MULTI-FRAME SUPER-RESOLUTION BY ENHANCED SHIFT & ADD

*Kassem Al Ismaeil**   *Djamila Aouada**   *Bruno Mirbach*[†]   *Björn Ottersten**

[*]SnT - Universtity of Luxembourg
{kassem.alismaeil, djamila.aouada, bjorn.ottersten}@uni.lu

[†]Advanced Engineering - IEE S.A.
bruno.mirbach@iee.lu

ABSTRACT

A critical step in multi-frame super-resolution is the registration of frames based on their motion. We improve the performance of current state-of-the-art super-resolution techniques by proposing a more robust and accurate registration as early as in the initialization stage of the high resolution estimate. Indeed, we solve the limitations on scale and motion inherent to the classical *Shift & Add* approach by upsampling the low resolution frames up to the super-resolution factor prior to estimating motion or to median filtering. This is followed by an appropriate selective optimization, leading to an enhanced *Shift & Add*. Quantitative and qualitative evaluations have been conducted at two levels; the initial estimation and the final optimized super-resolution. Results show that the proposed algorithm outperforms existing state-of-art methods.

Index Terms— Super-resolution, pyramidal optical flow, motion estimation, upsampling.

I. INTRODUCTION

Super-resolution (SR) is a common technique used to recover a high resolution (HR) reference image from a set of observed low resolution (LR) images subject to errors due the optical acquisition system such as noise and blurring, and to deviations from the reference image due to relative motion. The past two decades have witnessed a several contributions on SR for static scenes [1], [2], [3], [4], [5], [6], [7]. Most of the proposed methods are dedicated to a simple translational or affine motion. As presented in [8], these algorithms are numerically limited to small global motions even for an increased number of LR frames. Most SR techniques start with constructing the initial HR grid with sub-pixel accuracy by combining the LR frames by interpolation. These methods work effectively when a sufficient number[1] of LR images contain slightly different perspectives of the scene. It is critical to start with an initial HR image that is as accurate as possible. The initial image may be obtained by an operation commonly referred to as *Shift & Add* (*S&A*) [7] which includes a filling operation based on the motion of the considered LR images. Another method is by using some interpolation aligning the LR measurements on an HR grid and interpolating the missing points, the most successful method is the variational Bayesian SR (*VBSR*) [11]. Once

an initial HR image is designed, it is refined with an optimization process by minimizing a given cost function to finally reach the desired HR image. The main drawback of these methods is that the quality of the initial HR image is restricted to a specific range of motions related to the SR factor. Indeed, a weak motion diversity among the LR frames leads to undefined pixels in the initial HR image resulting in artifacts in the final solution and a strong deterioration of the SR performance. As a solution to this, example-based SR algorithms have been proposed [9], and their combinations with classical multi-frame SR [10]. Such algorithms rely on a heavy learning phase, and assume that images carry some redundancies.

In this paper, we propose to release the limitations on scale and the number of required frames of classical SR algorithms without prior assumptions on the data and without engaging in an additional learning stage. Our method is based on an accurate registration of frames to the reference frame resulting in an enhanced *S&A* algorithm. Our strategy consists in using the efficient pyramidal optical flow estimation starting from LR frames upsampled up to the SR factor. This is followed by a pixel-wise median operation which guarantees that no undefined pixels appear in the initial HR image and it is further refined by a selective optimization.

The paper is organized as follows: Section II presents the SR initialization step as performed by the *S&A* approach and explains the source of its limitations that consequently impact the full SR algorithm. We explain the improvements due to estimating motion from upsampled images in Section III and give our proposed algorithm in Section IV. Section V presents a thorough evaluation of our method and its comparison with state-of-the-art approaches. We give our summary and conclusion in Section VI.

II. MOTIVATION AND BACKGROUND

The aim of SR algorithms is to estimate an HR image $\mathbf{X}$ of size $(m \times n)$ from N observed LR images $\mathbf{Y}_k$, $k = 0, ..., (N-1)$, where each LR image is of size $(\acute{m} \times \acute{n})$ pixels, with $n = r \cdot \acute{n}$ and $m = r \cdot \acute{m}$, such that r is the SR factor. Every image $\mathbf{Y}_k$ may be viewed as a LR noisy and deformed realization of $\mathbf{X}$ caused by the imaging system at the kth acquisition. Considering their respective lexicographic vector forms $\mathbf{y}_k$ and $\mathbf{x}$, the data model is defined as follows:

$$\mathbf{y}_k = \mathbf{DHW}_k\mathbf{x} + \mathbf{n}_k, \qquad k = 0, ..., (N-1), \quad (1)$$

---

[1]Note that this number is bounded as proven in [8].

where $\mathbf{W}_k$ is an $(mn \times mn)$ matrix corresponding to the geometric motion between $\mathbf{x}$ and $\mathbf{y}_k$. In this framework, this motion is assumed to be global translational; hence, $\mathbf{W}_k$ represents a global shifting operator by u_k in x direction, and by v_k in y direction. The point spread function of the camera (PSF) is modeled by the $(mn \times mn)$ blurring matrix $\mathbf{H}$. The matrix $\mathbf{D}$ of dimension $(\acute{m}\acute{n} \times mn)$ represents the downsampling operator, and the vector $\mathbf{n}_k$ is the additive noise at k.

