affine bias model of (Chatterjee and Milanfar, 2009) based on a representation with patches, leading to an approximation of the UP-SR bias. This bias is related to the error due to gradient-based motion estimation (Robinson and Milanfar, 2003), and to the SR factor used in UP-SR as the upsampling factor. Few assumptions are

Aouada D., Al Ismaeil K. and Ottersten B..
Patch-based Statistical Performance Analysis of Upsampling for Precise Super–Resolution.
DOI: 10.5220/0005316001860193
In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (VISAPP-2015), pages 186-193
ISBN: 978-989-758-089-5

introduced for simplicity of analysis but are shown to still hold experimentally, both quantitatively and qualitatively. We give the variance of the UP-SR estimator considering an additive Laplacian noise model as it has been shown to better fit the SR problem as compared to a Gaussian noise model (S. Farsiu, 2003; S. Farsiu, 2004).

The remainder of the paper is organized as follows: Section 2 reviews the UP-SR estimation. An approximation of the corresponding MSE is derived in Section 3. Quantitative and qualitative results confirming the theoretical performance analysis are given in Section 4. The conclusion is given in Section 5.

2 UPSAMPLING FOR PRECISE SUPER RESOLUTION (UP-SR)

The dynamic multi–frame SR problem considers a sequence of N observed LR column images $\{\mathbf{y}_t, t = 1, \cdots, N\}$ of size m. The objective is to reconstruct the corresponding HR sequence $\{\mathbf{x}_t, t = 1, \cdots, N\}$ containing images of size n such that $n = r \times m$, with r being the SR factor. The dynamic SR problem may be simplified by reconstructing one HR image at a time using the full observed sequence. To that end, we fix the reference time to t_0, and focus on the reconstruction of $\mathbf{x}_{t_0}$ using the $N' = (N - t_0 + 1)$ preceding measurements. The operation may be repeated for $t_0 = 1, \cdots, N$. A noisy LR observation is modeled as follows:

$$\mathbf{y}_t = \mathbf{DHM}_{t_0}^t \mathbf{x}_{t_0} + \mathbf{n}_t,\ t_0 \leq t \text{ and } t, t_0 \in [1, N] \subset \mathbb{N}^*, \tag{1}$$

where $\mathbf{D}$ is a known constant downsampling matrix of dimension $(m \times n)$. The system blur is represented by the time and space invariant matrix $\mathbf{H}$. The $(n \times n)$ matrices $\mathbf{M}_{t_0}^t$ correspond to the motion between $\mathbf{x}_{t_0}$ and $\mathbf{y}_t$ before their downsampling. Without loss of generality, both $\mathbf{H}$ and $\mathbf{M}_{t_0}^t$ are assumed to be block circulant matrices. The additive noise vector $\mathbf{n}_t$ at time t follows a white multivariate Laplace distribution (S. Farsiu, 2003) defined as:

$$p(\mathbf{n}_t) = \prod_{i=1}^{m} \frac{\sqrt{2}}{2\sigma} \exp\left(-\frac{\sqrt{2}|\mathbf{n}_t(i)|}{\sigma}\right), \tag{2}$$

where $\frac{\sigma}{\sqrt{2}}$ is a positive Laplace scale factor leading to the diagonal covariance matrix $\Sigma = \sigma^2 \mathbf{I}_m$, with $\mathbf{I}_m$ being the identity matrix of size $(m \times m)$.

The UP-SR algorithm starts be upsampling the observed LR images. This leads to a more accurate and robust motion estimation which enhances the registration of frames. Moreover, it allows to directly solve the problem of undefined pixels in the SR initialization phase (K. Al Ismaeil, 2013b). We define the resulting r-times upsampled image as $\mathbf{y}_t \uparrow = \mathbf{U} \cdot \mathbf{y}_t$, where $\mathbf{U}$ is an $(n \times m)$ upsampling matrix. Due to the specifications of depth data, classical interpolation–based methods (e.g., bicubic) cannot be used as they lead to jagged values and to blurring effects especially for boundary pixels. Thus, the upsampling $\mathbf{U}$ has to be dense, which is also known as nearest neighbor upsampling.

