EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

Robust video super resolution algorithm using measurement validation method and scene change detection

Minjae Kim[1], Bonhwa Ku[1], Daesung Chung[1], Hyunhak Shin[1], David K Han[2] and Hanseok Ko[1*]

Abstract

Explicit motion estimation is considered a major factor in the performance of classical motion-based super resolution (SR) algorithms. To reconstruct video frames sequentially, we applied a dynamic SR algorithm based on the Kalman recursive estimator. Our approach includes a novel measurement validation process to attain robust image reconstruction results under inexplicit motion estimation. In our method, the suitability for high-resolution pixel estimation is determined by the accuracy of motion estimation. We measured the accuracy of the image registration result using the Mahalanobis distance between the input low-resolution frame and the motion compensated high-resolution estimation. We also incorporate an effective scene change detection method dedicated to the proposed SR approach for minimizing erroneous results when abrupt scene changes occur in the video frames. According to the ratio of well-aligned pixels (i.e., motion is compensated accurately) to the total number of pixels, we are able to detect sudden changes of scene and context in the input video. Representative experiments on synthetic and real video data show robust performance of the proposed algorithm in terms of its reconstruction quality even with errors in the estimated motion.

1. Introduction

In imaging devices and applications, we often have to deal with degraded low resolution (LR) images due to because of the theoretical and practical limits of imaging devices. In visual surveillance and satellite imaging systems, certain regions of interest in the input video must be magnified for more detailed analyses. However, it is difficult to obtain satisfactory images using conventional image zooming techniques and the interpolation methods. Expensive imaging devices capable of capturing images of higher resolution or higher quality may not be desirable for higher cost.

Nowadays, the super resolution (SR) algorithm has been considered one of the most promising methods to overcome the limits of imaging devices since it does not induce any additional expensive hardware. The SR algorithm is an image processing technique that can recover an HR image from multiple LR images.

Researchers have investigated a variety of SR approaches over the past last two decades in an attempt to achieve better image reconstruction results [1,2]. SR algorithms can be divided into two broad categories. The first is *motion-based* SR which considers movement between the LR image frames as a cue [3-9]. By making certain assumptions in the image acquisition model, this approach becomes straightforward and easy to implement. In this scheme, however, precise motion estimation and compensation are very important to reconstruct the HR image. Since the estimation of complex motions of multiple objects in LR video is difficult and time-consuming, new approaches have recently been developed to avoid the high dependency of motion-based SR on accurate motion estimation [10-14]. These approaches constitute the second category of SR algorithms and are referred to as *motion-free* SR [15]. Instead of directly estimating the motion, motion-free SR obtains spatial enhancement by incorporating cues such as blur.

Among the various motion-free SR approaches, the example-based SR algorithm [11] is one of the most promising methods. This method involves the concept

* Correspondence: hsko@korea.ac.kr
[1]School of Electrical Engineering, Korea University, Anam-dong, Seongbuk-Gu, Seoul 136-701, Korea
Full list of author information is available at the end of the article

of prior information to reconstruct HR image. They use learned data sets of image patches capturing the relationship between LR and HR images and find appropriate patches for estimating an HR image. However, because a large amount of training data is required to obtain a robust reconstruction results, example- or learning-based SR incurs an enormous computational load.

Daniel et al. [12] tried to handle this problem by combining the motion- based SR and example-based SR. Based on an assumption that patches in a single natural image tend to recur many times in an image, their approach uses LR/HR pairs of patches within and across the scales of a single image. However, the quality of the reconstructed image still depends on the accuracy of motion estimation when compensating motions of the patches. In addition, the desired LR/HR pairs of patches might be insufficient when the observed image is small or severely degraded. This makes it hard to apply their approach to practical applications such as video surveillance systems.

For the point of view of estimation criteria, SR algorithms may be divided into *static* and *dynamic* SR [8]. Static SR fuses multiple LR images to reconstruct a single HR image at a specific time point, while dynamic SR exploits the temporal evolution which reconstructs the HR image sequence. Dynamic SR requires relatively lower memory and numbers of computations than static SR, and is therefore regarded as being a more appropriate approach for real-time applications.

In this article, we propose a robust dynamic motion-based SR algorithm for LR video input. Our approach iteratively fuses the pixel data from an LR image sequence to estimate the pixel data of the HR image sequence based on the Kalman recursive estimation [8]. To deal with the performance degradation because of the inexplicit motion estimation, we suggest a validation process to filter out the irregularly registered pixels caused by inaccurate motion estimation. By implementing the proposed validation method, our SR approach was able to show robust HR image reconstruction results, even when the motion estimates were not accurate at the sub-pixel level. Moreover, abrupt changes in the scene input video can be detected in this validation process, so the fusion of pixels from two different scenes can be prevented. Since the quality of the reconstructed images is stable even with inaccurate motion estimation with low memory usage (requires only two frame memory) because of the sequential estimation, and each updated HR frame can be viewed during the estimation process, our approach is suitable for practical applications, especially in visual surveillance systems.

The remainder of this article is structured as follows. In Section 2, we describe the image acquisition modeling and basic concept of the dynamic SR process using the Kalman filter framework. In Section 3, we describe the proposed validation method for observed image data, and in Section 4 the scene change detection process developed for the robust sequential estimation of HR video has been described. In Section 5, we demonstrate both synthetic and real real-data experiments. Section 6 concludes this effort and discusses future study.

2. Dynamic SR

In this section, we review the dynamic SR approach proposed in [8], which is based on the Kalman recursive estimation. The main contribution of our approach will be described in Sections 3 and 4.

2.1. Image acquisition modeling

Among the many different image acquisition models, the following linear dynamic model is the most general and well represents the process of obtaining an LR image sequence:

$$\underline{X}(t) = M(t)\underline{X}(t-1) + \underline{U}(t), \tag{1}$$

$$\underline{Y}(t) = DB\underline{X}(t) + \underline{W}(t). \tag{2}$$

We used the underscore notation to indicate a vector derived from an image scanned in lexicographic order [8]. Thus, the HR frame at time t, $\underline{X}(t)$ with a size of $[r^2MN \times 1]$ is the warped version of the previous HR frame where r is the resolution-enhancement factor, since $M(t)$ with a size of $[r^2MN \times r^2MN]$, indicates the existing motions between the two neighboring frames. The $[r^2MN \times 1]$ vector, $\underline{U}(t)$, can be explained as the system noise that represents the accuracy of the motion estimation. In Equation 2, $\underline{Y}(t)$ with a size of $[MN \times 1]$ is the observed LR image at time t, and the $[r^2MN \times r^2MN]$ matrix, B, describes the blur operations resulting from the sensor's point spread function. The $[MN \times r^2MN]$ matrix, D, reflects the downsample operation in the image acquisition and saving. The $[MN \times 1]$ vector $\underline{W}(t)$ is the measurement noise.

To apply Kalman filtering for estimating $\underline{X}$from $\underline{Y}$, we constrain the model with the following assumptions:

(i) Only translational (planar) motion is considered in the input video.

(ii) The blur and downsampling operation are invariant in time. This is why there are no time indices in B and D.

(iii). Both the system and measurement noise are assumed to be additive white Gaussian noise.