In most methods, the estimation of $\mathbf{x}$ follows a maximum likelihood approach that, by considering $\mathbf{n}_k$ as a Laplacian white noise, leads to the following minimization:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\mathrm{argmin}}\Big(\|\mathbf{H}\mathbf{x} - \mathbf{x}_0\|_1 + \lambda\Gamma(\mathbf{x}) \Big), \qquad (2)$$

where $\mathbf{x}_0$ is an initial blurred guess for the original HR image $\mathbf{x}$. $\Gamma(\mathbf{x})$ is a regularization term added to compensate undetermined cases by enforcing prior information about $\mathbf{x}$ and λ is the regularization parameter.

The classical *S&A* approach [7] defines $\mathbf{x}_0$ by first setting its corresponding full HR image grid $\mathbf{X}_0$ to zeros as denoted below:

$$\mathbf{X}_0 = \mathbf{0}_{m \times n}. \qquad (3)$$

Then, all LR images $\mathbf{Y}_k$ are used to update the pixel values in $\mathbf{X}_0$. To that end, given a reference LR image $\mathbf{Y}_0$ chosen as the closest one to the target HR image $\mathbf{X}$, the global translational motions $\mathbf{w}_k = (u_k, v_k)$ between each image $\mathbf{Y}_k$ and $\mathbf{Y}_0$ are computed for $k = 1, \cdots, (N-1)$. These motions are used to register all LR images $\mathbf{Y}_k$ with respect to the reference image $\mathbf{Y}_0$. The resulting registered images $\overline{\mathbf{Y}}_k$ are simply defined at each pixel position $\mathbf{p} = (x, y)$ as follows:

$$\overline{\mathbf{Y}}_k(\mathbf{p}) = \mathbf{Y}_k(\mathbf{p} + \mathbf{w}_k). \qquad (4)$$

These images are then grouped into M sets based on their relative motions $\mathbf{w}_k$. Note that to avoid aliasing problems, the range of this motion is forced to be within the SR factor r by a simple modulo function, i.e., $u_k = u_k\mathrm{mod}(r)$ and $v_k = v_k\mathrm{mod}(r)$. The frames in one set are fused by median filtering resulting in one LR image $\overline{\overline{\mathbf{Y}}}_i$ per motion $\mathbf{w}_i$, with $1 \leq i \leq M \leq N$. Each frame is then used to update the pixels of $\mathbf{X}_0$ as follows:

$$\mathbf{X}_0 (r \cdot \mathbf{p} + \mathbf{w}_i) = \overline{\overline{\mathbf{Y}}}_i(\mathbf{p}). \qquad (5)$$

This operation is known as zero filling in the *S&A* approach. We note that for a successful filling, there should be enough motion diversity in the considered LR frames. Indeed, in order to further update the zero pixels in $\mathbf{X}_0$, an additional $(r \times r)$ median filtering is applied. Given that the median filter's breakdown point is $\frac{1}{2}$, a meaningful filling that does not leave pixels undefined should be achieved if the following condition is satisfied:

$$\mathrm{round}\left(\frac{r^2}{2}\right) \leq M. \qquad (6)$$

We show the effect of undefined pixels in classical *S&A* in Fig. 1(b). A similar phenomenon is observed using interpolation-based initialization such as *VBSR* as seen in
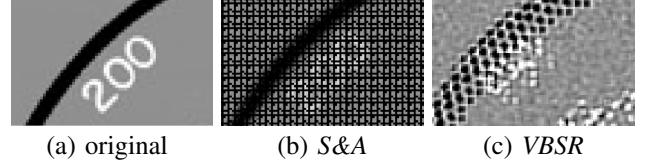

(a) original   (b) *S&A*   (c) *VBSR*

Fig. 1. Undefined pixels using state-of-the-art SR methods.

