

Adaptive Outlier Rejection in Image Super-resolution

Mejdi Trimeche,¹ Radu Ciprian Bilcu,¹ and Jukka Yrjänäinen²

¹ *Multimedia Technologies Laboratory, Nokia Research Center, Visiokatu 1, 33720 Tampere, Finland*

² *Symbian Product Platforms, Nokia Technology Platforms, Hermiankatu 12, 33720 Tampere, Finland*

Received 29 November 2004; Revised 10 May 2005; Accepted 27 May 2005

One critical aspect to achieve efficient implementations of image super-resolution is the need for accurate subpixel registration of the input images. The overall performance of super-resolution algorithms is particularly degraded in the presence of persistent outliers, for which registration has failed. To enhance the robustness of processing against this problem, we propose in this paper an integrated adaptive filtering method to reject the outlier image regions. In the process of combining the gradient images due to each low-resolution image, we use adaptive FIR filtering. The coefficients of the FIR filter are updated using the LMS algorithm, which automatically isolates the outlier image regions by decreasing the corresponding coefficients. The adaptation criterion of the LMS estimator is the error between the median of the samples from the LR images and the output of the FIR filter. Through simulated experiments on synthetic images and on real camera images, we show that the proposed technique performs well in the presence of motion outliers. This relatively simple and fast mechanism enables to add robustness in practical implementations of image super-resolution, while still being effective against Gaussian noise in the image formation model.

Copyright © 2006 Mejdi Trimeche et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Nowadays, digital cameras are being integrated into more versatile and portable computing platforms such as camera-phones or PDA's. Often, the intrinsic image quality is limited due to packaging and pricing constraints. On the other hand, the computational and memory resources on mobile devices are increasing all the time. It is already possible to consider the implementation of sophisticated and computationally intensive image processing algorithms.

Super-resolution (SR) [1–3] is considered to be one of the most promising techniques that can help overcome the limitations due to optics and sensor resolution. The technique consists in combining a set of low-resolution (LR) images portraying slightly different views of the same scene in order to reconstruct a high-resolution (HR) image of that scene. The idea is to increase the information content in the final image by exploiting the additional spatio-temporal information that is available in each of the LR images.

In practice, the quality of the super-resolved images depends heavily on the accuracy of the motion estimation; in fact, subpixel precision in the motion field is needed to achieve the desired improvement. Global parametric motion estimation using affine or projective models can provide accurate enough registration, which positively impacts the overall performance of the SR algorithms. If the images

exhibit optical distortions, higher-order polynomial models can be used to obtain better pixel correspondence within the LR images. One major problem with global registration techniques is that they are limited to the assumed parametric model, and more importantly, they completely fail in the presence of local outliers. For example, such outliers may be due to moving objects inside the scene or due to the presence of repetitive textures or localized noisy areas. In those cases, the super-resolved image can exhibit severe artifacts. Local registration techniques such as optical flow are capable of handling moving objects; however, their performance suffers from lack of precision [4] and the result is not completely prone to outliers. For these reasons, robustness towards registration errors is a critical requirement in super-resolution, especially if we target to realize commercial implementations. Moreover, if we consider current mobile devices, we can afford only a limited number of LR frames in the memory buffer; so it is useful to consider the optimized algorithms that reject localized outliers, but that are able to exploit the rest of the image areas to improve the final resolution.

Several solutions have been proposed to handle registration errors by solving them as a part of the regularization of the solution [5–7]. In [5, 6], motion error noise is incorporated as a priori information within the smoothness prior and the result image is obtained as the MAP solution.

In [7], a regularization functional is plugged in a constrained least-squares setting and solved by iterative gradient descent. This approach for handling the registration error as a part of the regularization certainly helps towards the conditioning of the ill-posed inverse problem. However, it is argued in [8] that for large magnification factors, and regardless of the number of LR images used, regularization suppresses useful high-frequency information and ultimately leads to smooth results. Note that in most of the literature, localized motion outliers are not properly handled in the model. Further, it is implicitly assumed that the extra resolution content is equally distributed among all LR images, and usually the result is obtained by averaging the contributions from all LR images, which propagates the outlier pixels from any of the LR images into the final HR image.

In [9], it was shown through simulations that in the presence of small errors due to motion estimation or due to inconsistent pixel areas in the consecutive frames, the combined noise is better modelled with a Laplacian distribution rather than a Gaussian distribution. So, if this is taken into consideration, the mixed noise model is best handled through the minimization of the L_p ($1 \leq p \leq 2$) norm. Specifically, if the L_1 norm is considered, the pixelwise median minimizes the corresponding cost function, and when used together with the bilateral prior regularization [10], the solution was robust towards errors and still preserved details near sharp edges. In the context of super-resolution reconstruction, the median filter was used earlier [11] in the fusing process of the gradient images. It was shown that together with a bias detection procedure, it is possible to increase resolution even for those regions that contained outlier objects. However, it is well known that the median operator is not optimal for filtering Gaussian noise. Also, the median tends to consistently eliminate those measurements that significantly deviate from the majority and which may contain most of the novel high-frequency information. So at least in principle, there is a delicate trade-off between outlier rejection performance, noise removal capability, and the capability to reconstruct aliased high frequencies. One possible approach is to consider studying, instead of the mean or median filters, the α -trimmed mean or $\{r, s\}$ -trimmed mean¹ in the fusing process. The generalized class of order statistics filters, or L -filters [12] constitute a suitable filtering framework to derive the desired balance between the different trade-offs that are involved in the fusing process of the LR images. We have used this approach [13] to super-resolve text images by emphasizing either the maximum or minimum values to enhance the contrast near character edges.

In order to efficiently handle localized outliers, we propose in this paper to use an adaptive FIR scheme that automatically reduces the contribution of the outliers and averages the rest of the pixels. As the scanning progresses over the image grid, the weights associated with each LR image are adapted using an LMS estimator. We used the median estimator as an adaptation criterion that tunes the FIR

coefficients to reject consistent outliers. Our approach is different in that we use the median estimator as an intermediate step in the adaptation process, and this inherently eliminates the need for a bias detection procedure [11], making the overall algorithm more robust to Gaussian noise in the image formation model.

The rest of the paper is organized as follows. In Section 2, we present the assumed imaging model. In Section 3, the general framework of the iterative super-resolution is presented. In Section 4, we review briefly the existing fusing techniques, and we explain the issues that need to be addressed in order to tune the SR algorithm for robustness against outlier regions. In Section 5, we introduce our approach that uses an adaptive FIR filter to combine the gradient images. In Section 6, we show the experimental results, and Section 7 concludes the paper.