By substituting $\underline{Z}(t) = B\underline{X}(t)$, we first estimate the *blurred version* of the HR image, $\underline{Z}(t)$, with a size of $[r^2MN \times 1]$ and then deblur it to obtain the final clear

HR image, $\underline{X}(t)$. The following two equations reflect the changes resulting from incorporating the blurred operation B to generate the measurement $\underline{Z}(t)$ into Equations 1 and 2, where the $[r^2MN \times 1]$ vector $\underline{V}(t)$ is the colored version of the measurement noise $\underline{U}(t)$:

$$\underline{Z}(t) = M(t)\underline{Z}(t-1) + \underline{V}(t), \tag{3}$$

$$\underline{Y}(t) = D\underline{Z}(t) + \underline{W}(t). \tag{4}$$

2.2. Kalman recursive for data fusion

Kalman filtering is the optimal method of estimating the dynamic state in linear modeling as described above [16]. The state to be estimated is the blurred HR image, i.e., $\underline{Z}(t)$. By means of the Kalman filtering theories [16,17], the *update* equations for the state vector and covariance matrix can be derived as follows:

$$\begin{aligned}\hat{\underline{Z}}(t) &= \underbrace{\hat{\underline{Z}}^M(t)}_{prediction} + \underbrace{K(t)}_{gain}\underbrace{[\underline{Y}(t) - D\hat{\underline{Z}}^M(t)]}_{innovation} \\ &= M(t)\hat{\underline{Z}}(t-1) + K(t)[\underline{Y}(t) - DM(t)\hat{\underline{Z}}(t-1)],\end{aligned} \tag{5}$$

$$\begin{aligned}\mathrm{Cov}(\hat{\underline{Z}}(t)) &= \underbrace{P(t)}_{prediction} - K(t)\underbrace{S(t)}_{innovation}K^T(t) \\ &= [\mathbf{I} - K(t)D]P(t),\end{aligned} \tag{6}$$

$$\begin{aligned}K(t) &= P(t)D^TS^{-1}(t) \\ &= P(t)D^T[DP(t)D^T + C_w(t)]^{-1},\end{aligned} \tag{7}$$

where $\hat{\underline{Z}}(t)$ denotes the estimated state vector, i.e., the blurred HR image. Equation 5 indicates that the final estimate of the blurred HR image is the sum of the *prediction* $\hat{\underline{Z}}^M(t)$ (i.e., motion compensated version of the previous estimate, $M(t)\hat{\underline{Z}}(t-1)$ and *innovation* or *measurement residual* (i.e., the difference between the new observation, $\underline{Y}(t)$, and prediction) multiplied by $K(t)$, which is the *Kalman gain* defined as the ratio of the *prediction covariance* $P(t)$ to the *innovation covariance* $S(t)$. Analogously, the updated covariance of $\hat{\underline{Z}}(t)$ can be derived as in Equation 6.

The procedures used to compute $P(t)$ and $S(t)$ are shown in Equations 8 and 9, respectively. The prediction covariance $P(t)$ in Equation 8 reflects the accuracy of the prediction for original HR image, $\hat{\underline{Z}}^M(t)$. The innovation covariance $S(t)$ in Equation 9 reflects the accuracy of prediction for an LR observation image, $D\hat{\underline{Z}}^M(t)$.

$$\begin{aligned}P(t) &= E\{[\underline{Z}(t) - \hat{\underline{Z}}^M(t)][\underline{Z}(t) - \hat{\underline{Z}}^M(t)]^T\} \\ &= M(t)\mathrm{Cov}(\hat{\underline{Z}}(t-1))M^T(t) + C_v(t),\end{aligned} \tag{8}$$

$$\begin{aligned}S(t) &= E\{[\underline{Y}(t) - D\hat{\underline{Z}}^M(t)][\underline{Y}(t) - D\hat{\underline{Z}}^M(t)]^T\} \\ &= DP(t)D^T + C_w(t).\end{aligned} \tag{9}$$

Since the inversion of the covariance matrix in Equation 7 is very cumbersome and requires substantial computation and memory, further assumptions are needed to achieve a faster implementation. As proven in [8], if the covariance matrix of $\underline{V}(t)$ denoted as $C_v(t)$ and the initial covariance $\mathrm{Cov}(\hat{\underline{Z}}(0))$ are *diagonal*, $P(t)$ and $\mathrm{Cov}(\hat{\underline{Z}}(t))$ become diagonal for all t. This enables a pixel-by-pixel implementation, so all of the procedures from Equations 1 to 9 can be computed as a single scalar value (i.e., single pixel). A more detailed description can be found in [8].

Once the covariances of the noise components $C_w(t)$, $C_v(t)$, and $\mathrm{Cov}(\hat{\underline{Z}}(0))$ are initialized at time $t = 0$, they are used to calculate $P(t)$, $S(t)$, and $K(t)$. After $K(t)$ is calculated, the estimation of the HR image $\hat{\underline{Z}}(t)$ and its covariance $\mathrm{Cov}(\hat{\underline{Z}}(t))$ is calculated recursively by the Kalman filter *update* equations in Equations 5 and 6. Since all of the covariance matrices are diagonal, we can convert them into general image matrices (not lexicographic ordered) to compute the Kalman gain on a pixel-by-pixel basis. The graphical procedures of Equations 7-9 are illustrated in Figure 1. The additions, multiplication, and inversion in Figure 1 are element-wise operations. Only MN elements of $K(t)$ have non-zero values, because of the up-sampling (zero-filling) of the innovation covariance, $S(t)$. This means that only MN pixels are updated in Equation 5 when the new input image frame $\underline{Y}(t)$ is measured.

To estimate and compensate the motions existing among the input frames modeled by $\underline{M}(t)$, we adopt the image registration method in frequency-domain [18] since it is simple and accurate for translational motions. It estimates the horizontal and vertical shifts in spatial domain by computing the phase shift in the frequency domain. Moreover, the frequency-domain approach benefits when the aliasing effect exists in input LR frames.

To handle color video input, we apply the same Kalman filtering process to each RGB channel. Once the blurred HR image, $\hat{\underline{Z}}(t)$, is estimated, the final clear HR image, $\hat{\underline{X}}(t)$, is reconstructed by the deblurring method. The flow chart of the conventional dynamic SR algorithm is illustrated in Figure 2.

3. Measurement validation

Explicit motion estimation is a major factor that affects the performance of the motion-based SR algorithm as mentioned in [13,14]. Various research efforts have been dedicated to enable precise (sub-pixel accuracy) motion estimation; however, the methods developed are

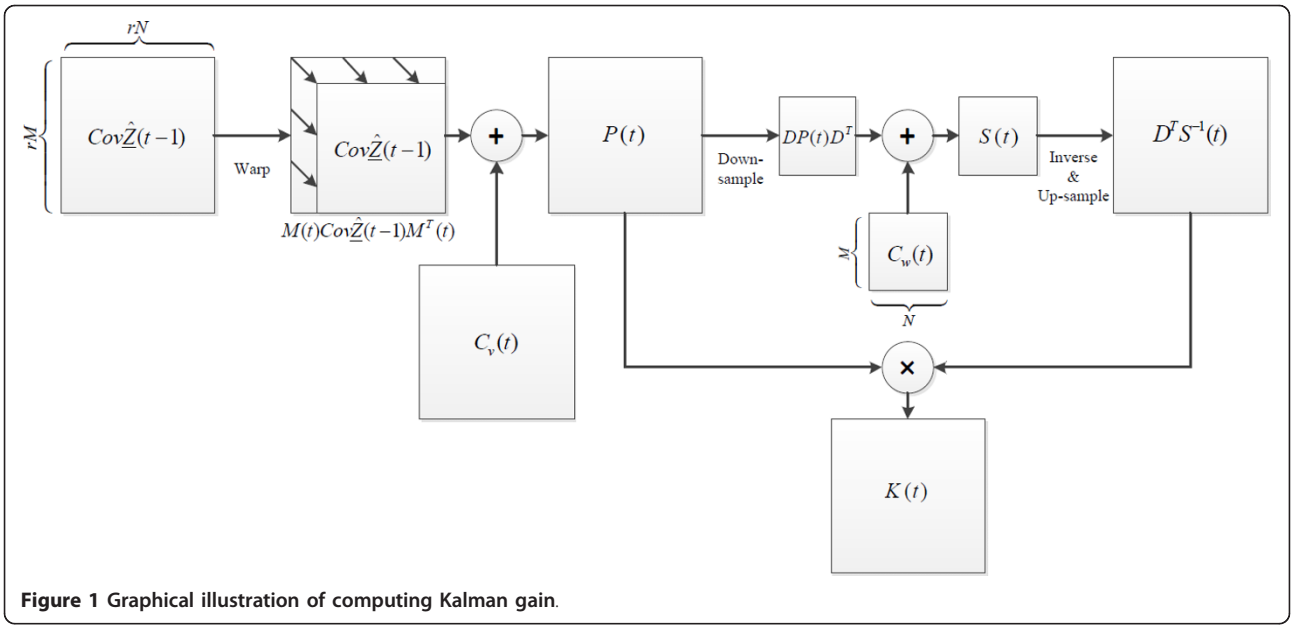

Figure 1 Graphical illustration of computing Kalman gain.

insufficient to guarantee perfect motion compensation and, even though perfect motion estimation is potentially possible, it usually requires a large amount of computation.

