

Time Invariant Steady-State Kalman Filter for Image Super-Resolution

C. Newland^{1,2}, and D. Gray^{1,2}

¹ CRC for Sensor, Signal and Information Processing (CSSIP).

²Department of Electrical and Electronic Engineering, University of Adelaide, SA 5005, Australia.

Email: {cnewland,dgray}@eleceng.adelaide.edu.au

Abstract

The Kalman filter is usually considered to be too computationally infeasible for image super-resolution due to the size of the images and thus the state space involved. This paper presents a computationally feasible, simplified time invariant steady-state Kalman filter approach to image super-resolution for the case of constant translational camera motion between low-resolution frames. Quantitative and aesthetic comparisons are made to the maximum *a-posteriori* (MAP) super-resolution approach, considered to be one of the standard super-resolution techniques, to demonstrate the Kalman filter's superiority despite the steady-state assumption.

Keywords: image super-resolution, time invariant steady-state Kalman filter, MAP estimator

1 Introduction

Image super-resolution is the process of estimating a high-resolution image from a sequence of noisy low-resolution image frames taken of the same scene but from marginally different perspectives. The differences in the camera position ensure that the scene has been sampled differently in each image causing the pixel values to differ between the image frames of the low-resolution sequence. Image super-resolution utilizes these differences to build a higher resolution composite image using all of the information contained within the individual low-resolution frames. Maximum *a-posteriori* (MAP) and projection onto convex sets (POCS) algorithms have been the dominant approaches to image super-resolution during the 1990's [1], with recent research focusing mostly on 'fast and robust' algorithms [2].

Kalman filters are widely acknowledged for creating the optimal mean square error estimate in the context of linear constraints. However in the field of image super-resolution they are equally recognised as being computationally unfeasible because the update of the Kalman gain matrix requires a large matrix inversion. Previously published approximations to the Kalman filter have included: recursive steepest-descent (R-SD) and recursive least-squares (R-LMS) algorithms that avoid the matrix inversion by approximating the Kalman filter itself [3, 4]; and reduced update Kalman filters that approximate the state-space into smaller regions to minimise the processing required for the matrix inversion [5, 6].

This paper presents a novel super-resolution Kalman filter approximation that uses a time-invariant steady-state error covariance matrix to negate the need for a matrix inversion with each update of the Kalman filter. Unlike previous approximations, this approach preserves both the structure of the Kalman filter and the complete state size, yet presents a computationally feasible algorithm. Future applications may include scanning devices and medical imaging systems with inbuilt motion control, or for use on unmanned aerial vehicles (UAVs).

In formulating a state space representation suitable for a Kalman filter solution, it has been assumed that:

- the motion between image frames is global, translational and of fixed high-resolution integer pixel increments;
- the point spread function of the imaging system can be modelled by a spatial averaging and decimation process;
- errors in the spatial modelling and other optical aberrations can be treated as a Gaussian measurement noise;
- motion estimation errors can be included within a Gaussian system noise.

The MAP implementation published by Schultz and Stevenson in [7] was chosen as a comparison algorithm as it is both a standard super-resolution approach and has a statistical basis similar to that of the Kalman filter. In the most general sense, MAP algorithms seek to find the super-resolution image estimate with the maximum probability of

being the underlying scene that was captured by the low-resolution sequence.

Section 2 of this paper details the super-resolution problem statement. In Section 3, the time-invariant steady-state Kalman filter developed. Section 4 provides a brief overview of the MAP algorithm from [7]. A theoretical justification of the expected results are provided in section 5. Simulation results are presented in section 6 to both quantitatively and aesthetically compare the steady-state Kalman filter to the MAP algorithm. A summary of the major findings and future directions of this work is provided in section 7.

2 Problem Statement

The measurements of the original imaging system are considered to be the individual low-resolution frames ($J \times J$ pixels) from a video sequence. The scene of interest is defined by smallest rectangular region that bounds the span of the entire sequence of low-resolution image frames. Between each frame of the video sequence, the imaging system experiences a constant translational motion, of a fixed integer of high resolution pixels, relative to the scene of interest.