2. IMAGING MODEL

In this section, we formulate the general model that relates the HR image to the LR observations. The degradation process involves consecutively, geometric transformation, sensor blurring, spatial subsampling, and an additive noise term. In continuous domain, the forward synthesis model can be described as follows: consider N observed LR images, we assume that these images are obtained as different views of a single continuous HR image. Following a similar notation as in [14], the i th LR image can be expressed as

$$g_i(x, y) = S \downarrow (h_i(u, v) * f(\xi_i(x, y))) + \eta_i(x, y), \quad (1)$$

where g_i is the i th observed LR image, f is the HR reference image, h_i the point spread function (psf), ξ_i the geometric warping, $S \downarrow$ the downsampling operator, η_i additive noise term, and $*$ denote the convolution operator. The overall degradation process is illustrated in Figure 1.

After discretization, the model can be expressed in matrix form as follows:

$$\bar{g}_i = \mathbf{A}_i \bar{f} + \bar{\eta}_i. \quad (2)$$

The matrix \mathbf{A}_i combines successively, the geometric transformation ξ_i , the convolution operator with the blurring parameters of h_i , and the downsampling operator $S \downarrow$ [15]. Note that in (2), \bar{g}_i , \bar{f} , and $\bar{\eta}_i$ are lexicographically ordered.

3. ITERATIVE SUPER-RESOLUTION

The super-resolution reconstruction problem can now be described as estimating the best HR image, which when appropriately warped and downsampled by the model in (2) will generate the closest estimates of the LR images \bar{g}_i . If we assume that $\bar{\eta}_i$ is Gaussian white noise, the least-squares solution also maximizes the likelihood that each LR image is the result of an observation of the original HR image. In other words, for each observation \bar{g}_i , the corresponding solution is a high-resolution image \bar{f} , which minimizes the following cost function:

$$\epsilon_i = \|\hat{g}_i - \bar{g}_i\|^2 = \|\mathbf{A}_i \bar{f} - \bar{g}_i\|^2, \quad (3)$$

¹ These filters are effective against impulsive outliers, and are relatively easy to tune.

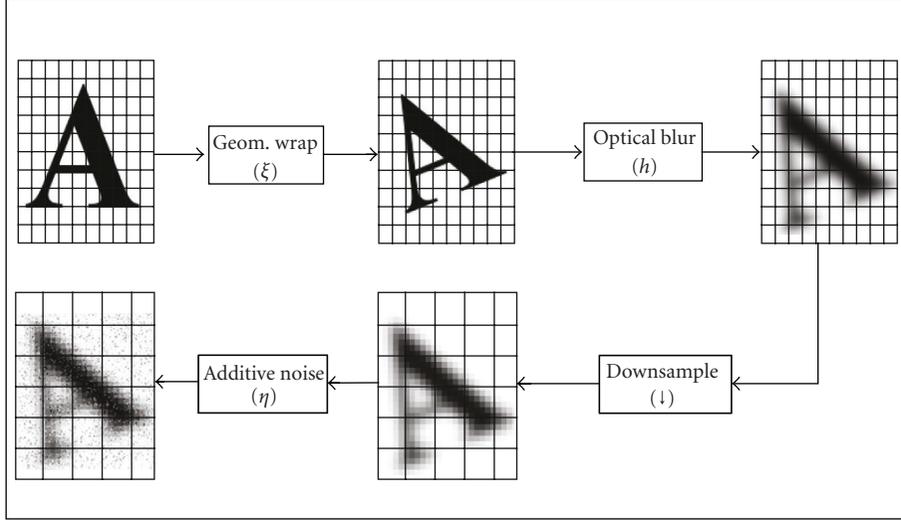


FIGURE 1: An illustration of the image degradation process following the model in (2).

with \hat{g}_i being the simulated LR image through the forward imaging model.

In order to minimize the error functional in (3), the method of iterative gradient descent is commonly employed. This optimization technique seeks to converge ϵ_i towards a local minimum following the trajectory defined by the negative gradient. That is, at iteration n , the high-resolution image according to observation \bar{g}_i , is updated as

$$\bar{f}^{n+1} = \bar{f}^n + \mu_i^n \bar{r}_i^n, \quad (4)$$

μ_i^n and \bar{r}_i^n are, respectively, the step size and the residual gradient at iteration n .

The residual gradient \bar{r}_i^n is computed as follows:

$$\bar{r}_i^n = \mathbf{W}_i (\bar{g}_i - \mathbf{A}_i \bar{f}^n). \quad (5)$$

The matrix \mathbf{W}_i combines successively the upsampling, and the inverse geometric warp ξ_i^{-1} . The step size μ_i^n that achieves the steepest descent is given by [16]

$$\mu_i^n = \frac{\|\bar{g}_i - \mathbf{A}_i \bar{f}^n\|^2}{\|\mathbf{A}_i \bar{r}_i^n\|^2}. \quad (6)$$

In (4), each scaled gradient term, $\bar{p}_i = \mu_i^n \bar{r}_i^n$, corresponds to the update image that verifies the reconstruction constraint for the i th observation \bar{g}_i . We define \mathbf{z}_k as the data vector that points to the values from all gradient images at pixel position k , $\mathbf{z}_k = \{p_i(k), i = 1, \dots, N\}$. In the process of SR reconstruction, we need to perform a temporal filtering operation that combines the observations in \mathbf{z}_k . For convenience of notation, we denote this filtering operator Φ . For each pixel k on the HR image grid, the resulting update value y_k is given as

$$y_k = \Phi(\mathbf{z}_k), \quad (7)$$

where Φ is a generic filtering operator that performs the fusing of the pixels from all available gradient images. Figure 2

depicts an illustration of the iterative SR implementation that we considered. Note that so far our formulation does not assume a proper regularization of the solution. Certainly, super-resolution is an ill-posed inverse problem, so regularization is necessary to obtain a stable solution. In the literature, there has been significant effort to formulate suitable prior models, and several solutions have been proposed for iterative super-resolution [6, 7, 10]. These solutions can be implemented in the iterative setting of Figure 2 by assuming a generic filter Γ that operates on the previous SR estimate \bar{f}^n or on the fused gradient image. If we denote s_k as the contribution that is due to the regularization process at pixel k , then at iteration n , the final output at each pixel k is updated as follows:

$$f_k^{n+1} = f_k^n + y_k + \mu^n \alpha s_k, \quad (8)$$

where α is the regularization parameter that controls the conditioning of the solution. In the rest of the paper, and in our experiments, we omitted the implementation of a regularization operator, that is, we assumed $s_k = 0$. We focus the discussion on the efficient implementation of the fusing process Φ in the presence of motion outliers.

4. FUSING THE GRADIENT IMAGES

Ideally, the fusing process defined by the operator Φ will retain the novel information from each LR frame, filter out the noise due to the image formation process, and of course reject the motion outliers. Thus, at least in principle, we shall consider all observations independently and design a filtering mechanism that adapts itself to instantly recognize and reject the outliers, while constantly adjusting its behavior according to the nonstationary noise distribution of the input images.