Two consecutive frames are better registered if the motion between them is estimated from their upsampled versions $\mathbf{y}_{t-1} \uparrow$ and $\mathbf{y}_t \uparrow$, by finding

$$\hat{\mathbf{M}}_{t-1}^t = \arg\min_{\mathbf{M}} \Psi\left(\mathbf{y}_{t-1} \uparrow, \mathbf{y}_t \uparrow, \mathbf{M}\right), \tag{3}$$

where Ψ is a dense optical flow-related cost function and

$$\mathbf{y}_t \uparrow = \mathbf{M}_{t-1}^t \mathbf{y}_{t-1} \uparrow + \mathbf{v}_t. \tag{4}$$

The vector $\mathbf{v}_t$ contains the innovation that we assume negligible in this framework. In addition, similarly to (Elad and Feuer, 1999), for analytical convenience, we assume that all pixels in $\mathbf{y}_t \uparrow$ originate from pixels in $\mathbf{y}_{t-1} \uparrow$ in a one to one mapping. Therefore, each row in $\mathbf{M}_{t-1}^t$ contains 1 for each position corresponding to the address of the source pixel in $\mathbf{y}_{t-1} \uparrow$. This bijective property implies that the matrix $\hat{\mathbf{M}}_{t-1}^t$ is an invertible permutation, s.t., $[\hat{\mathbf{M}}_{t-1}^t]^{-1} = \hat{\mathbf{M}}_t^{t-1}$. Furthermore, its estimate leads to the following registration to $\mathbf{y}_{t-1}$:

$$\bar{\mathbf{y}}_t \uparrow = \hat{\mathbf{M}}_t^{t-1} \mathbf{y}_t \uparrow. \tag{5}$$

Using a cumulative motion compensation approach, the registration of a non-consecutive frame $\mathbf{y}_t \uparrow$ to the reference $\mathbf{y}_{t_0} \uparrow$ is achieved as follows:

$$\bar{\mathbf{y}}_t^{t_0} \uparrow = \hat{\mathbf{M}}_t^{t_0} \mathbf{y}_t \uparrow = \underbrace{\hat{\mathbf{M}}_{t_0+1}^{t_0} \cdots \hat{\mathbf{M}}_t^{t-1}}_{(t-t_0) \text{ times}} \cdot \mathbf{y}_t \uparrow. \tag{6}$$

Choosing the upsampling matrix $\mathbf{U}$ to be the transpose of $\mathbf{D}$, the product $\mathbf{UD} = \mathbf{A}$, gives a block circulant matrix $\mathbf{A}$ that defines a new blurring matrix $\mathbf{B} = \mathbf{AH}$. Considering that $\mathbf{B}$ and $\mathbf{M}_{t_0}^t$ are block circulant matrices, we have $\mathbf{BM}_t^{t_0} = \mathbf{M}_t^{t_0}\mathbf{B}$. As a result, the estimation of $\mathbf{x}_{t_0}$ may be decomposed into two steps; estimation of a blurred HR image $\mathbf{z}_{t_0} = \mathbf{B}\mathbf{x}_{t_0}$, followed by a deblurring step. The data model in (1) becomes

$$\bar{\mathbf{y}}_t^{t_0} \uparrow = \mathbf{z}_{t_0} + \nu_t,\ t_0 \leq t \text{ and } t, t_0 \in [1, N] \subset \mathbb{N}^*, \tag{7}$$

where $\nu_t = \hat{\mathbf{M}}_t^{t_0} \mathbf{U} \cdot \mathbf{n}_t$ is an additive noise vector of length n. The permutation $\hat{\mathbf{M}}_t^{t_0}$ only reorders the elements of $\mathbf{n}_t$ while $\mathbf{U}$ leads to replicating each element r times. This results in a new $(n \times n)$ covariance matrix with a non-diagonal structure $\tilde{\Sigma} = \hat{\mathbf{M}}_t^{t_0} \mathbf{U} \Sigma \mathbf{D} \hat{\mathbf{M}}_{t_0}^t$. For simplicity of analysis, we will however assume an independent and identically distributed (i.i.d.) Laplace random vector with $\tilde{\Sigma} = \sigma^2 \mathbf{I}_n$. The error due to this simplification is a blurring effect that should be largely reduced in the deblurring step. The log-likelihood function associated with (7) becomes

$$\begin{aligned} \ln p(\overline{\mathbf{y}}_{t_0}^{t_0}\uparrow, \cdots, \overline{\mathbf{y}}_N^{t_0}\uparrow \mid \mathbf{z}_{t_0}) &= \\ &= \ln\left(\prod_{t=t_0}^{N} \frac{\sqrt{2}}{2\sigma} \exp\left(-\frac{\sqrt{2}\|\overline{\mathbf{y}}_t^{t_0}\uparrow - \mathbf{z}_{t_0}\|_1}{\sigma}\right)\right) \\ &= -N' \ln\frac{\sigma}{\sqrt{2}} - \frac{\sqrt{2}}{\sigma}\sum_{t=t_0}^{N} \|\mathbf{z}_{t_0} - \overline{\mathbf{y}}_t^{t_0}\uparrow\|_1, \end{aligned} \tag{8}$$