Fig. 1(c), suggesting that interpolation is not a sufficient solution to remove undefined pixels. Therefore, starting with an accurate initial guess $\mathbf{x}_0$ has a strong impact on the final solution of (2). It is common to face this serious problem of undefined pixels in practice. It is dealt with by restricting the SR factor to low values, e.g., $r = 2$, and by taking a relatively large number of frames, e.g., $N > 30$, thus indirectly attempting to satisfy the inequality (6), which, in turn, limits the practical usage of SR algorithms. In what follows, our aim is to increase motion diversity M to give more freedom in the choice of r without having to increase N. We propose to tackle the aforementioned problem by a new non-zero initialization of $\mathbf{X}_0$ along with a robust and accurate sub-pixel motion estimation as detailed in Section III.

III. PYRAMIDAL MOTION ESTIMATION FOR SR

Estimating the motions $\mathbf{w}_k$ with high sub-pixel accuracy is crucial in capturing the full diversity in motion as contained in the observed frames; hence, important in increasing M. Indeed, for two frames $\mathbf{Y}_i$ and $\mathbf{Y}_j$ with respective relative motions $\mathbf{w}_i$ and $\mathbf{w}_j$, such that $\|\mathbf{w}_i - \mathbf{w}_j\|_2 = \epsilon$; if the motion estimation approach has an accuracy that is smaller than ϵ, the two frames will be wrongly fused and labeled under the same motion.

Classical *S&A* uses pyramidal motion estimation (*PyrME*) ([17], [18]). This method represents state-of-art in motion estimation increasing both accuracy and robustness. In the SR context, we propose an enhancement of *PyrME* in order to serve our objective of increasing M.

Below, we start by describing *PyrME* as it is currently used in SR, specifically in *S&A*. We then present how we further improve its performance for SR.

Typically, motion between a frame $\mathbf{Y}_k$ and the reference frame $\mathbf{Y}_0$ at a given target point $\mathbf{p}$ is estimated by minimizing the following error:

$$\epsilon(\mathbf{w}_k) = \sum_{\mathbf{q}=\mathbf{p}-\boldsymbol{\mu}}^{\mathbf{p}+\boldsymbol{\mu}} \|\mathbf{Y}_0(\mathbf{q}) - \mathbf{Y}_k(\mathbf{q} + \mathbf{w}_k)\|_2^2. \qquad (7)$$

This dissimilarity function is computed on the neighborhood of radius $\|\boldsymbol{\mu}\|$ around the target pixel $\mathbf{p}$. The vector $\boldsymbol{\mu}$ defines the integration disc which corresponds to the maximum motion that can be detected within this framework. As a result, the estimation of $\mathbf{w}_k$ is characterized by two conflicting properties: accuracy and robustness. The first property is related to sub-pixel motion accuracy which is achieved by taking a small $\|\boldsymbol{\mu}\|$. The second property is related to robustness to large motions, for which a large $\|\boldsymbol{\mu}\|$

is preferable. *PyrME* was proposed as a trade-off solution for these conflicting characteristics. The principle is to follow a coarse to fine strategy that progressively downsamples the images $\mathbf{Y}_k$ and $\mathbf{Y}_0$ by a factor 2^l where l indicates the pyramidal level, $l = 0, \cdots, L$. This dyadic downsampling between consecutive levels may be defined as follows:

$$\mathbf{Y}_k^l(\mathbf{p}) = \mathbf{Y}_k^{l-1}(2\mathbf{p}) \quad s.t. \quad \mathbf{Y}_k^0 = \mathbf{Y}_k, \quad \forall k. \quad (8)$$

Practically, the size of the image at the highest level of the pyramid should not be less than (4×4) pixels. For the sake of simplicity, let us consider that the LR frames $\mathbf{Y}_k^0$ are square matrices ($\acute{m} = \acute{n}$). We may then set the maximal number of pyramidal levels as follows:

$$\frac{\acute{n}}{2^L} = 4 \quad \Rightarrow \quad L = \log_2(\acute{n}) - 2. \quad (9)$$

The motion estimation starts between the highest levels, i.e., the lowest resolutions, before progressively going back down to the initial level $l = 0$. At a given level l, the motion vector $\mathbf{w}_k^l$ is decomposed into an initial guess $\boldsymbol{\omega}_k^l$ and a residual motion vector $\boldsymbol{\phi}_k^l$. The initial guess $\boldsymbol{\omega}_k^l$ is obtained from the preceding level $(l+1)$ such that $\boldsymbol{\omega}_k^l = 2 \cdot \mathbf{w}_k^{(l+1)}$, and initially set to $\mathbf{0}_{1 \times 2}$ at $l = L$. This vector is used to pre-register the two images $\mathbf{Y}_k^l$ and $\mathbf{Y}_0^l$. Therefore, finding the optimal vector $\mathbf{w}_k^l$ is reduced to finding the optimal residual motion vector. The estimation of this smaller vector is, similarly to (7), defined by the following minimization:

$$\boldsymbol{\phi}_k^l = \underset{\boldsymbol{\nu}}{\mathrm{argmin}} \sum_{\mathbf{q}=\mathbf{p}-\boldsymbol{\mu}}^{\mathbf{p}+\boldsymbol{\mu}} \|\mathbf{Y}_0^l(\mathbf{q}) - \mathbf{Y}_k^l(\mathbf{q} + \boldsymbol{\omega}_k^l + \boldsymbol{\nu})\|_2^2. \quad (10)$$