One straightforward implementation of the fusing process would be to select Φ as the mean filter. In this case, if

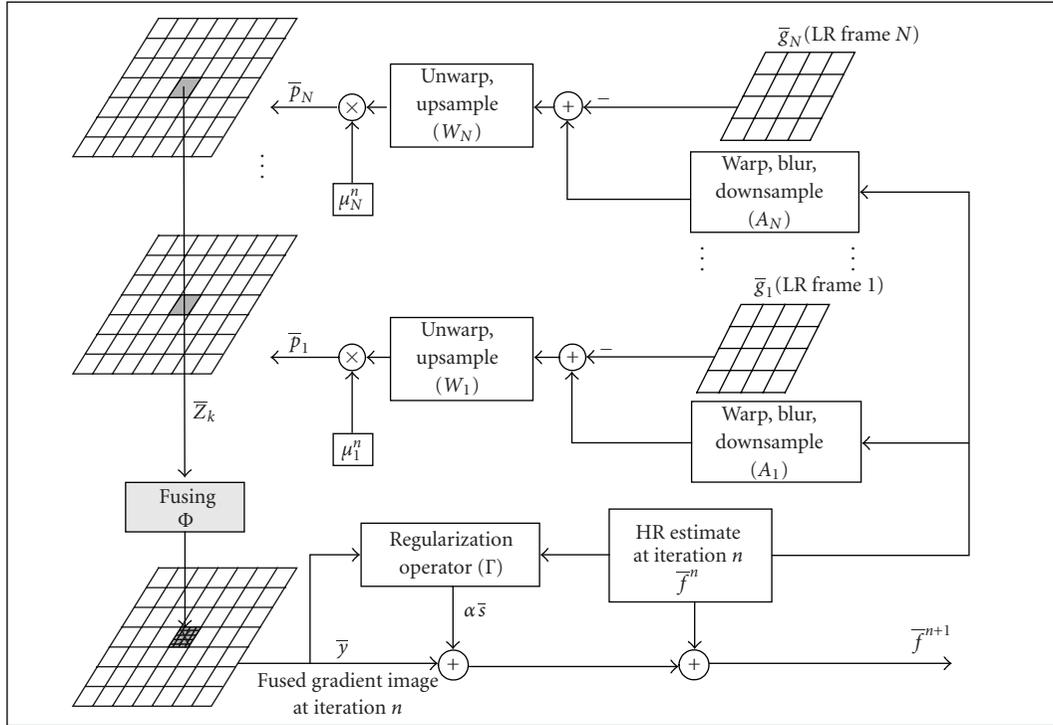


FIGURE 2: Generic block diagram of the iterative super-resolution process. The gradient images are combined using a filtering operator Φ that can be modulated depending on the application.

Gaussian noise is assumed in the imaging model, this implementation is equivalent to the maximum-likelihood solution. However, the solution is not robust against outliers. Another possibility is to select the median filter, which would be efficient against impulsive errors in \mathbf{z}_k . This idea was used earlier in iterative super-resolution [11] and was shown to improve the robustness against motion outliers. In fact, the median minimizes the L_1 cost function [10], which corresponds to the Laplacian distribution of the combined noise. However, in the case when the errors have a mixed distribution, for instance, Gaussian and impulsive, the class of *trimmed* mean filters might have better performance. Note that the filters discussed above can be derived as special cases of the generalized L -filters² which operate on the sorted data vector $\mathbf{z}_{(k)}$.

When we consider error modelling due to motion estimation, it is difficult in practice to assume a stationary distribution. This is especially true when dealing with local outliers, for example, due to moving objects inside the scene. More difficult is the case when the user tilts the camera, resulting in a significant perspective change. This situation is quite challenging for most motion estimation techniques, which may register parts of the image correctly, but may completely fail in some other regions. Hence, it is beneficial

to use an adaptive fusing strategy that is capable of automatically isolating localized outliers. In the following section, we introduce our approach which is based on spatially adaptive FIR filtering of the gradient images. We show that this technique enables the overall process to deal adequately with the outliers.

5. OUR APPROACH

5.1. Outlier rejection by adaptive FIR filtering

In (7), we chose to implement the fusing operator Φ as a weighted mean operator, that is, at each iteration, the update value y_k is calculated as the output of an FIR filter as follows:

$$y_k = \sum_{i=1}^N a_i p_i(k) = \mathbf{a}^T \mathbf{z}_k, \quad (9)$$

where \mathbf{a} is the FIR coefficient vector. The filter coefficients relate the contribution that each LR image brings into the fused image. In most conventional techniques, it is generally implied that all LR images contribute equally to the total gradient image, that is, $a_i = 1/N$, $i = 1, \dots, N$. However in the presence of outliers, the computed solution may be corrupted by the consistent presence of large projection errors coming from the same frames.

² For example, the median filter is a special case of the L -filters, which can be obtained by selecting all coefficients to be zero, except for the center coefficient that has unity value.

To take into account the presence of outlier regions at the fusing stage, we introduce an adaptation mechanism that modulates the weights associated with each input image. The coefficients of the FIR filter are varying with the pixel location k , that is in (9), we use \mathbf{a}_k instead of \mathbf{a} .

5.2. Coefficient adaptation

For its simplicity and computational efficiency, we chose to use the least mean-squared (LMS) estimator to adapt the filter coefficients. The coefficients are updated progressively according to a predetermined scanning pattern across the selected image region ($k = 1 \dots L$). Our proposed method for spatially adapting the FIR coefficients and simultaneously computing the update value is described below:

- (1) initialization: $\mathbf{a}_0 = [1/N, \dots, 1/N]$;
- (2) for $k = 1 \dots L$,
 - (2.1) filtering: $y_k = \mathbf{a}_{k-1}^T \mathbf{z}_k$,
 - (2.2) error computation: $e_k = d_k - y_k = \text{median}(\mathbf{z}_k) - y_k$,
 - (2.3) coefficient update: $\mathbf{a}_k = \mathbf{a}_{k-1} + \lambda e_k \mathbf{z}_k$,
 - (2.4) move to next pixel location $k + 1$.

In the LMS coefficient adaptation shown above, λ is the step-size parameter. We set the desired response of the LMS estimator (d_k) to be the median of all errors. In this setting, the median is used to point out those frames that consistently present error values that deviate from the majority. For example, if the scanning progresses through an area where the i th LR image contains an outlier region, then pixel after pixel, the error with respect to the median is going to be large, and the coefficient bias due to $\lambda e_k \mathbf{z}_k(i)$ is going to decrement the corresponding FIR coefficient $\mathbf{a}_k(i)$. Figure 3 depicts an illustration of the proposed filtering method.

When combined with a suitable step size, the LMS estimator gathers reliable statistics from the immediate pixel neighborhood. The resulting FIR coefficients tend to stabilize, rejecting the outlier contribution, while still averaging the rest of the error values. Given a sufficient set of samples, the median can approximate the mean quite well [12], however, with a reduced set of LR images (fewer samples), the result can be biased, and that is why we chose to set it only as an intermediate step for the coefficient adaptation. The experiments in the following section confirm that this fusing scheme is also efficient to filter the Gaussian noise assumed in the image formation model.