where $\|\cdot\|_1$ is the L_1–norm. Maximizing (8) with respect to $\mathbf{z}_{t_0}$, we obtain

$$\hat{\mathbf{z}}_{t_0} = \arg\min_{\mathbf{z}_{t_0}} \sum_{t=t_0}^{N} \|\mathbf{z}_{t_0} - \overline{\mathbf{y}}_t^{t_0}\uparrow\|_1, \tag{9}$$

which corresponds to the pixel-wise temporal median estimator, i.e., $\hat{\mathbf{z}}_{t_0} = \mathrm{med}_t\{\overline{\mathbf{y}}_t^{t_0}\uparrow\}_{t=t_0}^{N}$.
Then, as a second step, follows an image deblurring to recover $\hat{\mathbf{x}}_{t_0}$ from $\hat{\mathbf{z}}_{t_0}$. Considering a regularization term $\Gamma(\mathbf{x})$ added to compensate undetermined cases by enforcing prior information about $\mathbf{x}_{t_0}$, we finally find

$$\hat{\mathbf{x}}_{t_0} = \underset{\mathbf{x}}{\mathrm{argmin}}\Big(\|\mathbf{B}\mathbf{x} - \hat{\mathbf{z}}_{t_0}\|_1 + \lambda\Gamma(\mathbf{x})\Big), \tag{10}$$

where λ is the regularization parameter.

3 STATISTICAL PERFORMANCE ANALYSIS

Considering the data model in (7), we herein look into the performance of the median estimator $\hat{\mathbf{z}}_{t_0}$ in terms of MSE with respect to the SR factor r and the number of frames N'. The MSE may be decomposed into two parts, the variance var$(\cdot)$ and the bias denoted as bias$(\cdot)$. Given a known ground truth $\mathbf{x}_{t_0}$, we have

$$\mathrm{MSE}(\hat{\mathbf{z}}_{t_0}, \mathbf{x}_0) = \mathrm{var}(\hat{\mathbf{z}}_{t_0}) + \|\mathrm{bias}(\hat{\mathbf{z}}_{t_0})\|^2. \tag{11}$$

Below, we detail the computation of each term.

3.1 Bias Computation

The SR problem has been reformulated as a denoising problem in (7). The affine bias model of Chatterjee and Milanfar (Chatterjee and Milanfar, 2009) for image denoising may therefore be applied after modifications to fit the estimation in (9). This model is local where processing is done on patches. We start by decomposing the ground truth image $\mathbf{x}_{t_0}$ into n patches $\{\mathbf{q}_{t_0}(i), i = 1, \cdots, n\}$. Each patch $\mathbf{q}_{t_0}(i)$ is centered at the pixel $\mathbf{x}_{t_0}(i)$ and is chosen to be of the size of the upsampling factor r. Similarly, $\overline{\mathbf{y}}_t^{t_0}\uparrow$ are decomposed into n overlapping patches $\{\mathbf{p}_t(i), i = 1, \cdots, n\}$. The data model (7) can be rewritten for patches as:

$$\mathbf{p}_t(i) = \mathbf{q}_{t_0}(i) + \eta_t(i), \tag{12}$$

where $\eta_t(i)$ is the patch measurement error due to noise and to blur. Relating patches from frames at different times leads to rewriting (4) but between any two frames at t and t' as:

$$\mathbf{p}_{t'}(i) = \mathbf{W}_t^{t'}(i)\mathbf{p}_t(i) + \mathbf{w}_t^{t'}(i), \tag{13}$$

where $\mathbf{W}_t^{t'}(i)$ is a sub-block of $\hat{\mathbf{M}}_t^{t'}$ centered at position i, and $\mathbf{w}_{t'}^{t}(i)$ is a local innovation directly related to cumulated innovations neglected in (4). The estimation in (9) corresponds to locally selecting the element $\mathbf{p}_{t'}(i)$ with the highest ranking among the N' patches $\{\mathbf{p}_t(i), t = t_0, \cdots, N\}$ as the estimate $\hat{\mathbf{q}}_{t_0}(i)$. Thus, by combining (12) and (13), we may write

$$\hat{\mathbf{q}}_{t_0}(i) = \mathbf{W}_{t_0}^{t'}(i)\left(\mathbf{q}_{t_0}(i) + \eta_{t_0}(i)\right) + \mathbf{w}_{t_0}^{t'}(i) \tag{14}$$