Only a finite number, K , of low-resolution image frames will be used to estimate the super-resolution image. To simplify processing, the low-resolution image frames are converted into 1-dimensional vectors using column-wise lexicographical ordering. The vector containing the pixel values from the k^{th} low-resolution frame is denoted as \mathbf{y}_k , for $k = 1..K$. A sequence of 5 frames was reported to be an appropriate compromise between the MAP computational time and resulting super-resolution image quality in [7]. A sliding window of frames could be used in both the Kalman and MAP algorithms to sequentially produce an output sequence of super-resolution images, however this dynamic expansion will not be examined within this paper.

The desired super-resolution image spans the scene of interest, providing an improvement in resolution by the magnification factor, m , relative to the low-resolution pixels. Thus every frame in the low-resolution image sequence corresponds to an $m.J \times m.J$ pixel block within the super-resolution image, where each low-resolution pixel corresponds to the spatial average of an $m \times m$ block of high-resolution pixels.

The states of the Kalman filter, \mathbf{x}_k for $k = 1..K$, may be considered to be column-wise lexicographical vectors of high-resolution images, each with the same resolution and original dimensions as the

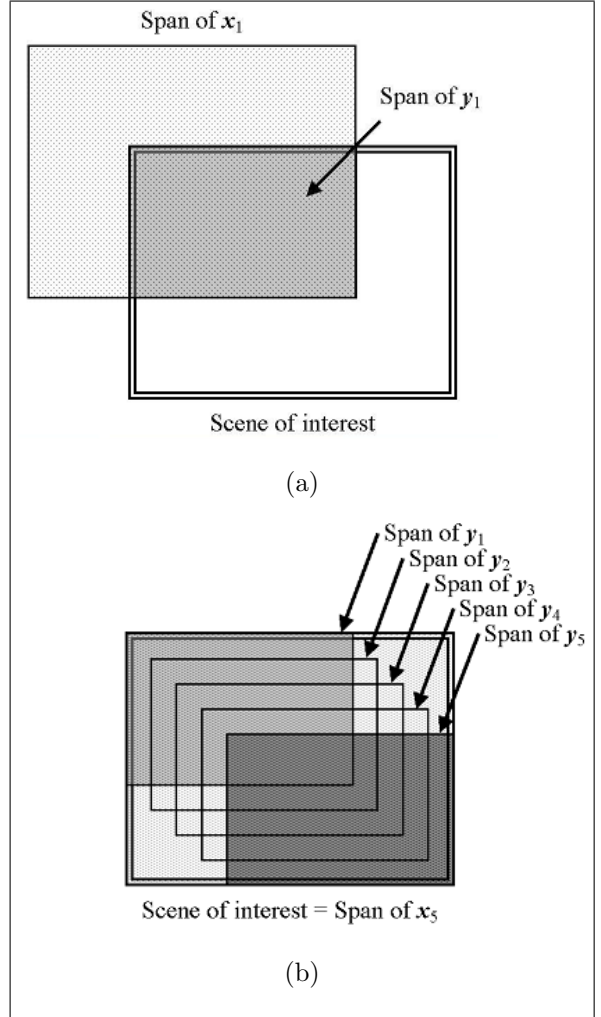


Figure 1: Example span of the initial state \mathbf{x}_1 , the final state \mathbf{x}_K (b), and the spans of the low-resolution frames \mathbf{y}_1 to \mathbf{y}_K relative to the scene of interest, for the case of a sequence of 5 images with a diagonally downward sensor motion. Note that the span of \mathbf{x}_5 is equal to the scene of interest and contains the span of all previous low-resolution image frames.

desired super-resolution image. The first state, \mathbf{x}_1 , will contain the span of the first low-resolution frame, \mathbf{y}_1 , in the corner (or edge) of the direction of the sensor motion, as shown in figure 1a for a downward and rightward motion. The remainder of the first state lies outside of the scene of interest and is filled with redundant zeros. Each of these 'high resolution images' will experience the same shift in the camera sensor as the low-resolution image frames, such that each state \mathbf{x}_k will contain the span of the k^{th} low-resolution frame in the corner (or edge) of the direction of the sensor motion. This positioning causes each progressive state, \mathbf{x}_k , to contain the span of all low-resolution image frames from \mathbf{y}_1 to \mathbf{y}_k , allowing each new low-resolution image frame to correct all previous pixel estimates in the super-resolution image. The

final state, \mathbf{x}_K , will exactly span the scene of interest and provide the final super-resolution estimate. This is demonstrated in figure 1 for a diagonally downwards sensor motion with $K = 5$.