Note that the desired response of the LMS estimation (d_k) can be changed to modulate the performance of the super-resolution process. In this case, we used the median estimator to tune the algorithm for robustness against local outliers. Other functions might be studied and plugged in d_k to obtain a specific property of the fusing process. For example, to speed up the reconstruction property for all input images, we can set $d_k = 0$. In this case, since we are fusing gradient images, the algorithm will favor the contribution of those LR images that consistently present most of the novel information.

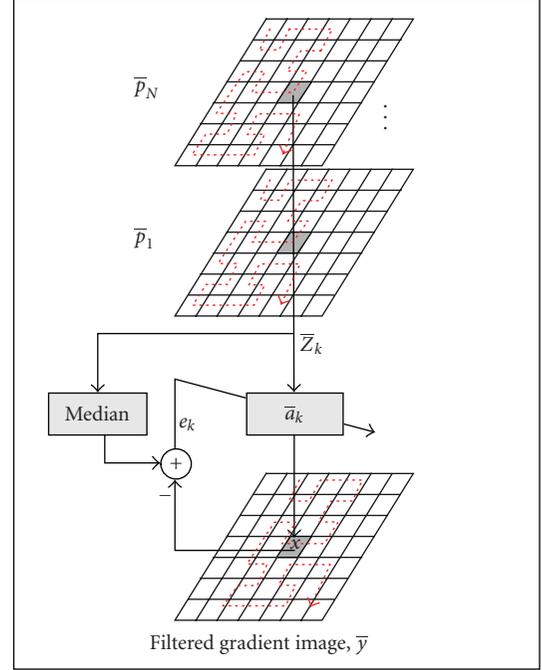


FIGURE 3: Block diagram of the proposed fusing method. The gradient images are combined with a spatially varying FIR filter. The coefficients of the FIR are chosen with an LMS estimator that is tuned to reject outliers.

5.3. Stability of LMS adaptation

Despite its simplicity and good adaptation performance, the LMS has also some sensible points that must be addressed. The first issue is the initialization of the step size λ . It is well known that the value of λ provides a tradeoff between the speed of convergence and quality of adaptation. If its value is large, the convergence is fast but at the expense of an increased adaptation error. On the contrary, a small step size provides good adaptation performance, but the transient time is increased.

The problem of stability and adaptation speed for the LMS estimator is well studied in the literature [17]. Several modified solutions have been proposed to solve the problem for 1D signals. To ensure the stability of the LMS estimator, the step size must be bounded:³

$$0 < \lambda < \frac{2}{3 \text{tr}[\mathbf{R}]}, \quad (10)$$

where $\mathbf{R} = E\{\mathbf{z}_k \mathbf{z}_k^T\}$ is the cross-correlation matrix of the input vector, $E\{\cdot\}$ denotes the expectation operator, and $\text{tr}[\mathbf{R}]$ is the sum of the diagonal elements of matrix \mathbf{R} .

³ For several applications, relaxed boundary conditions may be used for λ . However, the stability condition in (10) has been shown to ensure stability for a wider class of input statistics, including nonstationary signals.

The above stability criterion is valid and easy to implement when the input sequence is stationary. However, for nonstationary inputs, as it is often the case with image data, the cross-correlation matrix \mathbf{R} changes when scanning through the image. As a consequence, the stability interval in (10) is not fixed throughout the entire image. To overcome this difficulty, the simplest solution consists in selecting a small value of λ , such that it is always within the stability bounds for all pixel locations. However, such a small step size will significantly slow down the convergence. Moreover, although in some parts of the image, a small step size will be beneficial to avoid fast and unnecessary variations in the FIR coefficients, a larger value of λ will be required in regions containing outliers.

To overcome those difficulties and to simplify the setup of the algorithm, we have implemented the normalized LMS (NLMS). The gradient step factor is normalized by the energy of the data vector. In our case, λ_k is modified depending on the pixel location, and is given by the following equation:

$$\lambda_k = \frac{\gamma}{\|\mathbf{z}_k\|^2}, \quad (11)$$

where $\|\mathbf{z}_k\|$ is the Euclidean norm of the vector \mathbf{z}_k .

With this setup, the stability condition of (10) becomes

$$0 < \gamma < \frac{2}{3}. \quad (12)$$

As it can be seen from (11), the algorithm maintains a step-size value that is inversely proportional to the input power. As a result, the normalized algorithm converges faster within fewer samples in many cases. To overcome the possible numerical problems when $\|\mathbf{z}_k\|^2$ is very close to zero, the step size of the Normalized LMS in (11) is usually modified as follows [17]:

$$\lambda_k = \frac{\gamma}{c + \|\mathbf{z}_k\|^2}, \quad (13)$$

with $c > 0$. Note that the stability interval of γ remains unchanged, and is the same as in (12). In (13), the constant c can be used to prevent very large changes of the step size. If we use a relatively large value, we decrease the speed of coefficient adaptation, but on the other hand, we improve the robustness of the employed NLMS adaptation against fast changing edges and other local image details that are present in the gradient images.

5.4. Scanning pattern

To better handle outlier regions, especially those due to moving objects, the proposed fusing algorithm is most efficient when the coefficient adaptation procedure stays localized around the 2D outlier patterns. Ideally, we would like the scanning path to satisfy the following constraints:

- (1) cover the entire image area,
- (2) pass through each point only once,
- (3) stay in the highly correlated image areas as long as possible.

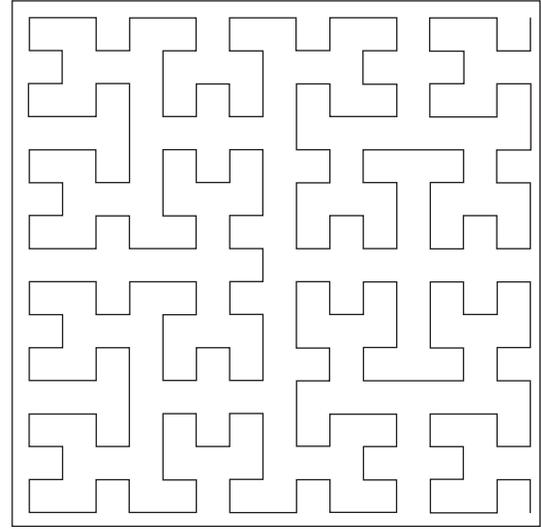


FIGURE 4: Hilbert scanning pattern is used to maximize the efficient adaptation of the FIR coefficients.