Therefore, given the expectation operator $\mathbb{E}(\cdot)$ and $\mathbf{I}_r$ the identity matrix of size $(r \times r)$, the local bias per patch can be calculated as:

$$\begin{aligned} &\mathrm{bias}(\hat{\mathbf{q}}_{t_0}(i)) = \mathbb{E}(\hat{\mathbf{q}}_{t_0}(i)) - \mathbf{q}_{t_0}(i) \\ &= \left(\mathbb{E}\left(\mathbf{W}_{t_0}^{t'}(i)\right) - \mathbf{I}_r\right)\mathbf{q}_{t_0}(i) + \mathbb{E}\left(\mathbf{W}_{t_0}^{t'}(i)\eta_{t_0}(i) + \mathbf{w}_{t_0}^{t'}(i)\right) \\ &= \mathbf{S}_i\mathbf{q}_{t_0}(i) + \mathbf{u}_i. \end{aligned} \tag{15}$$

The result in (15) is a local affine model inspired by, but different from, the model in (Chatterjee and Milanfar, 2009). The final bias is

$$\|\mathrm{bias}(\hat{\mathbf{z}}_{t_0})\|^2 = \sum_{i=1}^{n} \|\mathrm{bias}(\hat{\mathbf{q}}_{t_0}(i))\|^2. \tag{16}$$

It is interesting to note that for the simple case where the average motion per patch as well as its innovation $\mathbf{w}_{t_0}^{t'}(i)$ are close to zero, the expected value of the matrix of local motion is close to the identity matrix, i.e., $\mathbb{E}\left(\mathbf{W}_{t_0}^{t'}(i)\right) \approx \mathbf{I}_r$, and $\mathbf{S}_i$ becomes a zero matrix. The per–patch bias term becomes $\mathbb{E}(\eta_t(i))$ which represents the combined effect of blur and noise per patch. The statistical properties of the noise part are the same as those of ν_t, i.e., of zero mean. The blur part is due to the $(r-1)$ pixels per patch that resulted from dense upsampling. Assuming that they induce a fixed mean error ε, the total bias may be simplified as follows:

$$\|\mathrm{bias}(\hat{\mathbf{z}}_{t_0})\|^2 = \sum_{i=1}^{n} \|\mathbb{E}(\eta_t(i))\|^2 = n \cdot (r-1)\varepsilon^2. \tag{17}$$

Note that in (17), for $r = 1$, there is no blur due to upsampling, and the UP-SR estimation becomes unbiased. In the general case, however, the bias term is data dependent because of $\mathbf{q}_{t_0}(i)$ in (15). It also depends of the SR factor r, and the statistics of the local motions and noise. We note that the bias is proportional to the squared SR factor r^2 and to the image size n. These results are also data dependent as expressed by the pixel values $\mathbf{p}_k(i)$ and the structural decomposition of an image to patches. As can be seen next, the variance term is proportional to the noise variance σ^2 and the number of measurements N'.

3.2 Variance Computation

Assuming an i.i.d. n–multivariate Laplace distribution, we may write: $\mathrm{var}(\hat{\mathbf{z}}_{t_0}) = \mathrm{tr}(\mathrm{cov}(\hat{\mathbf{z}}_{t_0})) = n \cdot \mathrm{var}(\hat{\mathbf{z}}_{t_0}(i))$, where $\mathrm{tr}(\cdot)$ and $\mathrm{cov}(\cdot)$ are the trace and covariance functions, respectively. Therefore, using the result of (Beaulieu and Jiang, 2010), we find

$$\mathrm{var}(\hat{\mathbf{z}}_{t_0}(i)) = 2\sigma^2 f(N'), \quad i = 1, \cdots, n, \qquad (18)$$

where for N' even,

$$f(N') = \frac{4N'!}{\left(\left(\frac{N'-1}{2}\right)!\right)^2}\left(\frac{1}{2}\right)^{\frac{N'+1}{2}} \sum_{k=0}^{\frac{N'-1}{2}} \frac{\binom{\frac{N'-1}{2}}{k}\left(-\frac{1}{2}\right)^k}{(N'+1+2k)^3}, \qquad (19)$$