Some novel approaches not involving accurate motion estimation were recently suggested in [10-14], but they are not suitable for practical real-time surveillance system applications because of their computation requirements. In this article, we added a validation method in the sequential estimation process to enhance erroneous reconstructed HR images caused by inexplicit motion estimations.

When the motion estimation result is inaccurate (i.e., the reference and target frames are misaligned), the difference in the pixel intensity between the two corresponding frames will be increased as depicted in Figure 3. With the dynamic linear modeling described in Section 2, this difference in the pixel intensity can be represented by the distance in Equation 10:

$$d^2(t) = [\underline{Y}(t) - D\hat{\underline{Z}}^M(t)]^{\mathrm{T}} S^{-1}(t)[\underline{Y}(t) - D\hat{\underline{Z}}^M(t)]. \quad (10)$$

$$d^2(t) = \sum_{k=1}^{MN} d_k^2, \quad (11)$$

where $d_k^2(t) = [y_k(t) - D_k\hat{\underline{Z}}^M(t)]^{\mathrm{T}} S_k^{-1}(t)[y_k(t) - D_k\hat{\underline{Z}}^M(t)]$.

Since we assume that all covariance matrices including $S(t)$ are diagonal, computing the distance of one measured frame at time t, $d(t)$ which is referred to as the *'Mahalanobis distance'* or *'Statistical distance'*, is the same as computing the sum of the distances of each pixel in that frame, $d_k(t)$, in Equation 11. $y_k(t)$ is the kth pixel in a measured frame $\underline{Y}(t)$ and $S_k(t)$ is the kth diagonal element of $S(t)$. D_k is the kth row of the downsampling operator D size of $[1 \times r^2MN]$.

When the Kalman filter has at least been initialized and the state vector is being estimated, the true observation at time t, given the measurements $\underline{Y}^{t-1} = \{\underline{Y}(1), ..., \underline{Y}(t\text{-}1)\}$,, is normally distributed.

$$p[\underline{Y}(t)|\underline{Y}^{t-1}] = N[D\hat{\underline{Z}}^M(t), S(t)]. \quad (12)$$

$\underline{Y}(t)$ in Equation 12 is the measurement at time t and $\underline{Y}^{t-1}$ is the sequence of measurements from the initial time to time t - 1. Thus, Equation 12 represents that the conditional probability of $\underline{Y}(t)$ given the measurements up to time t - 1, namely $\underline{Y}^{t-1}$ is normally distributed with the mean equal to the predicted measurement $D\hat{\underline{Z}}^M(t)$ and the covariance equal to the innovation covariance $S(t)$. The theoretical description for this can be found in the sections on the Kalman filter in [16,17].

In the proposed SR algorithm, we attempt to detect any *'misalignment'* at the pixel level but not at the frame level, meaning that we want to exclude only those pixels that are misaligned in the measured frame, not all of the pixels in the measured frame that are misaligned. By incorporating the concept from [17] and from the ideas of the validation methods or data association for target tracking field in [19,20], we may define a *validation region* $V(\gamma)$ for a measured pixel as in Equation 13:

$$V(\gamma) = \{y_k(t) : d_k^2(t) \le \gamma\}, \quad k = 1, 2, ..., MN. \quad (13)$$

By fixing the threshold γ at all times for every pixel, the validation region $V(\gamma)$ is dependent only on the

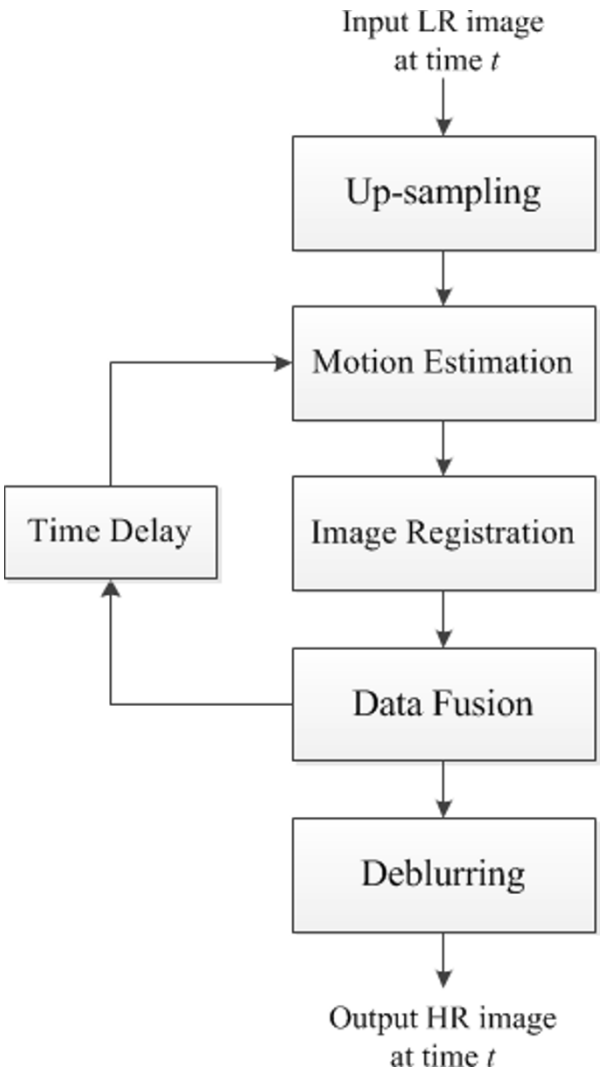

Figure 2 Flow chart of conventional dynamic SR algorithm.

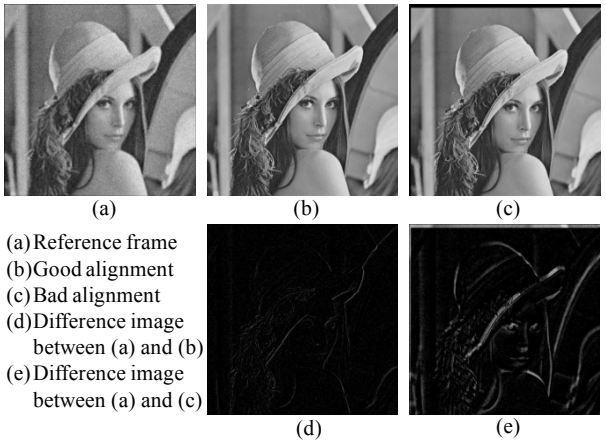

Figure 3 Pixel intensity difference increases when misalignment occurs.

threshold γ, but not on the time t or pixel index k. Whenever the pixel data from the input LR image at each time instant (i.e., $y_k(t)$ for all k) are observed, we compute each distance $d_k(t)$ in Equation 11 and filter out the pixels falling out of the region in Equation 13. In other words, only those pixels whose distance is below the threshold are considered *valid*. So, this procedure regards the pixels that lie outside of the validation region as outliers, i.e., misaligned, hence they are excluded from the data fusion process. This is the so-called *'Measurement Validation'* method and it is applied right before the pixel data fusion process in Equations 5 and 6 in our SR approach illustrated in Figure 4.