By default, if we use the simple raster scan over the entire HR image, we fail to satisfy condition (3). One immediate solution is to divide the image into areas of equal size, and to apply the filtering in these areas independently, with careful handling of the borders. Instead of the raster scan, space-filling curves can be used to traverse the image plane during the filtering process. These curves have been successfully used in several other applications such as image coding [18]. This mode of scanning through the pixels, though more complicated, has the important advantage of staying localized within areas of similar frequencies before moving to another area. Figure 4 shows the Hilbert scanning pattern for a rectangular window of 16×16 . Notice that the filtering following the Hilbert path will stay longer in regions having 2D correlation than the one following the raster scan. In our implementations, we tested the Hilbert space filling curves of 64×64 , as well as 16×16 . It was clear to us that applying this type of scanning pattern significantly enhanced the coefficient adaptation and allowed to use smaller values of λ , thus resulting in better stability of the LMS estimator. It is worth mentioning that these scanning patterns are easily integrated in the overall implementation using predefined look-up tables.

The typical space filling patterns (such as Peano, Hilbert [18]) are defined over grid areas that are powers of 2. To confine with this restriction, we divided the image area into smaller tiles that are powers of 2. This option is rather a limitation to the performance of the LMS estimator. Moreover, if the tiles happen inside an outlier area, some artifacts might appear at the borders of the tiles, and may get amplified with the iterations. To avoid these artifacts, one immediate solution is to slow down the LMS adaptation by decreasing λ . Another solution is to smooth the coefficients at the borders of adjacent tiles, but this procedure makes the overall implementation rather cumbersome. Better solution would be

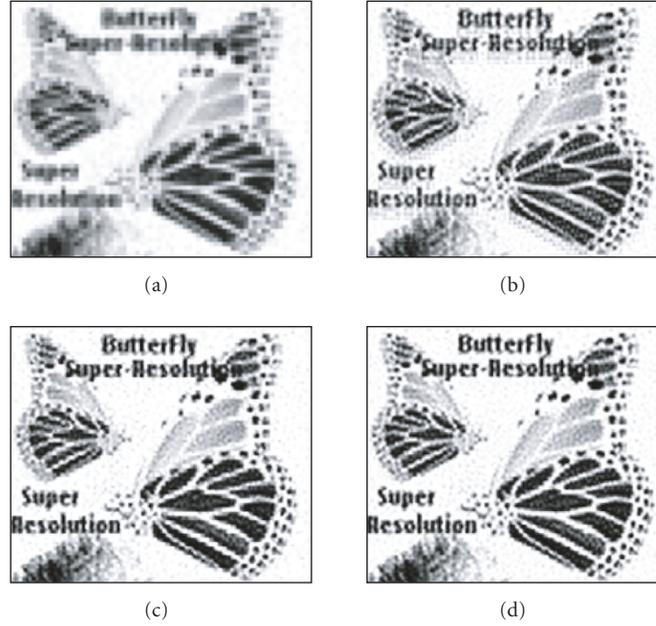


FIGURE 5: Five noisy LR were synthetically generated by random warp and downsampling by 2, additive Gaussian noise ($\sigma^2 = 40$), and 1 outlier image. (a) Reference LR image, SNR = 11.85. (b) SR result with mean fusing (ML solution) after 10 iterations, SNR = 14.12. (c) Iterative median fusing after 10 iterations, SNR = 15.32. (d) SR using adaptive FIR filtering after 10 iterations, SNR = 15.99.

to apply space filling curves that are defined over arbitrary sized images, for example the scanning technique that is proposed in [19] provides an elegant method for preserving two-dimensional continuity.

To further enhance the stability of the LMS estimator, the adapted FIR coefficients are saved in between successive iterations of the super-resolution algorithm. These are used to initialize the input coefficients at the beginning of each scanning block. In fact, in the presence of consistent outliers, the coefficients tend to stabilize quickly after scanning through a small part of the image (see Figure 6), and the outlier regions can be pointed out, since their corresponding coefficients are much smaller than the rest. The detected outlier regions can be thrown away when processing the following iterations to reduce the computational complexity of the overall algorithm.

6. SIMULATION RESULTS

In this section, we show the performance of the proposed technique. First, we tested the algorithm on a sequence of synthetic test images. The images, 5 in total, were generated from a single HR image according to the imaging model described in (1). The original HR image was randomly warped using an 8 parameter projective model. The registration parameters were saved for the reconstruction experiments. We used a continuous Gaussian psf ($\text{psf} = 0.5$) as the blurring operator, and we downsampled the images by 2 to obtain the 5 LR images. All images were contaminated with additive Gaussian noise ($\sigma^2 = 40$). Out of the 5 obtained images, we singled out one image, and we introduced a deliberate error

in its registration parameter corresponding to a translation error of 1.5 pixels on the LR image grid.

We ran the algorithm on the resulting set of images. Figure 6 shows the trajectory of the adapted coefficients through the first iteration. In this experiment, we fixed a small LMS step size, $\gamma = 5 \cdot 10^{-7}$. Although the step size is relatively small, the LMS estimator successfully singles out the outlier image (third image) by decreasing its corresponding FIR coefficient $a(3)$ after scanning through a small part of the image.

We compared the results of iterative super-resolution obtained using the proposed fusing process against the mean and median filters. For the three compared techniques, we used the same step size μ_i^n in the update (4). Figure 5 shows the result images; both our fusing technique and the median fusing successfully singled out the outlier image and improved the robustness of the overall SR process. Compared to median fusing, the proposed filtering has shown better robustness towards noise, and was able to reconstruct finer character details. Figure 7 shows the corresponding SNR values across the iterations. The SNR figures confirm that the proposed filtering scheme consistently performs better than the mean and median filters. It is worth mentioning that the intermediate result was truncated in between iterations, which helped to constrain the solution and achieve steadier convergence for this set of almost binary images. Note that in all experiments, we have not used a regularization operator because we are mainly interested to isolate the effect of the fusing strategy. We assume that it would be possible to enhance the final result when we correctly assert some prior knowledge about the image content in the regularization step.

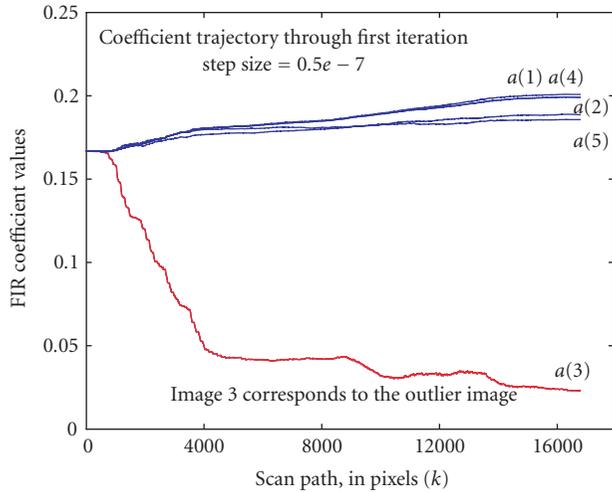


FIGURE 6: Adaptation of the filter coefficients during the first iteration corresponding to the image shown in Figure 5(d). The coefficient $a(3)$ reflecting the contribution of the outlier image is automatically decreased.