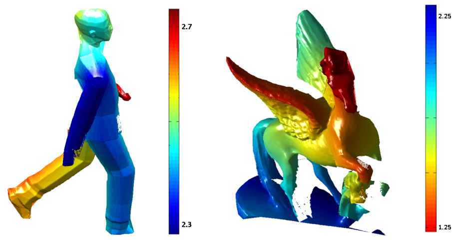
and for N' odd,

$$f(N') = \frac{N'!}{\left(\frac{N'}{2}\right)!\left(\frac{N'}{2}-1\right)!}\left(\frac{1}{2}\right)^{\frac{N'}{2}}\left(\frac{1}{N'^3}\left(\frac{1}{2}\right)^{\frac{N'}{2}} + \sum_{k=0}^{\frac{N'}{2}-1}\binom{\frac{N'-1}{2}}{k}\left(-\frac{1}{2}\right)^k \frac{7N'^2+8N'(k+1)+4(k+1)^2}{N'^2(N'+2k+2)^3}\right). \qquad (20)$$

We note that in addition to assuming that the noise is i.i.d., we also assume that the effect of overlapping patches is expressed in the bias term. Thus, the variance is independent of r, which means that it is the same for a simple denoising operation where no SR is involved and $r = 1$.

4 EXPERIMENTAL VALIDATION

In order to illustrate the statistical analysis of the UP-SR algorithm with quantitative evaluation, we set up the following experiment. We use the publicly available toolbox V-REP (V-REP, 2014) to create synthetic data with fully known ground truth for both dynamic and static scenes, Figure 1 (a), and Figure 1 (b), respectively. Three depth cameras with the same field of view are fixed at the same position. These cameras are of different resolutions, namely, (512×512), (256×256), and (128×128) pixels. They are used to capture three sequences for each subject. These sequences are further degraded with additive Laplacian noise with a standard deviation σ varying from 0 mm to 60 mm. Each sequence is super-resolved using UP-SR by considering 9 successive frames.



(a) (b)

Figure 1: Ground truth data used for the statistical performance analysis. (a) Dynamic scene with a moving person, (b) Static scene.

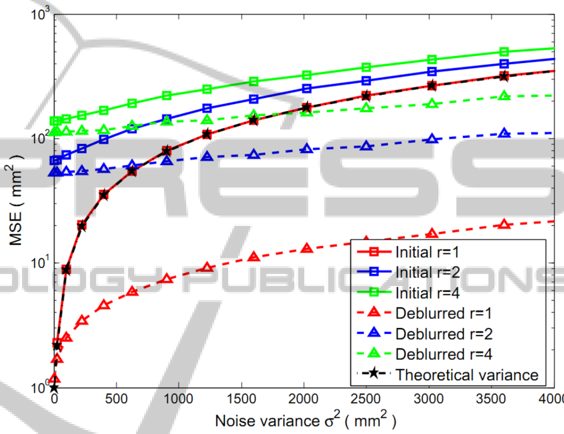


Figure 2: MSE of UP–SR versus noise variance for the static scene in Figure 1 (b).

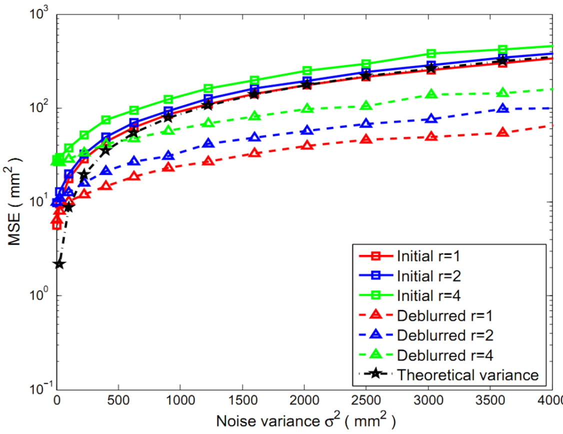


Figure 3: MSE of UP–SR versus noise variance for the dynamic scene in Figure 1 (a).