The optimal motion vector at level l is then found by:

$$\mathbf{w}_k^l = \boldsymbol{\omega}_k^l + \boldsymbol{\phi}_k^l. \quad (11)$$

The result above is obtained with sub-pixel accuracy thanks to the residual refinement operation defined in (10) which is ensured by imposing a relatively small neighborhood disc radius $\|\boldsymbol{\mu}\|$. This operation is repeated until the finest motion vector is obtained at $l = 0$ defining $\mathbf{w}_k$ as $\mathbf{w}_k = \boldsymbol{\omega}_k^0 + \boldsymbol{\phi}_k^0$. It may also be expressed using the residuals at all levels as follows:

$$\mathbf{w}_k = \sum_{l=0}^{L} 2^l \boldsymbol{\phi}_k^l. \quad (12)$$

The refinement in (10) implies that the maximal pixel motion that can be estimated at each level of the pyramid is within a disc of amplitude $\|\boldsymbol{\mu}\|$. Given (12), the maximal overall pixel motion that can be estimated by *PyrME* is:

$$\boldsymbol{\mu}_{\max} = \mathcal{G}(L)\,\boldsymbol{\mu} \quad \text{with} \quad \mathcal{G}(L) = 2^{(L+1)} - 1. \quad (13)$$

The gain $\mathcal{G}(L)$ is a function of the height of the pyramid L. It follows that for more levels, *PyrME* may estimate large motions up to $\boldsymbol{\mu}_{\max}$ and hence verifies the robustness and accuracy properties simultaneously.

In the case of SR, we note that the target resolution at which we want to land is the HR, $(n \times n)$ in the square case. This gives us a natural way to further improve the performance

of *PyrME*. We thus propose to start by upsampling the LR frames up to the SR factor r prior to any motion estimation. By doing so, we increase the size of the basis of the pyramid by a factor r. Changing the starting point in *PyrME* leads to an increased pyramid height $L \uparrow^r$ by $\log_2(r)$ which results in a new gain $\mathcal{G}(L \uparrow^r)$:

$$\mathcal{G}(L \uparrow^r) = r \cdot \mathcal{G}(L) + (r - 1). \quad (14)$$

This result is important as we have shown that, in the SR context, the performance of *PyrME*, in terms of accuracy and robustness, may further be enhanced with a new gain $\mathcal{G}(L \uparrow^r)$.

IV. PROPOSED ALGORITHM

With the result of Section III at hand, we go back to define an *enhanced S&A* (*eS&A*) by taking the following steps:

Step 1: Non-zero Initialization

In order to estimate $\mathbf{w}_k$ more accurately, we now work with $\mathbf{Y}_k \uparrow^r$, $k = 0, \cdots, (N - 1)$, the N LR frames upsampled by the factor r. Performing the registration process as in (4) on the upsampled images $\mathbf{Y}_k \uparrow^r$ gives:

$$\overline{\mathbf{Y}}_k \uparrow^r (\mathbf{p}) = \mathbf{Y}_k \uparrow^r (\mathbf{p} + \mathbf{w}_k). \quad (15)$$

Registering upsampled frames as in (15) guarantees a better result with a higher accuracy than registering the LR images followed by upsampling them. This is due to the fact that registration parameters are approximated by rounding the motion vectors $\mathbf{w}_k$ with an expected error of $\pm \frac{1}{2}$ pixel. The effect of this error is related to the size of the registered images. By prior upsampling, this rounding error is reduced by a factor r; from $\pm \frac{1}{2\acute{n}}$ in the LR case to $\pm \frac{1}{2\acute{n}}$ in (15). The effect of this improvement is illustrated in Fig. 2. The registered frames are then fused by median filtering in order to initialize the HR grid $\mathbf{X}_0$ such that:

$$\mathbf{X}_0 = \mathrm{med}_k \left\{ \overline{\mathbf{Y}}_k \uparrow^r \right\}_{k=0}^{(N-1)}. \quad (16)$$