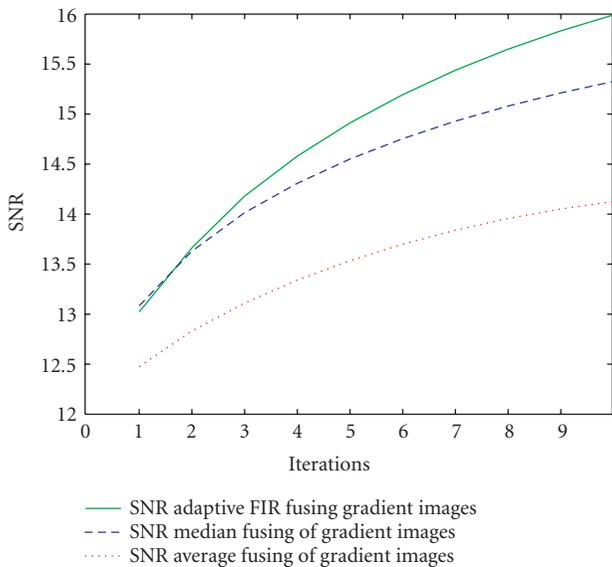


FIGURE 7: SNR comparison across the first 10 iterations for the super-resolved images shown in Figure 5. SNR curves for (a) proposed adaptive solution, (b) median fusing of the gradient images, and (c) average fusing of the gradient images.

In Figure 8, we repeated the same experiment. We generated 4 LR images with the same parameters described above, but in this setting, we selected the last LR frame, and we inserted several outlier objects. Figure 10 shows the SNR values across the iterations for the three fusing techniques. The convergence of the SR algorithm is fast during the first 4 iterations of the steepest descent (SD), but in the following iterations, the SNR starts to oscillate without significant improvement. This example illustrates the need for a regularization step in order to ensure the convergence of the solution. Early abortion of the iterations is the only available option to avoid over-amplified edges. In Figure 8, we

display the results after 4 iterations, again, both the median and the proposed solution eliminated the outlier areas, whereas the mean failed. Better SNR performance, as well as better visual result, was obtained with our fusing method (Figure 8(f)). Figure 9 shows the trajectory of the adapted coefficients through the last iteration. The coefficient $a(4)$ reflecting the contribution of the last LR image is automatically decreased when stepping inside an outlier area. When the scanning steps outside the outlier area, the coefficient increases again. The other coefficients corresponding to the nonoutlier images are kept around the same level. As indicated in Figure 9, basically our method operates as a weighted mean filter, except for the detected outlier areas. So, compared to median fusing, an improved performance against Gaussian noise is predictable. In Figure 8(f), the expert eye will notice some artifacts near the borders of the Hilbert scanning blocks that contain outlier regions. These are due to the fast and abrupt change of the coefficient values on the borders of the subareas that were used for scanning. To reduce this effect, some implementation enhancements can be designed, such as the use of larger scanning areas or the smoothing of the coefficients near adjacent blocks.

Figure 11 shows the super-resolved images obtained using 5 LR scenery images taken with a camera phone (Nokia 6600). To register the pixels on the reference HR grid, we used hierarchical block matching in the central parts of the image, followed by the estimation of the global projective motion parameters. In one of the images, the registration failed due to a significant perspective change. Figure 11(a) shows the interpolated reference frame (pixel replication). Figure 11(b) shows the result when simple mean fusing is used; note the picture of a ghost car that does not belong to the original scene. Figures 11(c) and 11(d) show, respectively, the results after 5 iterations when fusing with the median and with the proposed technique. For both images, the sharpness of the scene detail is significantly enhanced and the outlier region in the bottom of the image is successfully eliminated. In this specific set of input images, the clouds were particularly difficult to register because they were deformed from one shot to the next. In fact, for the corresponding area, the only information that needs to be considered is the one that comes from the reference frame. This specific example illustrates the inadequacy of the median filter to fuse this kind of fuzzy regions (Figure 11(c)). Since the input samples do not constitute a reliable majority to obtain a correct vote, the median filter picks borders randomly from any one of the input images. The proposed filtering does not solve the problem completely, however, it prevents the formation of excessive artifacts in those regions (clouds in Figure 11(d)). The reason is that similar FIR coefficients are employed when filtering adjacent pixels, unless a clear outlier frame is consistently voted after scanning through several consecutive pixels, which is not the case in this example. Note that Zomet et al. [11] have tackled this problem and proposed to use a bias detection procedure in conjunction with the median. The detection procedure outputs a binary mask indicating where to perform the filtering. However it is unclear how the thresholds and the windows would be selected.

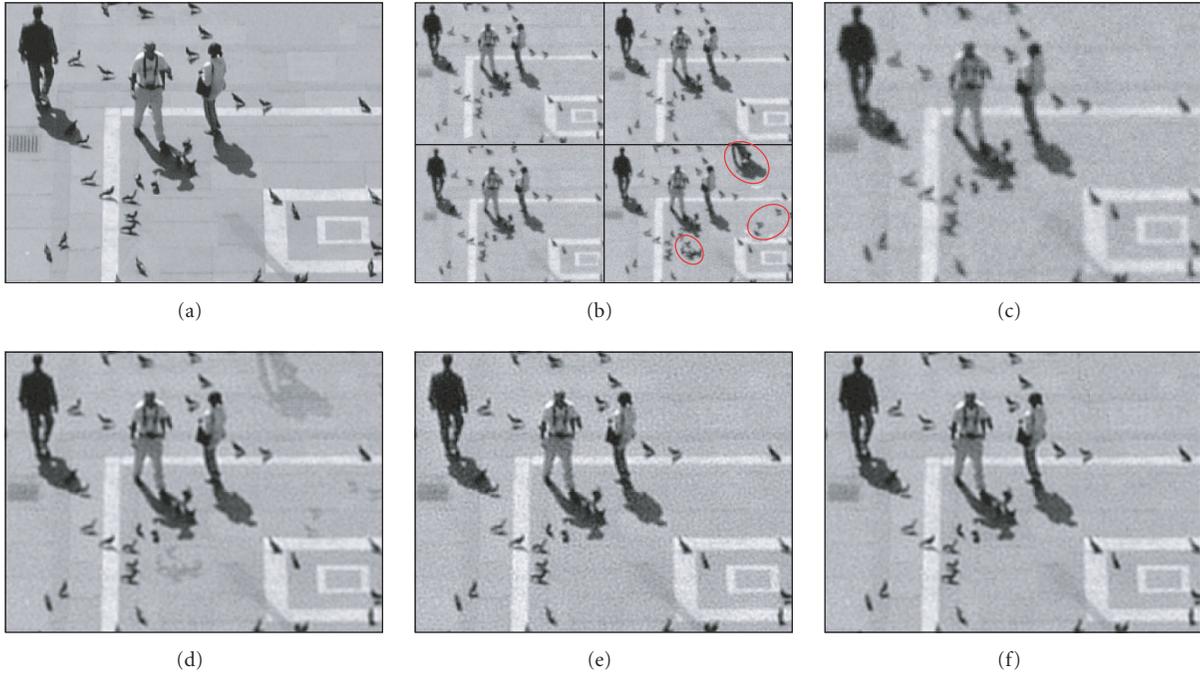


FIGURE 8: (a) Original HR image. (b) The set of LR images used in the experiment: 4 noisy LR were synthetically generated from the original HR image. The last image was generated from the same image with artificial objects inserted. All images were shifted, downsampled by 2, and contaminated with additive Gaussian noise ($\sigma^2 = 40$). (c) Interpolated reference image (pixel replication), SNR = 8.6. (d) SR result using iterative mean fusion after 4 iterations, SNR = 11.4. Remark the shaded outlier regions. (e) SR result using iterative median fusion after 4 iterations, SNR = 11.3. (f) SR using adaptive FIR filtering after 4 iterations, SNR = 12.1.