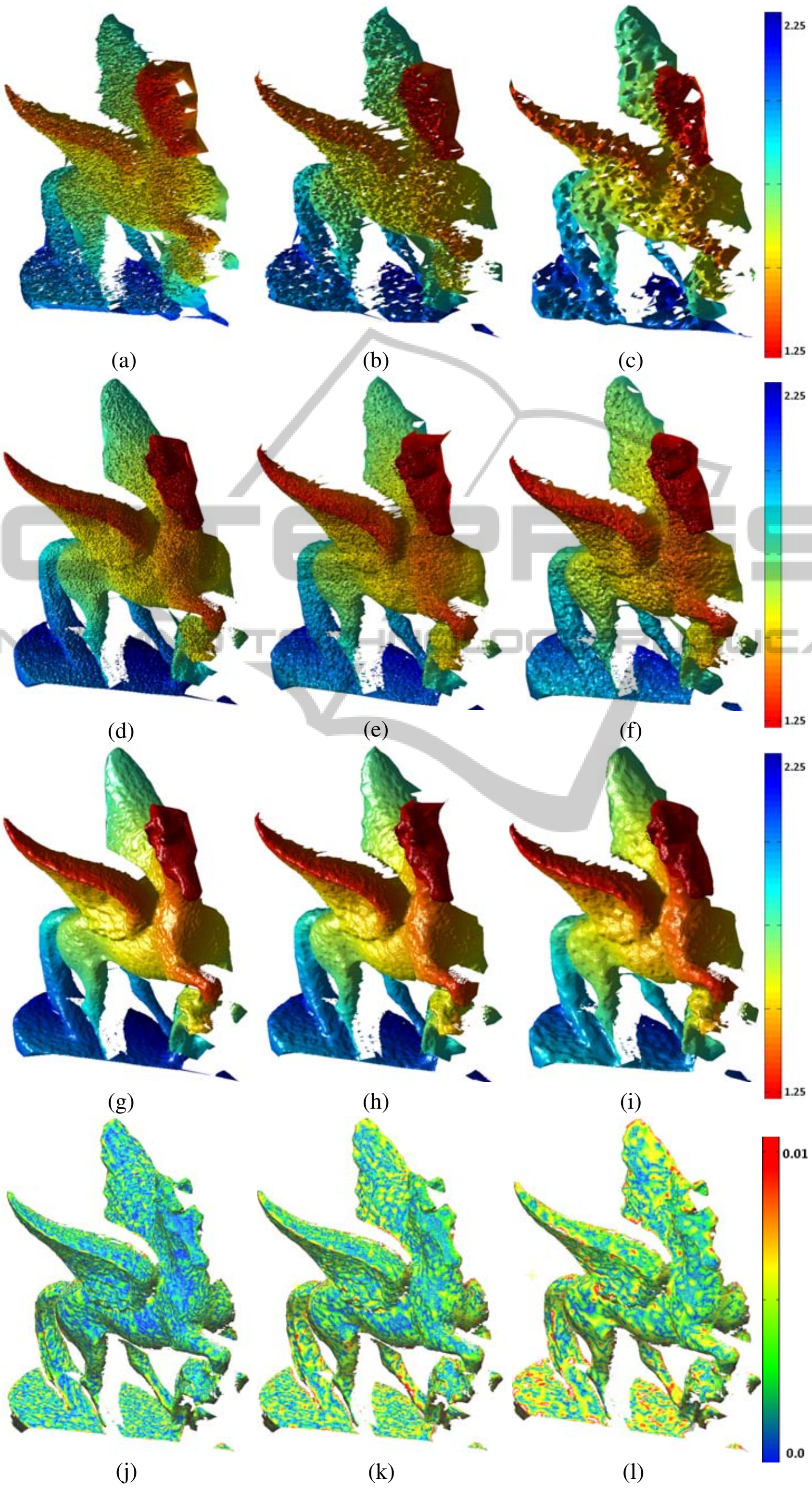


Figure 4: Statistical performance analysis of UP-SR for static depth scenes. First, second and third columns correspond respectively to $r = 1$, $r = 2$, and $r = 4$ where (a), (b) and (c) are the noisy LR observations; (d), (e), and (f) are the result of the Initial of UP-SR; (g), (h), and (i) are the result of deblurring step of UP-SR. The corresponding error maps as compared with the ground truth Figure 1. (b) are given in (j), (k), and (l).