As represented in Equations 5 and 6, $K(t)$ determines the amount of updates required for estimating $\hat{\underline{Z}}(t)$ and $\mathrm{Cov}(\hat{\underline{Z}}(t))$. In the proposed measurement validation method, only valid pixel values should be used in the update equations. When $K(t)$ is equal to zero, no updates will be made in Equations 5 and 6, thus the estimations for $\hat{\underline{Z}}(t)$ and $\mathrm{Cov}(\hat{\underline{Z}}(t))$ are only dependent on the prediction terms. In our implementation, after the new measurement is obtained, i.e., MN pixels are observed at time t, each pixel is investigated to determine whether or not it falls inside the validation region in Equation 13. After we determine the misaligned pixels among MN pixels, we can prevent them from being used in the update equations by setting those elements

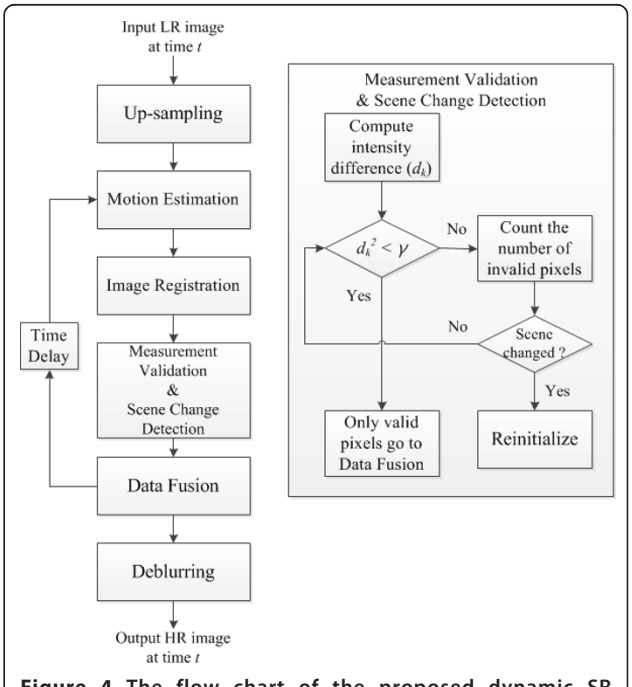

Figure 4 The flow chart of the proposed dynamic SR algorithm.

of $K(t)$, whose indices correspond to the indices of misaligned pixels, to zero.

Under the Gaussian assumption, the validation region $V(\gamma)$ is *chi-square* distributed with the number of degrees of freedom equal to the dimension of the measurement. The chi-square distribution table gives the probability mass:

$$P(\gamma) = p\{y_k(t) \in V(\gamma)\}, \quad k = 1, 2, ..., MN. \tag{14}$$

$P(\gamma)$ is the probability that the measurement will fall inside the validation region for various values of γ and dimensions of $y_k(t)$. Since the degree of freedom (DoF) for a single pixel is one, we can select the threshold γ in Table 1. Therefore, we can control the range of the valid region by varying the threshold value, γ, obtained from the chi-square table for the desired confidence level [17]. For example, if we set γ to 2.71,[a] the probability that the measurement falls inside of the validation region will be 90%. In the proposed method, the threshold is set to 15.1 which means that there is a 99.99% chance that $d_k^2(t)$ will be less than or equal to 15.1. So, the threshold value is not directly related to the image dynamic range, but to the range of the statistical distance of the image pixel. The bigger the threshold that is selected, the wider the validation region. In other words, the probability that the measured pixels are determined as misaligned will decrease as the threshold becomes larger.

4. Scene change detection

Since the dynamic SR algorithm recursively fuses the pixel data from the sequentially observed images, it is highly likely for an erroneous HR estimation result to occur when the scene or contents of two adjacent frames are totally different. This problem arises frequently when the input LR video contains many different scenes or the motions in it are too large to be estimated. There is no possible motion between different frames from different scenes and, hence, these frames can never be aligned correctly. Even though the measurement validation method can detect and filter out misaligned pixels, fusing pixels from two different scenes is not a desired situation.

Instead of applying one of the conventional scene change detection methods [21,22], we suggest a simple but effective way to detect a sudden change of scene in the input LR video by exploiting the statistical distance already discussed in the previous section.

Table 1 Chi-square distribution table

DoF	$P = 0.9$	$P = 0.99$	$P = 0.999$	$P = 0.9999$
1	2.71	6.63	10.8	15.1
2	4.61	9.21	13.8	18.4
10	16.0	23.2	29.6	35.6
100	118	136	149	161

The proposed method detects abrupt scene changes between adjacent frames by computing the proportion of invalid pixels with respect to the total number of pixels in the observed LR frame of size $[M \times N]$:

$$\frac{1}{MN}\sum_{k=1}^{MN} I(d_k(t)) \geq Th, \text{ where } I(d_k(t)) = \begin{cases} 1 & \text{if } d_k^2(t) > \gamma. \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

In this article, we set the threshold value, Th to 0.3, which means that about 30% of the pixels from the current input LR frame are different from those of the previous frame. This threshold value is determined experimentally with more than ten real video data containing scene changes. If a sudden scene change is detected with this method, we reset the estimation process (i.e., reinitialize the Kalman filter). The procedure is summarized in Figure 4.

5. Experimental results

We evaluated the performance of the proposed dynamic SR algorithm with synthetic and real video data. The threshold for measurement validation was set to 15.1 for all experiments, which represents that a confidence probability of 99.99% according to the chi-square distribution table. For the deblurring method in the last step of the proposed SR algorithm, we used the classical but effective *Wiener filter* approach with a constant noise-to-signal ratio (NSR) to reduce the computation complexity. The parameter NSR for the Wiener filter was tuned to obtain the best performance in all experiments.

5.1. Synthetic video data test

In this experiment, we tested the proposed algorithm with synthetic LR video data. We generated LR color videos by simulating the image acquisition procedure described in Section 2.1. The test video in Figure 5 was downloaded from the website of the author in [8][b] and the test videos in Figures 6 and 7 were captured by a commercial surveillance camera, SHC-730N, courtesy of Samsung Techwin Co., Ltd., Korea. We downsampled the original videos by a factor of two after blurring them with a 3×3 Gaussian kernel whose variance was equal to 1. Finally, we generated LR videos by adding Gaussian noise to achieve its signal-to-noise ratio (SNR) of 30 dB. The size of all three LR videos was 160×120 and they contained only global translational motions. The test LR videos are super-resolved by a factor of two through the proposed algorithm and the method in [8].

The method in [8] was implemented directly from the MATLAB GUI (http://users.soe.ucsc.edu/~milanfar/software/superresolution.html). According to [8], they used