It should be noted that there is a large redundancy in state space in all but this last state, and even here some pixels may still remain unobserved by the sensor due to the characteristics of the sensor motion. Each state space could be reduced to cover only the high-resolution span of the corresponding low-resolution image frame with little detriment to the final super-resolution image.

The measurement matrix, \mathbf{H} , models the spatial averaging process used to create the low-resolution image frame \mathbf{y}_k from a segment of the high-resolution image \mathbf{x}_k . Errors in the assumption of the simple spatial averaging point spread function, as well as any further system aberrations and noise, are modelled as Gaussian measurement noise, \mathbf{v}_k , with covariance matrix $\mathbf{R} = \sigma_R^2 \mathbf{I}$ and zero mean. The relationship between the low-resolution frames and the high-resolution image is given by

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k. \quad (1)$$

The motion of the sensor between frames is assumed to be known and is modelled by a system matrix, \mathbf{F} . Possible errors in the motion estimation are modelled as Gaussian system noise, \mathbf{w}_k , with covariance matrix $\mathbf{Q} = \sigma_Q^2 \mathbf{I}$ and zero mean. The change in location of the sensor's field of view between frames is modelled by equation (2) where \mathbf{x}_k corresponds to the current high-resolution image containing the spatial span of \mathbf{y}_1 to \mathbf{y}_k , and \mathbf{x}_{k+1} corresponds to the high-resolution image after the next camera motion when \mathbf{y}_{k+1} enters the field of view.

$$\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_k + \mathbf{w}_k \quad (2)$$

3 Time-Invariant Steady-State Kalman Filter

The video capturing equations, defined in equations (1) and (2), serve as measurement and transition equations that fully describe the image capturing process. Since the measurement and transition matrices within these equations are time-invariant, the Kalman gain can be precomputed (for known motion).

Each time increment, k , of the Kalman filter will correspond to the sequential input of a low-resolution image frame, \mathbf{y}_k , from a video sequence of finite length. The prediction equation in (3) of the time-invariant Kalman takes the

current super-resolution image after processing frame k , $\hat{\mathbf{x}}_{k|k}$, and applies the transition matrix \mathbf{F} to predict the next super-resolution image, $\hat{\mathbf{x}}_{k+1|k}$. The time-invariant correction equation in (4) updates the prediction from equation (3), with a weighted error signal. The error signal is equal to the difference in low-resolution pixel values between the current frame, \mathbf{y}_k , and those which would be created from the predicted super-resolution image when passed through the measurement matrix \mathbf{H} . The weighting is provided by the Kalman gain matrix, \mathbf{K} , as defined later within this section.

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}\hat{\mathbf{x}}_{k-1|k-1} \quad (3)$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}(\mathbf{y}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1}) \quad (4)$$

In both standard and time-invariant Kalman filters, the error covariance matrices, $\Sigma_{k|k-1}$ and $\Sigma_{k|k}$, are also calculated in a similar prediction and correction process. However, each correction step requires the inversion of a $(\mathbf{H}\Sigma\mathbf{H}^T + \mathbf{R})$ term and requires an unfeasible amount of processing for standard sized images. It is the avoidance of this inversion that has led to approximations of the Kalman filter in past literature [3, 4, 5, 6]. Instead, the use of a steady-state error covariance matrix, Σ , is proposed here. This matrix needs only to be calculated once, either offline or in advance, by solving the Algebraic Ricatti Equation (ARE) given in equation (5). The covariance matrix, \mathbf{Q} , modelled in equation (5) will be higher than the predicted system noise to incorporate a driving term to boost the gain of the new pixels. After solving the ARE, the Kalman gain matrix then only needs to be calculated once using equation (6).

$$\Sigma = \mathbf{F}\Sigma\mathbf{F}^T - \mathbf{F}\Sigma\mathbf{H}^T(\mathbf{H}\Sigma\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\Sigma\mathbf{F}^T + \mathbf{Q} \quad (5)$$

$$\mathbf{K} = \Sigma\mathbf{H}^T(\mathbf{H}\Sigma\mathbf{H}^T + \mathbf{R})^{-1} \quad (6)$$

To approximate a zero mean process, the midpoint of available pixel values ('127' for an 8 bit image) is subtracted the low-resolution image pixel values. At the conclusion of the Kalman filtering, the midpoint value is added back onto the pixels within the final super-resolution image estimate, $\hat{\mathbf{x}}_{k|k}$.