The proposed non-zero initialization in (16) releases the condition in (6), thus solving the problem of undefined
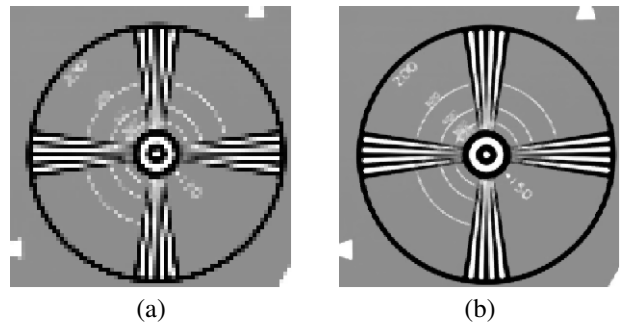
Fig. 2. Registration error. (a) Registered LR frames followed by upsampling process. (b) Registered upsampled frames using up-scaled LR motion vectors.

pixels. In order not to fall under the same artifacts as those present with interpolation-based SR approaches, e.g., *VBSR* (Fig. 1(c)), it is necessary to perform the filling operation from registered and clustered LR images as in (5). Indeed, the values from LR frames remain more reliable sources of information than the ones due to upsampling. We further point out that the higher accuracy in the estimation of $\mathbf{w}_k$ leads to a higher discrimination between motions, resulting in a higher diversity M and a better update of the pixel values in $\mathbf{X}_0$ as compared to the case of classical *S&A*. We note, that in our algorithm, it is more accurate to refer to this operation as initialization update rather than filling.

Step 2: Selective Optimization

We have now obtained a more accurate initial guess $\mathbf{X}_0$ that may be refined by optimizing its lexicographic form $\mathbf{x}_0$ using (2). In our work, we adopt the robust bilateral total variation Γ_{BTV} as a regularization term as defined in [7], and solve the following optimization:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}}\Big(\|\mathbf{\Lambda}\mathbf{H}\mathbf{x} - \mathbf{x}_0\|_1 + \lambda\Gamma_{BTV}(\mathbf{x})\Big). \qquad (17)$$

The matrix $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal corresponds to the elements of a weighting matrix $\mathbf{A}$ of the same size as the HR image $\mathbf{X}$. The matrix $\mathbf{A}$ assigns a weight to each pixel in $\mathbf{X}$ such that its contribution in the optimization (17) is proportional to the number of measurements used in initializing its value in $\mathbf{X}_0$ during Step 1. For an efficient optimization, we propose to further use $\mathbf{A}$ as a mask that, similarly to [16], only selects a subset of pixels to contribute in the optimization. In our framework, this selection actually improves the estimation by only accounting for reliable pixels, those that were updated, and setting to zero the effect of the pixels that were not updated during the initialization of $\mathbf{X}_0$. We illustrate the effect of this new selective optimization in Section V.

V. EXPERIMENTAL RESULTS

We compare the performance of the proposed enhanced *S&A* with the two state-of-the-art methods, that are currently, to the best of our knowledge, the best performing SR algorithms, namely, *S&A* [7], and *VBSR* [11]. We tested these methods using the softwares provided in [14] and [15].

Starting with the HR image *EIA* [12], we generated LR images by downsampling with a factor $r = 4$, and simulating a (3×3) Gaussian PSF with a standard deviation $\sigma = 0.4$, and further degrading by additive white Gaussian noise (AWGN). We evaluate the robustness of the proposed algorithm against two parameters: number of considered LR images N, and image contamination with noise using the signal to noise ratio (SNR). We measure the quality of the estimated HR image using peak signal to noise ratio (PSNR) defined as: $\text{PSNR} = 10\log_{10}\frac{m\times n}{\|\mathbf{x}-\hat{\mathbf{x}}\|_2}$.

Fig. 3(a) shows the average PSNR for 100 different noise realizations, and N progressively increasing from 4 to 20. In order to evaluate for relatively large motions, translation parameters are generated randomly between 0 and 9. Note

(a)