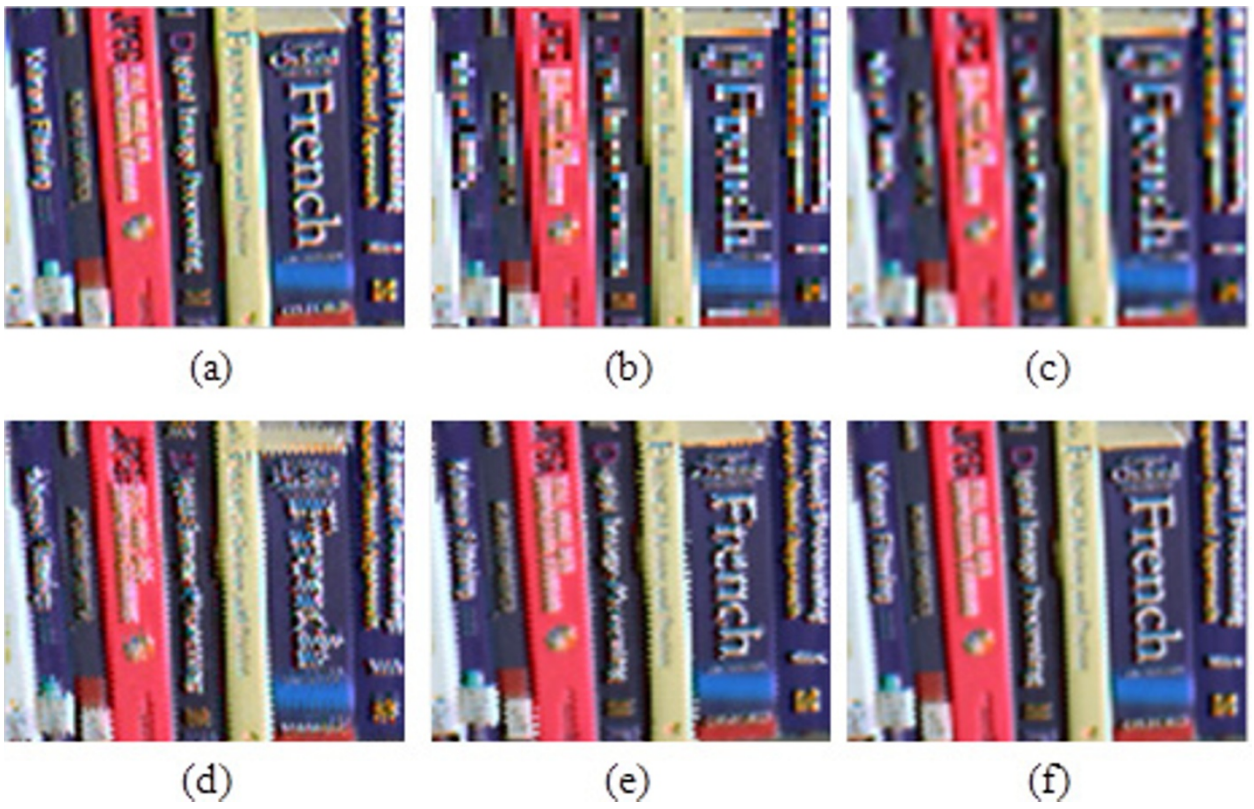

Figure 5 The synthetic webcam video data result: **(a)** Original frame. **(b)** LR frame. **(c)** Bicubic interpolated frame. **(d)** Reconstructed HR frames by applying the method in [8] with the image registration algorithm in [23]. **(e)** Reconstructed HR frames by applying the method in [8] with the image registration algorithm in [18]. **(f)** Reconstructed HR frames by applying the proposed method.

the image registration algorithm in [23] which is different from the algorithm we exploited. As mentioned in the earlier sections and previous related studies, the major factor contributing to the reconstruction image result of the multi-frame SR algorithm is the accuracy of the image registration. Thus, if a different image registration algorithm is used in the reference method, we cannot say that the improved HR image result is completely because of the proposed measurement validation. For a fair comparison, we also implemented the method in [8] using the frequency-domain image registration algorithm [18] which is used in the proposed method. Therefore, we compared the proposed method with two reference methods, one from the author's website and the other from our own implementation by modifying the image registration part. In addition, we applied the Wiener filter to the method in [8], instead of the bilateral-total variation (BTV) regularization to see the effect of the measurement validation only. The quality of the reconstructed HR image is evaluated quantitatively with the $PSNR^c$ (Peak SNR) metric.

We enlarged the 100 × 80 sections of the original, simulated LR, bicubic interpolated, and reconstructed video frames for better visual quality evaluation. The images in Figure 5 are the 90th frames and the images in Figure 6 are the 60th frames of each input video. In the reconstructed HR frames in Figures 5 and 6, there are some artifacts because of the motion estimation error, such as periodic teeth along horizontal and vertical lines or stair-case phenomena along diagonal lines. The motion estimation error may become large when the size of an image is too small, or the motion is too large. Because the only difference between the methods in Figure 5d,e is the image registration algorithm, the slightly better quality of Figure 5e can be attributed to the better performance of the algorithm in [18]. As shown in Figures 5f and 6f, the image quality of the HR result with the proposed method is enhanced more than the results in Figures 5e and 6e. The corresponding PSNR values are listed in Table 2. When compared to the results obtained with the method in [8], the jaggedness of the edges and corners is substantially reduced. Even though the same image registration algorithm was used for the results in Figure 5e,f, the result obtained with the proposed method is visually superior. This demonstrates the effectiveness of the proposed

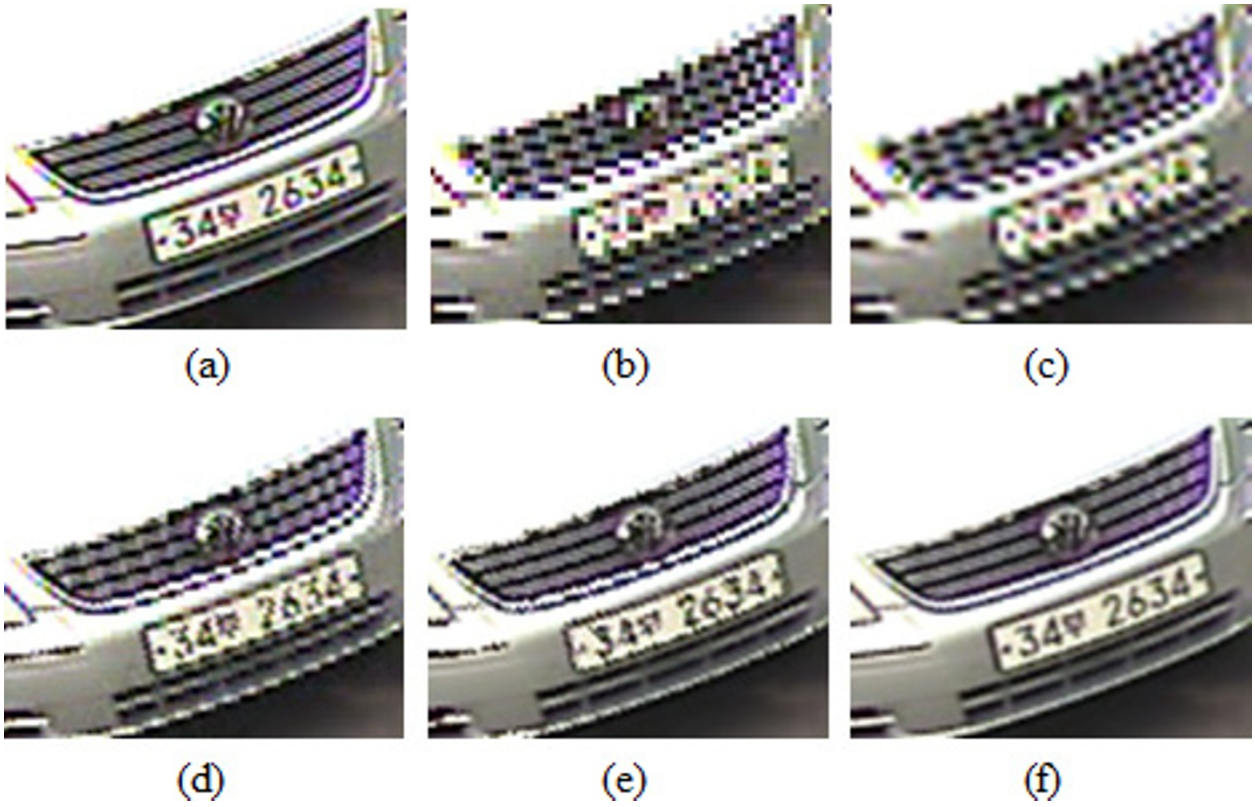

Figure 6 The synthetic surveillance video data result: **(a)** Original frame. **(b)** LR frame. **(c)** Bicubic interpolated frame. **(d)** Reconstructed HR frames by applying the method in [8] with the image registration algorithm in [23]. **(e)** Reconstructed HR frames by applying the method in [8] with the image registration algorithm in [18]. **(f)** Reconstructed HR frames by applying the proposed method.