For the best final image quality, the initial super-resolution estimate, $\hat{\mathbf{x}}_{1|0}$, has experimentally been determined to be a zero matrix containing only the correctly positioned bilinear interpolation of one low-resolution frame after applying the midpoint pixel offset. The pixels in the initial estimate of the super-resolution image that are unobservable by this low-resolution frame are set to zero.

4 MAP Summary

The maximum *a posteriori* (MAP) algorithm used as a comparison within section 5 of this paper is detailed in [7]. A brief outline of the major steps are provided within this section.

The MAP algorithm presented in [7] aims to create a super-resolution image of the middle (k^{th}) frame of a low-resolution video sequence. It begins by isolating that central frame and uses a zero-order hold on this frame to create the initial estimate of the super-resolution image, \mathbf{x}_k . (Note that unlike the states of the Kalman filter, there is no redundancy within the MAP estimate). The remaining image frames are bundled into one stacked observation vector, \mathbf{y}' , and their relationship to the super-resolution image is modelled in a stacked motion-compensated subsampling matrix, $\hat{\mathbf{H}}'$. Any low-resolution pixel that is not fully contained within the span of the super-resolution estimate of the middle frame is discarded.

The objective function to be minimised in the MAP algorithm within [7] is given by equation (7). This function is the sum of two distinct terms, each of which are defined in turn below.

$$f(\mathbf{x}^{(k)}, T, \Lambda) = \sum_{a,b} \sum_c^4 \rho_T \left(d_{a,b,c}^t \mathbf{x}^{(k)} \right) + \left\| \Lambda^{1/2} \left(\mathbf{y}' - \hat{\mathbf{H}}' \mathbf{x}^{(k)} \right) \right\|^2 \quad (7)$$

The first term in the objective function provides an edge preserving smoothing of the current super-resolution estimate. Second order finite differences $d_{a,b,c}^t$ are calculated at each pixel (a, b) of the super-resolution estimate $\mathbf{x}^{(k)}$ in each of the four possible neighbouring pixel directions (horizontal, vertical and both diagonals) as given by

$$d_{a,b,1}^t \mathbf{x}^{(k)} = x_{a,b-1} - 2x_{a,b} + x_{a,b+1} \quad (8)$$

$$d_{a,b,2}^t \mathbf{x}^{(k)} = x_{a-1,b} - 2x_{a,b} + x_{a+1,b} \quad (9)$$

$$d_{a,b,3}^t \mathbf{x}^{(k)} = 0.5x_{a+1,b-1} - x_{a,b} + 0.5x_{a-1,b+1} \quad (10)$$

$$d_{a,b,4}^t \mathbf{x}^{(k)} = 0.5x_{a-1,b-1} - x_{a,b} + 0.5x_{a+1,b+1} \quad (11)$$

Each of those directional second order finite differences is then passed into the Huber function given by equation (12). Small discontinuities are smoothed by the quadratic cost of the first term, but significant edges and discontinuities above the threshold T are preserved by the linear weighting of the second term. The resulting output of the Huber function is summed over each of the four possible second order finite difference directions

at each pixel throughout the super-resolution estimate to form the first term of the objective function in equation (7).

$$\rho_T(z) = \begin{cases} z^2, & |z| \leq T \\ 2T|z| - T^2, & |z| > T \end{cases} \quad (12)$$

The second term within the objective function in (7) constrains the super-resolution image estimate to maintain fidelity to the measurements within \mathbf{y}' . The matrix Λ defined by equation (13) provides a weighting to the constraint, based on the individual confidence weightings, $\lambda^{i,k}$ for $i \neq k$, chosen by the user for each of the low-resolution image frames, \mathbf{y}_i , with respect to the middle frame, \mathbf{y}_k .