(b)
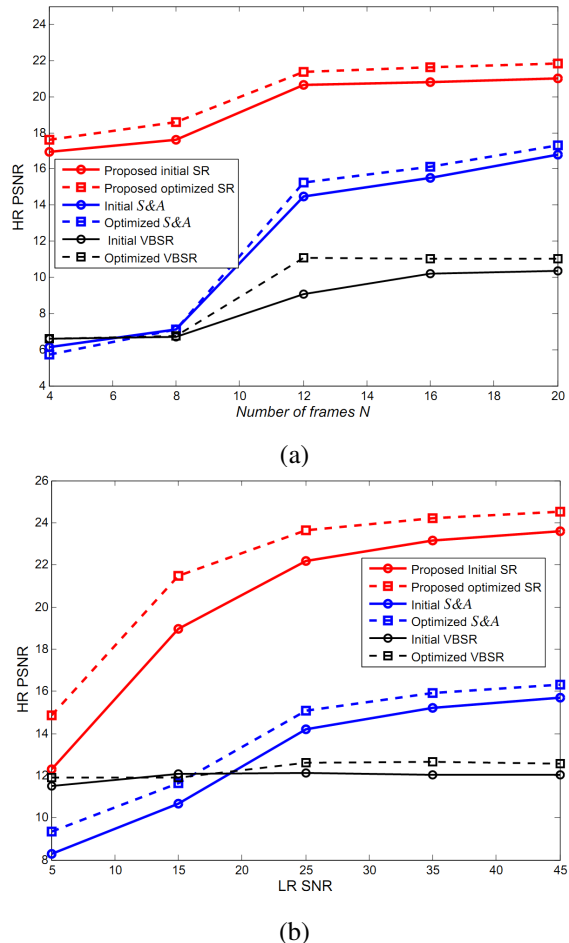
Fig. 3. PSNR for different SR methods applied on (75×75) LR frames with $r = 4$ (a) for increasing N, (b) for increasing SNR levels.

that in [7], smaller motions have been used which explains the difference with the result obtained in this paper using the same *S&A* algorithm. In the overdetermined case where $M \geq r^2$, and for small motions, both methods *eS&A* and *S&A* give comparable results.

We first provide the results of the non-zero initialization step using the proposed *eS&A* (solid lines). Then we give the final results obtained after applying the selective optimization (dashed lines) whose starting point is the output of the previous step. To avoid any increase in computational cost, upsampled frames can be registered using an approximation by upscaling corresponding LR motion vectors. Please note that for a fair comparison, we use the same set of parameters in the optimization step for both the proposed method and the method in [7]. From Fig. 3(a), it is clear that the proposed method provides significant improvements as compared to existing methods for any choice of N, even as small as 4 images. This observation holds for both the initial estimation and for the iteratively optimized solution. Note that the initial estimate considerably outperforms *VBSR* and *S&A*, initial and optimized solutions.

Fig. 4 illustrates an example of a HR estimated image using 8, 12, and 20 LR images. Due to the condition (6), it is not surprising to see the artifacts caused by the undefined pixels, where the number of images is not sufficient to cover the motion range. Moreover, it is clear that the proposed method provides the best visually enhanced HR images as seen in Fig. 4 (c), (f), and (i) with sharper edges compared to other methods.

Next, we conducted a second round of experiments to evaluate the performance of the proposed *eS&A* at different noise levels. We used the same 12 frames generated previously and further degraded them by AWGN with SNR of 5, 15, 25, 35, and 45 dB.

One may note that for a fair comparison we use this number of frames as it guarantees an initial HR image without undefined pixels for all methods (see Fig. 4(d), (e), and (f)). Mean PSNR values of 100 different noise realizations are plotted in Fig. 3(b) showing that the proposed method provides the best results among discussed SR methods across all noise levels. It is important to note that the optimization of initial estimates ensures a final result that is consistently more robust to noise.

Similar results and conclusions were obtained using real data as we used 20 (57×49) LR images of the *disk* dataset [12]. Fig. 5 presents the SR results of *disk* images with SR factor $r = 6$. It is visually clear that the proposed method provides better results with sharper edges and less ringing artifacts than other methods in addition to solving the undefined pixels problem.

Finally, in order to illustrate the effect of the proposed selective optimization as compared to the optimization proposed in [7], we ran an experiment on three initial HR images by performing 50 iterations. Initial images are obtained by applying *eS&A* with $r = 5$ on three different sequences. Each sequence consists of 11 (96×96) LR frames further degraded by AWGN with SNR = 25dB. As shown in Fig. 6, we may see that the proposed optimization method results in an increase in $PSNR$ gain as compared with [7]. Moreover, the number of processed pixels decreases and varies from an image to another depending on the number of selected pixels (e.g., 82944 pixels (36 %) and 73728 pixels (32%) processed pixels per iteration for the *cameraman* and *Lena* images, respectively). In contrast, the objective function in [7] processes all pixels with a minimum weight of value 1 for unreliable pixels (e.g., for a (480×480) image, the number of processed pixels are 230400 pixels per iteration).