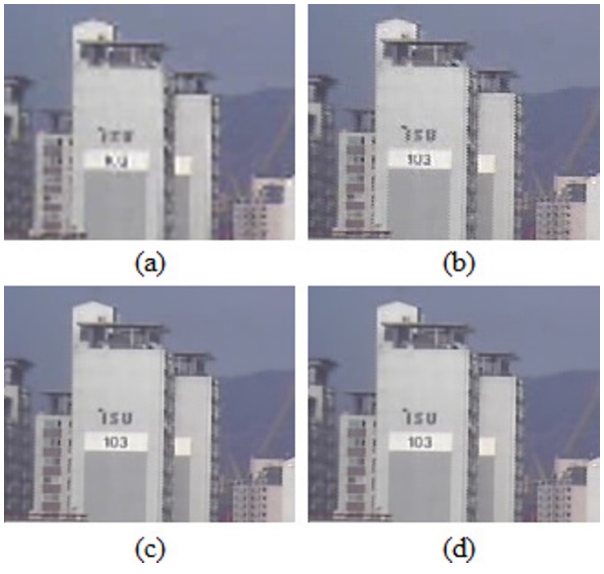

Figure 7 The synthetic video data result: **(a)** Bicubic interpolated frame. **(b)** Reconstructed 90th HR frame using the method in [8,23]. **(c)** Reconstructed 90th HR frame using the method in [8,18]. **(d)** Reconstructed 90th HR frame using the proposed method. **The PSNR are 19.91, 21.09, 23.94, and 24.02 dB, respectively**.

measurement validation method. Analogously, the same analysis can be applied to the results in Figure 6.

In the experiment corresponding to the results in Figure 7, we enhanced the spatial resolution of the LR video by a factor of two. In Figure 7, only 160 × 130 zoomed sections of the results are depicted. There is little difference in performance between the results obtained with and without the measurement validation (Figure 7c,d, respectively) because the image registration was quite accurate. To test the performance of the measurement validation, we intentionally added alignment errors to the aligned LR frames beyond the 60th frame. The HR image at the 90th frame without the measurement validation in Figure 8a was significantly degraded because of the registration errors. On the contrary, the resulting HR image obtained with the measurement validation was less affected by the registration errors as shown in Figure 8b. In Figure 8c, one can see that the number of misaligned pixels determined by the threshold in Equation 13 increases after the 60th frame. This tells us that the measurement validation method becomes more effective when a large amount of image registration errors occurs.

Table 2 PSNR of experiment in Figures 5 and 6.

Output size	Bicubic interpolation	Farsiu [8] + [21](without MV)	Farsiu [8] + [17](without MV)	Proposed (with MV)
320 × 240	5(c), 19.44 dB	5(d), 19.33 dB	5(e), 18.84 dB	5(f), 23.95 dB
	6(c), 19.79 dB	6(d), 20.98 dB	6(e), 21.56 dB	6(f), 24.73 dB

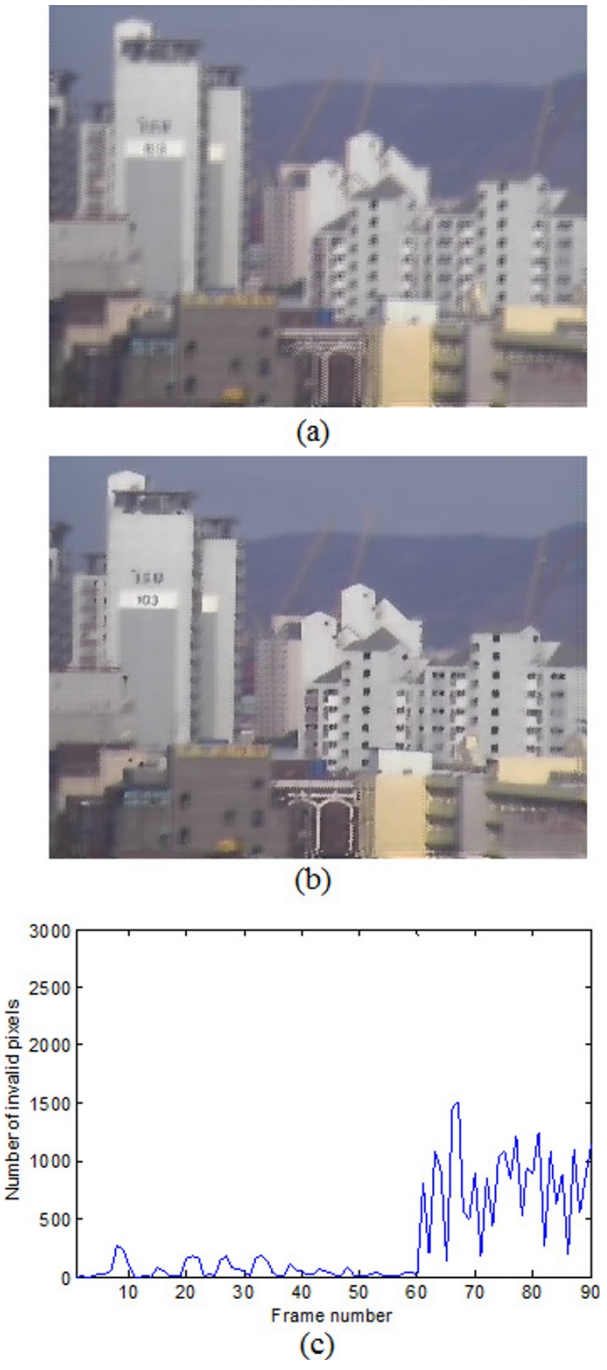

Figure 8 The synthetic video data result: **(a)** Reconstructed 90th HR frame using the method without measurement validation. **(b)** Reconstructed 90th HR frame using the method with measurement validation. **(c)** The number of misaligned pixels for each frame. **We artificially added registration errors from the 60 to 90th frames**.

5.2. Real video data test

In the next experiment, our algorithm is evaluated with real video data captured by a surveillance camera, courtesy of Adyoron Intelligent Systems Ltd., Tel Aviv, Israel. We increased the spatial resolution of the real LR video by a factor of two in the vertical and horizontal directions. The input size of the video frame was 138 × 115 and, therefore, the resulting size of the reconstructed video frame is 276 × 230, as shown in Figure 9. Figure 9d demonstrates the superior performance of the proposed algorithm compared to the conventional methods in Figure 9b,c. Especially, the jagged edges because of the wrong translational motion estimation are clearly reduced in Figure 9c. This is the contribution of the measurement validation process.

In the case of a small input size, the effect of filtering misaligned pixels becomes more remarkable, as shown in the experimental results of Figure 10. In general, precise motion estimation is more difficult when the input image is small, since the number of pixels, i.e., *features* or *information* is insufficient to achieve a good alignment. The visual quality of the results without the measurement validation in Figure 10c,g is worse than the Bicubic interpolated results in Figure 10b,f.

Assuming that a sufficient number of LR frames are available and the proper image registration algorithm is used for compensating the motions existing among the LR frames, multi-frame SR generally outperforms the single image interpolation method. In the extreme case where we do not register the LR frames at all, the estimated HR image result will be worse than the Bicubic interpolation result. However, if we apply the measurement validation while still not registering all LR frames, the HR image result will be almost the same as the initial estimated HR image since most of the unregistered LR pixels will be regarded as *invalid*. Thus, if we set the initial estimated HR image as the Bicubic interpolated one of the initial LR frames, the HR image result obtained with the proposed method cannot be worse than the Bicubic interpolation result even though most of the LR data are excluded.