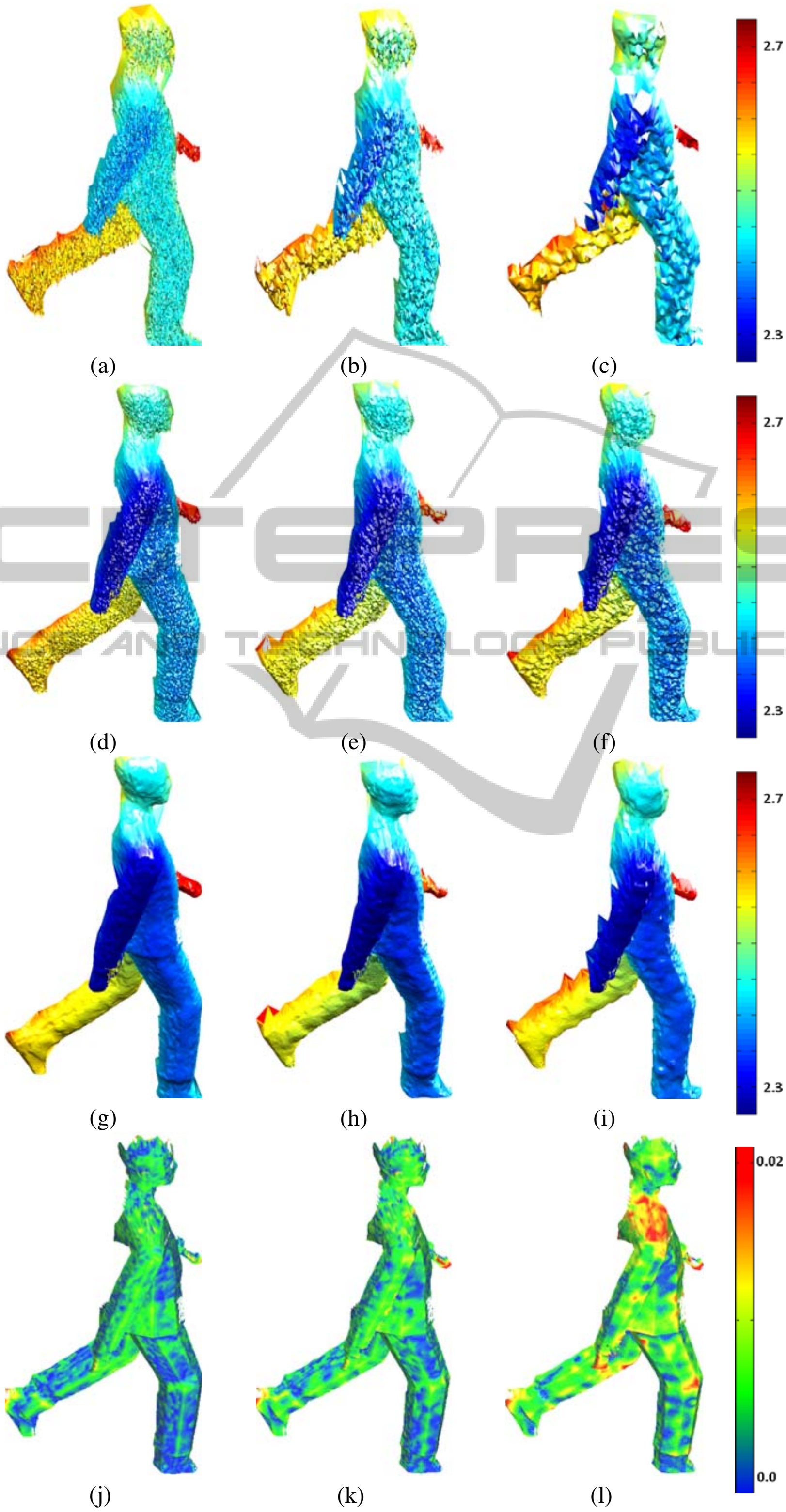


Figure 5: Statistical performance analysis of UP-SR for dynamic depth scenes. First, second and third columns correspond respectively to $r = 1$, $r = 2$, and $r = 4$ where (a), (b) and (c) are the noisy LR observations; (d), (e), and (f) are the result of the initialization step of UP-SR; (g), (h), and (i) are the result of the deblurring step of UP-SR. The corresponding error maps as compared with the ground truth Figure 1. (a) are given in (j), (k), and (l).

Starting with the static case, the corresponding MSE performance of the initialization step and the second deblurring step of UP-SR are reported in Figure 2 in solid and dashed lines, respectively. In the simple case where $r = 1$, the SR resolution problem is merely a denoising one where the ground truth is estimated from 9 noisy measurements. In other words, the objective is not to increase resolution, and hence there is no blur due to upsampling. Indeed, as seen in Figure 2, the solid red line overlaps with the dashed-dotted black line which corresponds to the theoretical variance for the odd case obtained using (20). A non-zero bias is found for $r = 2$ and $r = 4$ where the corresponding blue and green solid lines are above the theoretical variance. This suggests a correlation between motion and upsampling blur as expressed by the vector $\mathbf{u}_i$ in (15). We note an increased bias for a larger SR factor r. This is justified by a larger blur effect due to the dense upsampling and to motion. Finally, the dashed lines in Figure 2 confirm the performance enhancement after applying the optimization in (10); thus, ensuring an effective deblurring. We used an exhaustive search to find the best parameters for Γ. These quantitative results can be appreciated visually in Figure 4 where the noise level is fixed at $\sigma = 30$ *mm*. The effective resolution enhancement, with a SR factor of $r = 4$, and denoising power of UP-SR for a static depth scene is seen in 3D in Figure 4 (i). The average RMSE in 3D is shown in Figure 4 (l).