VI. CONCLUSION

We presented a new enhanced *S&A* algorithm which improves the quality of the initialization of the HR image in the context of the SR problem. The proposed algorithm is based on upsampling LR images before registering them. We demonstrated that this new approach for SR provides a more accurate motion estimation and registration. Experimental results with both synthetic and real images demonstrate that the proposed algorithm gives results superior to existing state-of-the-art methods such as classical *S&A* and *VBSR*
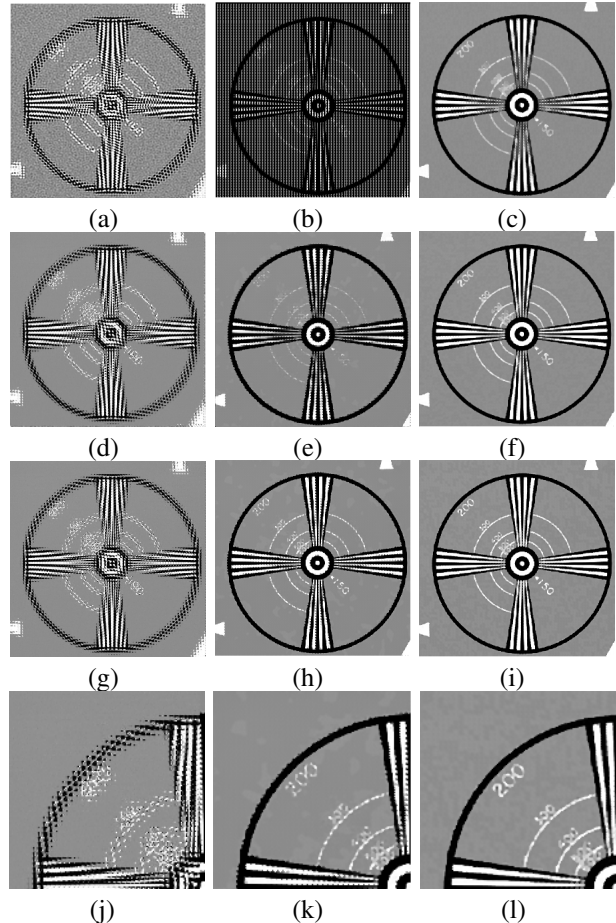
Fig. 4. Results of different SR methods applied to a (75×75) LR sequence of a static scene with $r = 4$ and different frame numbers. *VBSR* for (a) $N = 8$, (d) $N = 12$, (g) $N = 20$, and by *S&A* for (b) $N = 8$, (e) $N = 12$, (h) $N = 20$ and by proposed *eS&A* for (c) $N = 8$, (f) $N = 12$, (i) $N = 20$.

under various conditions; low number of input LR images, and different noise levels. In addition to being robust, the proposed approach showed that it can be reliably used as an initial guess for SR algorithms. Further optimizing this initialization ensures a strong resilience to noise without additional computational cost.
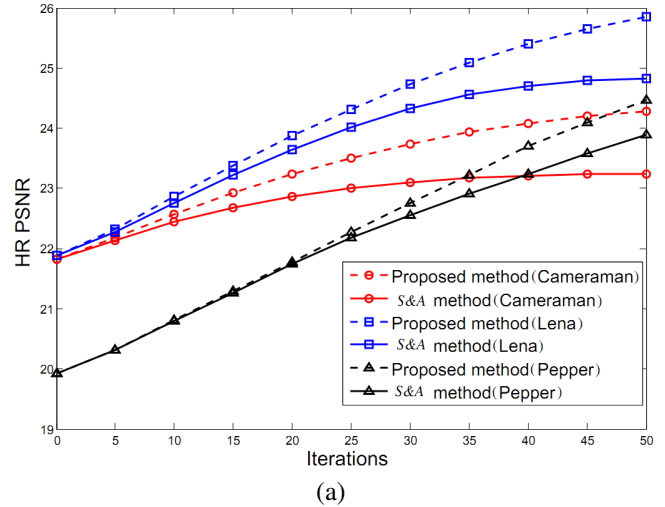
VII. REFERENCES

[1] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration", in Advances in Computer Vision and Image Processing, vol. 1, chapter 7, pp. 317-339, JAI Press, Greenwich, Conn, USA, 1984

[2] N. K. Bose, H. C. Kim, H. M. Valenzuela, "Recursive implementation of total least squares algorithm for

(a)   (b)