If all of the frames are aligned perfectly or well enough to fall in the preset validation region, all of the measured pixel values will contribute to the HR image estimation process. The benefit of the measurement validation process is that it prevents the misaligned pixel values from contributing to the HR image estimation. By setting the confidence level for the image

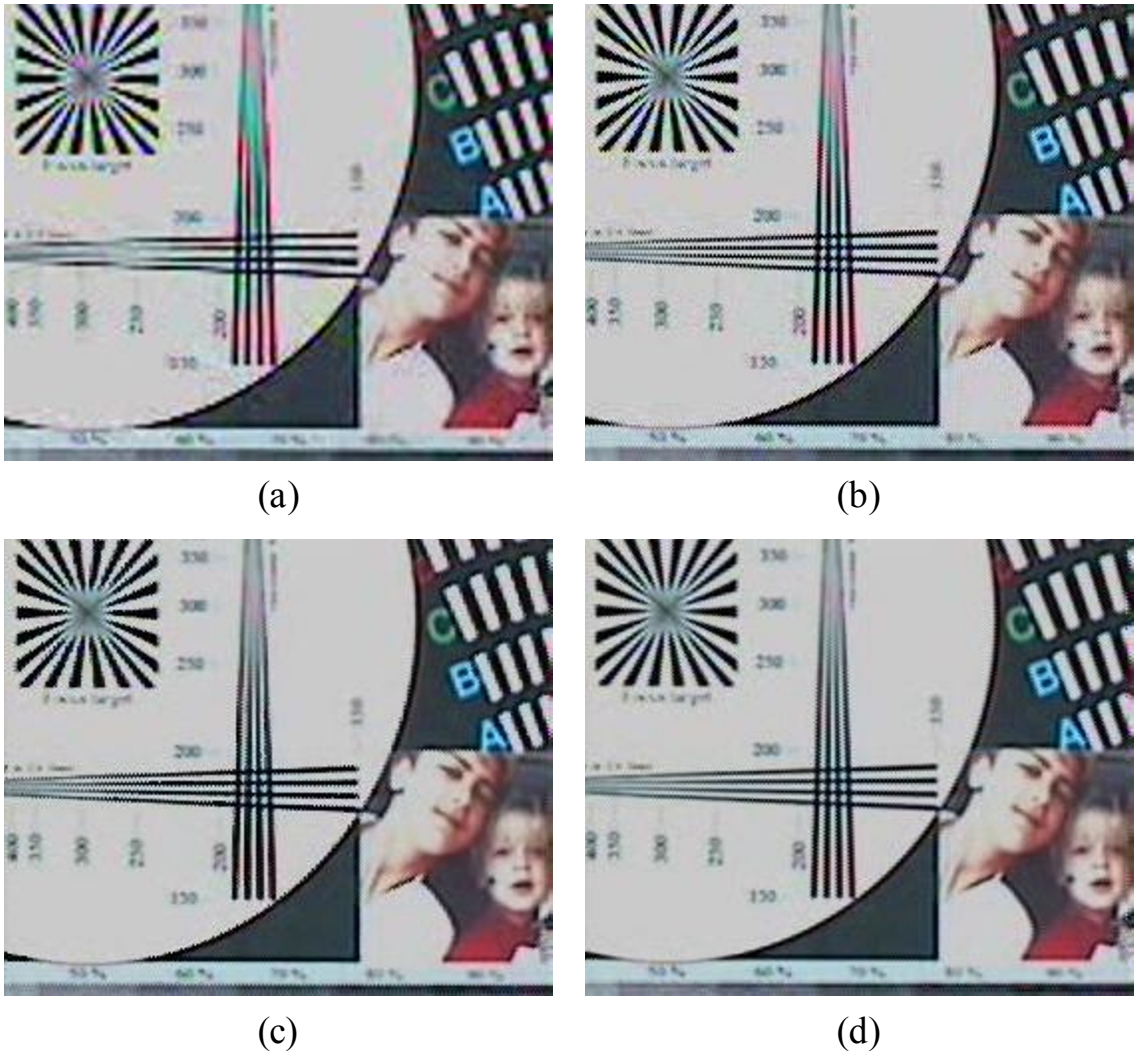

Figure 9 Real video data result: **(a)** Bicubic interpolated frame. **(b)** Reconstructed 40th HR frame using the method in [8,23]. **(c)** Reconstructed 40th HR frame using the method in [8,18]. **(d)** Reconstructed 40th HR frame using the proposed method. Note that the artifact because of misalignment around the edges are effectively removed in **(d)**.

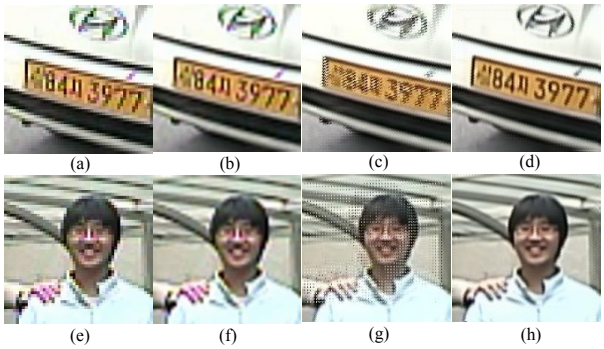

Figure 10 Small size real video data result: **(a, e)** 90th LR frames with sizess of 50 × 50. **(b, f)** Bicubic interpolated frames. **(c, g)** Super-resolved by a factor of four with the methods in [8,23]. **(d, h)** Reconstructed frames using the proposed method.

registration result (i.e., the threshold for the validation region), we can exclude undesired updates of the pixel values. Thus, it becomes more beneficial when there is a higher possibility of misalignment because of the poor performance of the image registration algorithm or because of the existence of LR frames with fast motion. This is the reason why the results obtained with the proposed method in Figure 10d,h show more robust performance when large motion estimation errors occur frequently.

5.3. Scene change detection performance test

In this experiment, we evaluate the proposed scene change detection method. We created LR videos containing four different scenes. The input size is 50 × 50 and the spatial resolution ratio was increased by a factor

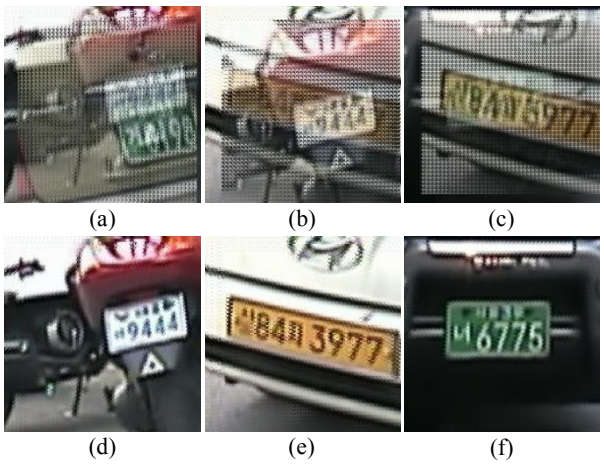

Figure 11 Effect of scene change detection method: **(a-c)** 120, 270, and 390th reconstructed frames without the scene change detection method, respectively. **(d-f)** Well-reconstructed frames exploiting proposed scene change detection method.

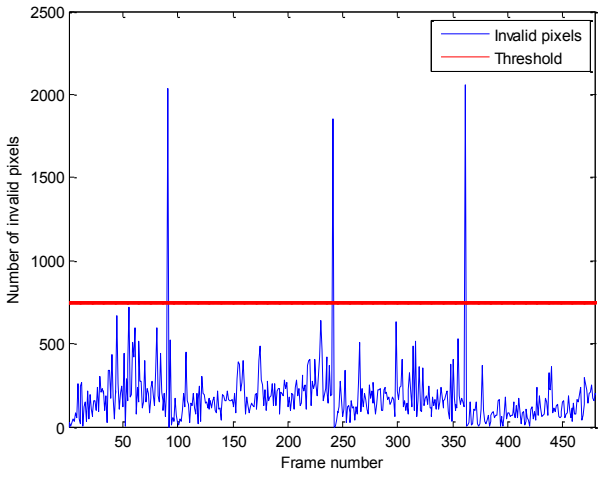

Figure 12 Invalid pixels of each input frame. Scene changes occurr at the 90, 241, and 361th frames.

of four. The upper images in Figure 11 are 30 frames after the scene change occurred without using the proposed scene change detection method. When the incoming frames are different from the previously reconstructed frame, two different scenes are overlapped with each other. This artifact can easily be addressed by resetting the SR process when the current input frame belongs to a different scene. In this case, most of the pixels from the changed scene will be considered as invalid ones by the proposed measurement validation method. Consequently, the ratio of invalid pixels will cross the preset threshold with high probability.