In the dynamic case, a similar behaviour has been observed with some differences related to the local motion estimation and data type. We can see that even for the simple case with $r = 1$, a non-zero bias from the theoretical variance is found for both the initial and optimized results, represented by the solid and dashed red lines in Figure 3, respectively. This bias is mainly due to the error caused by self-occlusions and errors in registration. In the case of resolution enhancement with SR factors $r = 2$ and $r = 4$, we can see that the non-zero bias in Figure 3 follows the same behaviour as the one for the static case. The difference is a smaller shift from the theoretical variance, especially for low noise levels as can be seen in the corresponding blue and green solid lines. This is directly related to the data type. Whereas in the dynamic case we used a simple CAD object (Figure 1 (a)), in the static case we used a scanned object containing more geometric details (Figure 1 (b)). Therefore, the downsampling process has more effect on the static object and leads to a larger loss in details, hence a larger bias.

5 CONCLUSION

We have proposed to adapt the affine bias model proposed by (Chatterjee and Milanfar, 2009) to approximate the bias of a depth multi–frame super–resolution algorithm using a patch–based representation. Specically, the *Upsampling for Precise Super–Resolution* (UP–SR) algorithm has been considered. With an additional step to handle the effect of downsampling, this derived statistical analysis may be applied to any multi–frame super–resolution algorithm. The application to UP–SR has the advantage that it does not need to handle downsampling separately because it directly transfers the super–resolution problem to a denoising one. We provided a theoretical performance analysis of UP–SR in terms of mean square error, including the variance and the bias terms. We validated these results experimentally using a synthetic simulation setup. This analysis gave insights on the effect of the different parameters: noise level, the number of observed low resolution frames, and the super–resolution factor. In summary, the performance of UP–SR or any multi-frame super–resolution algorithm increases with the increase of the number of observations. In the case of dynamic scenes, this performance decreases due to local motions and errors of registration. In the case of UP–SR, there is an additional error due to the upsampling effect. It can be reduced thanks to the final deblurring phase.

ACKNOWLEDGEMENT

This work was supported by the National Research Fund (FNR), Luxembourg, under the CORE project C11/BM/1204105/FAVE.

REFERENCES

3d MLI (2014). http://www.iee.lu/technologies.

Beaulieu, N. and Jiang, S. (2010). Ml estimation of signal amplitude in laplace noise,. In *Global Telecommunications Conference*. IEEE.

Chatterjee, P. and Milanfar, P. (2009). Bias modeling for image denoising,. In *43rd Asilomar Conference on Signals, Systems and Computers*.

Elad, M. and Feuer, A. (1999). Super-resolution reconstruction of continuous image sequences,. In *Transaction Pattern Analysis and Machine Intelligence*. IEEE.

K. Al Ismaeil, D. Aouada, B. M. B. O. (2013a). Depth super-resolution by enhanced shift & add,. In *15th International Conference on Computer Analysis of Images and Patterns*. Springer.

K. Al Ismaeil, D. Aouada, B. M. B. O. (2013b). Dynamic super-resolution of depth sequences with non-rigid motions,. In *20th IEEE International Conference on Image Processing*. IEEE.

Milanfar, P. (2010). Super-resolution imaging,. In *CRC Press*.

pmd CamBoard nano (2014). http://www.pmdtec.com/.

Rajagopalan, A. and Kiran, P. (2003). Motion-free super-resolution and the role of relative blur,. In *J. Opt. Soc. Amer.*

Robinson, D. and Milanfar, P. (2003). Bias-minimizing filters for gradient-based motion estimation,. In *37th Asilomar Conference on Signals, Systems and Computers*. IEEE.

Robinson, D. and Milanfar, P. (2006). Statistical performance analysis of super-resolution,. In *Transaction Image Processing*. IEEE.

S. Farsiu, D. Robinson, M. E. P. M. (2003). Robust shift and add approach to super–resolution,. In *International symposium on Optical Science and Technology, SPIE's 48th Annual Meeting*.

S. Farsiu, D. Robinson, M. E. P. M. (2004). Fast and robust multi-frame super-resolution,. In *Transaction Image Processing*. IEEE.

V-REP (2014). http://www.k-team.com/mobile-robotics-products/v-rep.