(c)   (d)
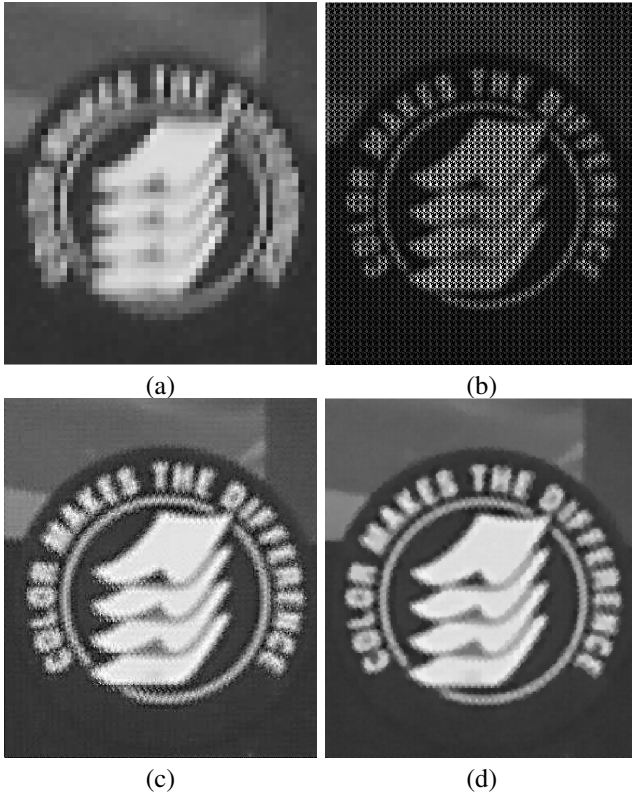
Fig. 5. Results of different SR methods with $r = 6$ applied on real data of a static scene, (a) *S&A*, (b) *VBSR*, and (c) proposed method.

(a)

Fig. 6. PSNR gain per iteration of the proposed selective optimization against *S&A* [7] using 11 (96×96) LR images with SNR = 25 and SR scale factor $r = 5$.

image reconstruction from noisy, undersampled multi-frames", ICASSP-93., vol.5, no., pp.269,272 vol.5, 27-30 April 1993

[3] M. Elad, Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," IEEE Transactions on Image Processing, vol.10, no.8, pp.1187,1193, Aug 2001.

[4] M. Chiang and T. E. Boult, "Efficient super-resolution via image warping", Image and Vision Computing, vol. 18 no. 10, pp. 761-771, 2000.

[5] M. Irani and S. Peleg, "Improving resolution by image registration", Graphical models and image processing, 1991, vol. 53, pp. 231-239.

[6] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust Super-Resolution", CVPR, 2001, vol. 1, pp. 645-650.

[7] S. Farsiu, M. D. Robinson, M. Elad, P. Milanfar, "Fast and robust multiframe super resolution," Image Processing, IEEE Transactions on , vol.13, no.10, pp.1327,1344, Oct. 2004

[8] L. Zhouchen ; S. Heung-Yeung , "Fundamental limits of reconstruction-based superresolution algorithms under local translation," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.26, no.1, pp.83,97, Jan. 2004

[9] O. M. Aodha, N. Campbell, A. Nair, G. Brostow.

"Patch Based Synthesis for Single Depth Image Super-Resolution". European Conference on Computer Vision, 2012.

[10] D. Glasner, S. Bagon, M. Irani, "Super-resolution from a single image," Computer Vision, 2009 IEEE 12th International Conference on , vol., no., pp.349,356, Sept. 29 2009-Oct. 2 2009

[11] S. D. Babacan, R. Molina, A. K. Katsaggelos, "Variational Bayesian Super Resolution", IEEE Transactions on Image Processing, vol.20, no.4, pp.984,999, April 2011

[12] MDSP Super Resolution and demosaicing Datasets. Available: http://users.soe.ucsc.edu/ milanfar/software/ sr-datasets.html.

[13] L. Xu, J. Jia, S. B. Kang, "Improving sub-pixel correspondence through upsampling", Computer Vision and Image Understanding, Volume 116, Issue 2, February 2012, Pages 250-261, ISSN 1077-3142.

[14] S. Farsiu, Resolution Enhancement Software: http://www1.idc.ac.il/toky/videoproc-07/projects/superres/ srproject.html.

[15] S. Villena, M. Vega, D. Babacan, J. Mateos, R. Molina, A. K. Katsaggelos, Superresolution Software, http://decsai.ugr.es/pi/superresolution/software.html.

[16] L. J. Karam, N. G. Sadaka, R. Ferzli, Z. A. Ivanovski, "An Efficient Selective Perceptual-Based Super-Resolution Estimator", IEEE Transactions on Image Processing, vol.20, no.12, pp.3470,3482, Dec. 2011

[17] J. Y. Bouguet, "Pyramidal implementation of the Lukas Kanade feature tracker. Description of the algorithm". http://robots.stanford.edu/cs223b04/algo_tracking.

[18] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. " Hierarchical Model-Based Motion Estimation". In Proceedings of the Second European Conference on Computer Vision (ECCV '92)