The lower images in Figure 11 are the reconstructed frames at the same time instant as the upper ones. The scene change was detected when the ratio of invalid pixels is below the threshold value, Th, in Equation 15 which was set to 0.3; hence, the SR process was reinitialized. The artifact in Figure 11a-c is eliminated by fusing the pixel data from the same scene. In Figure 12, the blue line represents the number of invalid pixels for the input video frames and the red line is the preset threshold. The scene changes abruptly three times at frames #91, #241, and #361.

6. Conclusions

In this article, we proposed a robust dynamic SR algorithm to alleviate the performance degradation because of inaccurate motion estimation and sudden scene changes. We adopted the dynamic SR algorithm based on the Kalman filter approach, because of its effectiveness when applied to real-time applications. When the size of the output super-resolved image is about 200 × 200, the proposed dynamic SR algorithm estimates the HR images sequentially at a speed of over 20 fps while necessitating a memory size corresponding to only two frames.

In the case of misalignment caused by motion estimation error, the proposed measurement validation method determines whether each of the pixels is suitable for data fusion or not with the statistical distance of intensity. It is preferable to set the pixels with a large distance as invalid and filter them out after the estimation process enters the steady state. Otherwise, the estimated HR pixels tend to remain the same as the previous LR pixel since every input pixel with a large intensity difference would be filtered out and, hence, the update process in Kalman filtering would be prevented. The starting point of the measurement validation and the appropriate threshold remain as an ongoing research topic.

In addition, we developed a scene change detection method to handle various input videos containing one or more scene changes. By virtue of the proposed scene change detection method, we can handle input video containing more than one scene. Adaptive threshold setting for the scene change detection method is preferable for robust detection performance, and so this remains as a future study. Throughout this study, we fixed, defined a relatively large validation region, $V(\gamma)$, whose threshold is equal to 15.1, because we assumed that the image registration algorithm performs well enough to align most of the LR frames correctly. If we can predict the accuracy of image registration, we can control the validation region by varying the threshold, γ.

As shown in the several representative experiments, a considerable degree of enhancement and the restoration of the deteriorated visual information can be achieved by the proposed SR algorithm. Especially, for input

images of small size, such as human face and license plate images, the proposed SR algorithm is appropriate for real-time visual surveillance applications considering the processing speed and the visual quality of the reconstructed image.

Endnotes

[a]The threshold value has no digital unit since $d(t)$ is a normalized random variable (i.e., statistical distance). [b]The test video can be downloaded from http://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html. [c]The PSNR of two images $\underline{X}$and $\underline{Y}$of size M by N is defined as $\mathrm{PSNR(dB)} = 10\log_{10}((255^2 \times MN)/\left\|\underline{X} - \underline{Y}\right\|_2^2)$.

Acknowledgements
This research was supported by the Seoul R&BD Program (WR080951).

Author details
[1]School of Electrical Engineering, Korea University, Anam-dong, Seongbuk-Gu, Seoul 136-701, Korea [2]Office of Naval Research, Arlington, VA, USA

Competing interests
The authors declare that they have no competing interests.

Received: 28 February 2011 Accepted: 15 November 2011
Published: 15 November 2011

References

1. SC Park, MK Park, MG Kang, Super-resolution image reconstruction: a technical overview. IEEE Signal Proc. Mag. **20**(3), 21–36 (2003). doi:10.1109/MSP.2003.1203207
2. M Elad, A Feuer, Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. IEEE Trans. Image Process. **6**(12), 1646–1658 (1997). doi:10.1109/83.650118
3. M Elad, Y Hel-Or, A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur. IEEE Trans. Image Process. **10**(8), 1187–1193 (2001). doi:10.1109/83.935034
4. S Farsiu, D Robinson, M Elad, P Milanfar, Fast and robust multiframe super resolution. IEEE Trans. Image Process. **13**(10), 1327–1344 (2004). doi:10.1109/TIP.2004.834669
5. M Elad, A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur. IEEE Trans. Image Process. **10**(8), 1187–1193 (2001). doi:10.1109/83.935034
6. S Farsiu, M Elad, P Milanfar, Multiframe demosaicing and super-resolution of color images. IEEE Trans. Image Process. **15**(1), 141–159 (2006)
7. M Elad, A Feuer, Super-resolution reconstruction of image sequences. IEEE Trans. Pattern Anal. Mach. Intell. **21**(9), 817–834 (1999). doi:10.1109/34.790425
8. S Farsiu, M Elad, P Milanfar, Video-to-video dynamic super-resolution for grayscale and color sequences. EURASIP J. Appl. Signal Process, 1–15 (2006). Article ID 61859
9. B Narayanan, RC Hardie, KE Barner, M Shao, A computationally efficient super-resolution algorithm for video processing using partition filters. IEEE Trans. Circuits Syst. Video Technol. **17**(5), 621–634 (2007)
10. M Protter, M Elad, H Takeda, P Milanfar, Generalizing the nonlocal-means to super-resolution reconstruction. IEEE Trans. Image Process. **18**(1), 36–51 (2009)
11. W Freeman, T Jones, E Pasztor, Example-based super-resolution. Comput. Graph. Appl. **22**(2), 56–65 (2002). doi:10.1109/38.988747
12. D Glasner, S Bagon, M Irani, Super-resolution from a single image, in *International Conference on Computer Vision (ICCV)* (2009)
13. M Protter, M Elad, Super resolution with probabilistic motion estimation. IEEE Trans. Image Process. **18**(8), 1899–1904 (2009)
14. H Takeda, P Milanfar, M Protter, M Elad, Super-resolution without explicit subpixel motion estimation. IEEE Trans. Image Process. **18**(9), 1958–1975 (2009)
15. S Chaudhuri, J Manjunath, *Motion-free Super-Resolution* (Springer, 2005)
16. L Louis, *Statistical Signal Processing* (Scharf, Addison-Wesley Pub. Co, 1991)
17. Y Bar-Shalom, TE Fortmann, *Tracking and Data Association* (Academic Press, Inc, 1988)
18. P Vandewalle, S Susstrunk, M Vetterli, A frequency domain approach to registration of aliased images with application to super-resolution. EURASIP J. Appl. Signal Process, 1–14 (2006). Article ID 71459
19. BH Ku, YH Lee, WY Hong, H Ko, Suppressing ghost targets via gating and tracking history in Y-shaped passive linear array sonars. IEEE Trans. AES. **47**(3), 1605–1616 (2011)
20. H Ko, IK Lee, JH Lee, D Han, Effective multi-vehicle tracking in nighttime condition using imaging sensors. IEICE Trans-Inform. Syst. **E86-D**(9), 1887–1895 (2003)
21. E El-Qawasmeh, Scene change detection schemes for video indexing in uncompressed domain. Informatica **14**(1), 19–36 (2003)
22. C Ngo, T Pong, R Chin, H Zhang, Motion-based video representation for scene change detection. Int. J. Comput. Vis. **50**(2), 127–142 (2002). doi:10.1023/A:1020341931699
23. JR Bergen, P Anandan, KJ Hanna, R Hingorani, Hierarchical model-based motion estimation, in *Proceedings of European Conference on Computer Vision (ECCV '92)*, Santa Margherita Ligure, Italy 237-252 (1992)

doi:10.1186/1687-6180-2011-103
Cite this article as: Kim *et al.*: Robust video super resolution algorithm using measurement validation method and scene change detection. *EURASIP Journal on Advances in Signal Processing* 2011 **2011**:103.