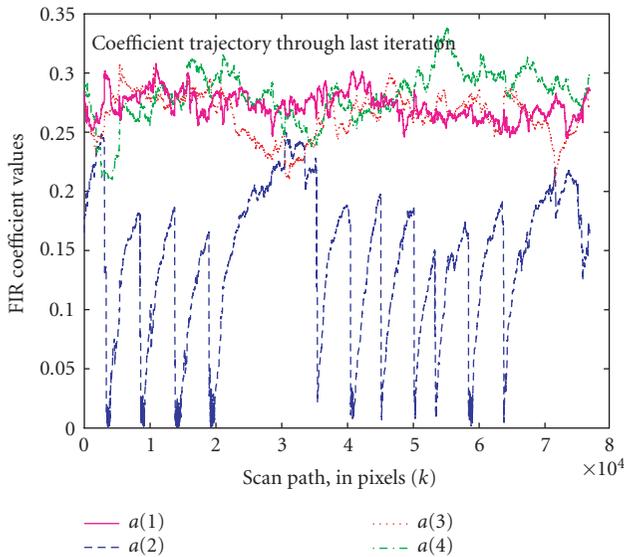


FIGURE 9: Adaptation of the filter coefficients during the fourth and last iteration corresponding to the result in Figure 8(f). The coefficient $a(4)$ reflecting the contribution of the last LR image is automatically decreased when inside an outlier region, when the scanning steps outside the outlier area, the coefficient increases again. 16×16 Hilbert scanning is used in this example.

Figure 12 shows a similar example depicting the performance of the proposed algorithm on real image scenes. We used 5 LR images that were cropped from VGA pictures

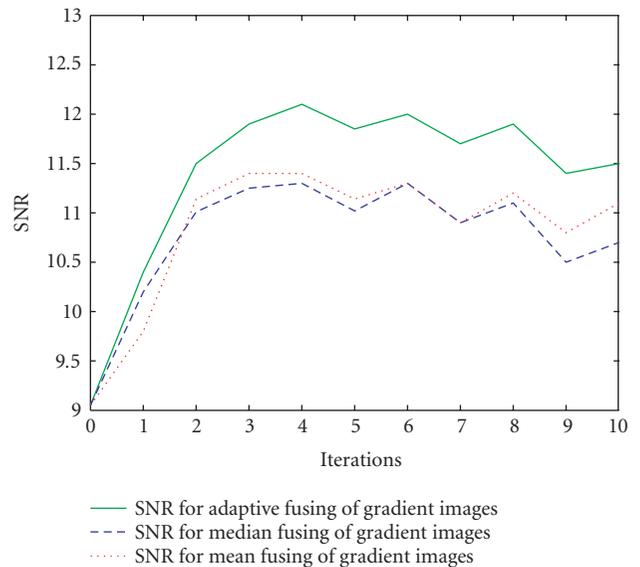


FIGURE 10: SNR comparison across the first 10 iterations for the super-resolved images shown in Figure 8. SNR curves for (a) proposed adaptive solution, (b) median fusion of the gradient images, and (c) average fusion of the gradient images.

imaged at close range (the images are JPEG compressed at 90%). The last frame contained an outlier object. Again, note that the median fusion (c) and our technique (d) successfully

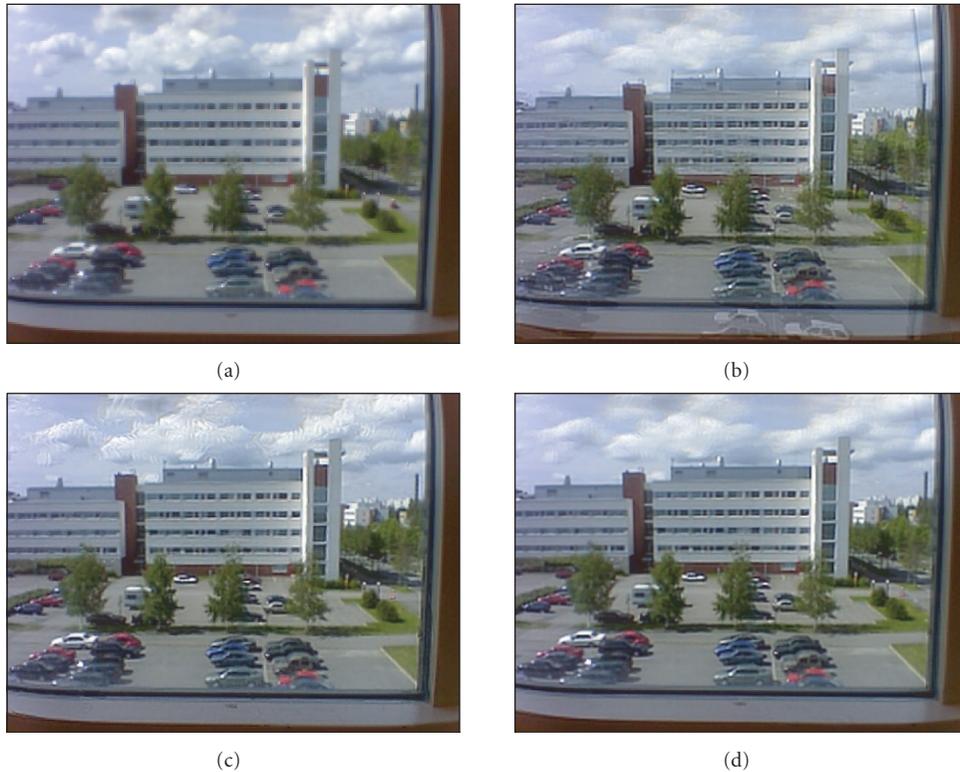


FIGURE 11: The super-resolved images using the proposed implementation. Five LR images were used. The global motion estimation failed to register at least one frame. (a) Interpolated reference frame, zoom factor 2; (b) result using mean fusing; (c) result using median fusing; and (d) super-resolved image using the proposed algorithm.

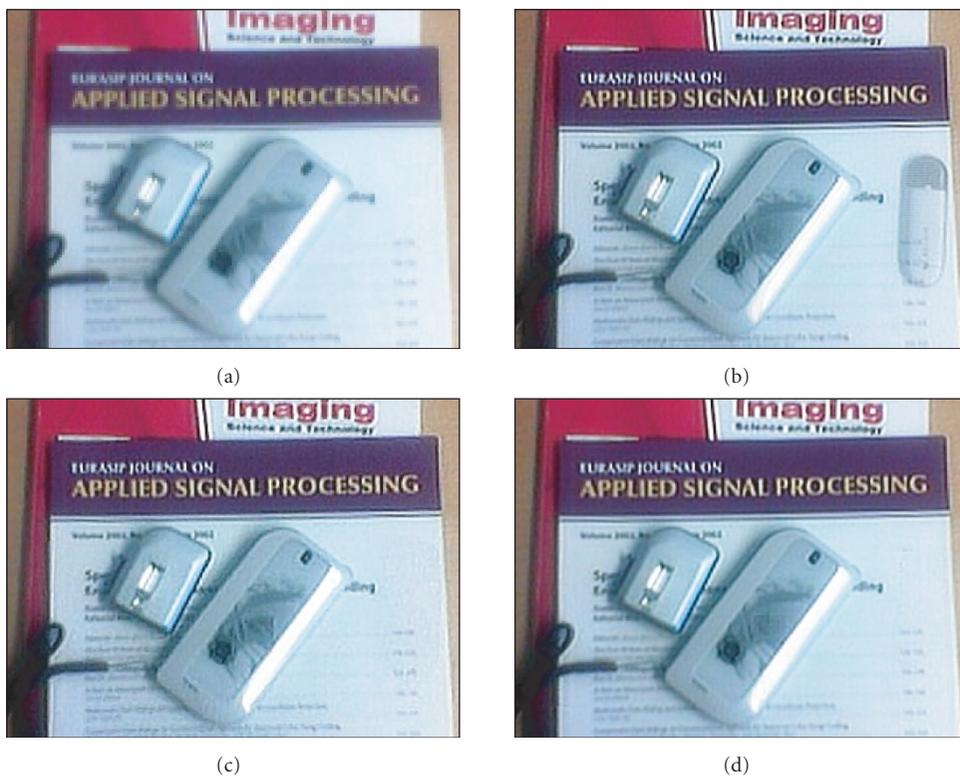


FIGURE 12: The super-resolved images using the proposed implementation. We used 5 LR Images that were cropped from VGA images taken with a camera phone (Nokia 9500). One outlier object appears in the last frame. (a) Zero-order interpolated reference frame, zoom factor 2; (b) result using mean fusing, (c) using median fusing, and (d) super-resolved image using the proposed algorithm.

wiped out the outlier object from the reconstructed scene. Looking more closely, we can notice that the result image of the proposed filtering method has less noise artifacts, especially on smooth areas.

7. CONCLUSION

In this paper, we have proposed to use adaptive FIR filtering of the gradient images in iterative super-resolution. The FIR coefficients are adapted using an LMS estimator that is tuned to detect motion outliers. The algorithm performs adequately in the presence of Gaussian noise, and is capable of automatically isolating outlier regions, which are due to registration errors. The proposed method is useful to enhance the robustness of super-resolution implementations.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and insightful steering to enhance the content of this paper.

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] S. Chaudhuri, Ed., *Super-Resolution Imaging*, Kluwer Academic, Boston, Mass, USA, 2001.
- [3] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.
- [4] S. Baker and T. Kanade, "Super-resolution optical flow," Tech. Rep. CMU-RI-TR-99-36, Robotics Institute, Carnegie Mellon University, Pittsburgh, Pa, USA, 1999.
- [5] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 996–1011, 1996.
- [6] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, 1997.
- [7] E. S. Lee and M. G. Kang, "Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration," *IEEE Transactions on Image Processing*, vol. 12, no. 7, pp. 826–837, 2003.
- [8] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.
- [9] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Robust shift and add approach to super-resolution," in *Applications of Digital Image Processing XXVI*, vol. 5203 of *Proceedings of SPIE*, pp. 121–130, San Diego, Calif, USA, August 2003.
- [10] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super-resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.
- [11] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. 645–650, Kauai, Hawaii, USA, December 2001.
- [12] J. Astola and P. Kuosmanen, *Fundamentals of Nonlinear Digital Filtering*, CRC Press, New York, NY, USA, 1997.
- [13] M. Trimeche and J. Yrjänäinen, "Order filters in super-resolution image reconstruction," in *Image Processing: Algorithms and Systems II*, vol. 5014 of *Proceedings of SPIE*, pp. 190–200, Santa Clara, Calif, USA, January 2003.
- [14] D. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 2, pp. 627–634, Kauai, Hawaii, USA, December 2001.
- [15] M. Elad and A. Feuer, "Super-resolution reconstruction of image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 817–834, 1999.
- [16] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*, chapter 6, Institute of Physics Publishing (IOP), Bristol, UK, 1998.
- [17] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1996.
- [18] N. Max, "Visualizing Hilbert curves," in *Proceedings of IEEE Visualization '98*, pp. 447–450, 564, Research Triangle Park, NC, USA, October 1998.
- [19] A. Perez, S. Kamata, and E. Kawaguchi, "Peano scanning of arbitrary size images," in *Proceedings of 11th IAPR-IEEE International Conference on Pattern Recognition (ICPR '92)*, vol. 3, pp. 565–568, The Hague, the Netherlands, August–September 1992.

Mejdi Trimeche received the B.S. degree in electrical engineering from Bilkent University, Ankara, Turkey, in 1998. In 2000, he received the M.S. degree with distinction from the Department of Information Technology, Tampere University of Technology (TUT), Tampere, Finland. He is currently a Ph.D. student with the same university. He joined Nokia in 2000. Since then, he has been working in various topics on image and video processing. Currently, he works as a Senior Research Engineer in Nokia Research Center. His research interests include image and video processing, in particular, algorithms for image restoration and enhancement. Other active topics of interest include computer vision and content-based indexing and retrieval.



Radu Ciprian Bilcu received the B.S. and M.S. degrees from the Technical University of Cluj-Napoca, Romania, in 1995 and 1996, respectively, and the Dr.Tech. degree from Tampere University of Technology, Tampere, Finland, in 2004. He has published 28 papers in international journals and conferences and holds three patents. From 1999 to 2004 he was a Researcher at the Institute of Signal Processing, Tampere University of Technology, where he was also involved in teaching. In 2004, he joined Nokia Research Center as a Research Engineer in the field of image processing. His research interests include adaptive systems, adaptive algorithms and applications to image processing, communications, echo control, audio, and speech.



Jukka Yrjänäinen studied for the degree of Engineering Diploma in Tampere University of Technology. He joined Nokia in 1992. Currently he works as a Senior Technology Manager in Nokia Technology Platforms, Tampere, Finland. He is responsible for imaging and video related R&D activities. His interests include image and video signal processing, multimedia applications, camera technologies, and processing architectures for multimedia.