$$\Lambda = \text{blockdiag}(\lambda^{i,k} \mathbf{I}, \dots, \lambda^{k-1,k} \mathbf{I}, \lambda^{k+1,k} \mathbf{I}, \dots) \quad (13)$$

The MAP algorithm in [7] then uses a gradient projection algorithm to minimise the objective function whilst maintaining fidelity to the central low-resolution frame, and creates the corresponding super-resolution image estimate. Note, however, that [7] uses an adaptive step size within their gradient projection algorithm, whereas the implementation within this paper uses a fixed step size (experimentally optimised).

5 Theoretical Expectations

The Kalman filter and the MAP estimator are derived from the same statistical basis and can both produce the minimum variance solution, but only under the strict requirement that every variable for the MAP estimator follows a Gaussian probability distribution [8]. Note that this condition is unnecessary for the Kalman filter [9].

However, images do not commonly follow a Gaussian probability distribution, and consequently any MAP estimator used for super-resolution purposes will deviate away from the minimum variance solution of the Kalman filter. To minimise this effect, within [7] a Huber Markov Random Field (with Gibbs distribution) is introduced into the MAP cost function, deviating away from the Kalman filter formulation. Furthermore, the MAP algorithm in [7] discards all low-resolution pixels that are not fully contained within the span of the central frame, reducing the information available for the super-resolution process.

The Kalman filter is therefore expected, in theory, to outperform the MAP algorithm of [7].

6 Simulation Results

A sequence of five low-resolution image frames, each 16 x 16 pixels, were created by spatially averaging and decimating a model image by a factor

of 2 along each dimension. In between frames, the modelled sensor moved down and to the right by one high resolution pixel. In order to compare the results of the MAP and Kalman algorithms, the super-resolution images created by Kalman filter were cropped to the span of the third low-resolution frame only. The third low-resolution frame from the sequence and the model image corresponding to the span of this frame are shown in figure 2b and 2a respectively.

Gaussian noise was added to each frame in the image sequence in two steps. System noise, with variance σ_Q^2 , was added to the model high-resolution image to represent motion estimation errors. A low-resolution image frame was created from this noisy model image. Measurement noise, with variance σ_R^2 , was then added to represent errors caused by the sensor itself. Four levels of noise are considered here: ‘no noise’, ‘low noise’, ‘moderate noise’, and ‘high noise’ as defined in table 1. At each noise level the Kalman filter and the MAP estimator were run over a wide range of parameters to find the images with the lowest MSE. The ‘best’ model parameters at each noise level for the MAP and Kalman filter are provided in table 1.

Table 1: MAP & Kalman model parameters used

Noise Level	Applied		MAP		Kalman	
	σ_Q^2	σ_R^2	λ 's	T	σ_Q^2	σ_R^2
No Noise	0	0	all 1	10	0.1	0
Low Noise	0.01	0.01	all 1	10	0.1	0.01
Medium Noise	0.1	0.1	all 1	12	1	0.1
High Noise	1	1	all 1	25	2	1

The MSE of each super-resolution estimate was calculated at each noise level by using equation (14), as per [10], where $\mathbf{x}_{a,b}$ is the pixel in row a and column b of the model high-resolution image and $\hat{\mathbf{x}}_{a,b}$ is the corresponding pixel in the super-resolution image.

$$MSE = \frac{\sum_{a=0}^{mJ-1} \sum_{b=0}^{mJ-1} (\hat{\mathbf{x}}_{a,b} - \mathbf{x}_{a,b})^2}{\sum_{a=0}^{mJ-1} \sum_{b=0}^{mJ-1} (\mathbf{x}_{a,b})^2} \quad (14)$$

For the ‘no noise’ case, the resultant MAP and Kalman filter super-resolution image estimates are shown in figure 2d and 2e respectively. Figure 2b and 2c provide additional visual comparisons created using nearest-neighbour and bilinear interpolations of the noise-free third low-resolution frame. MSE values for all four approaches are provided in table 2. The MAP super-resolution estimate provides almost a tenfold reduction in the MSE values of the interpolation techniques.

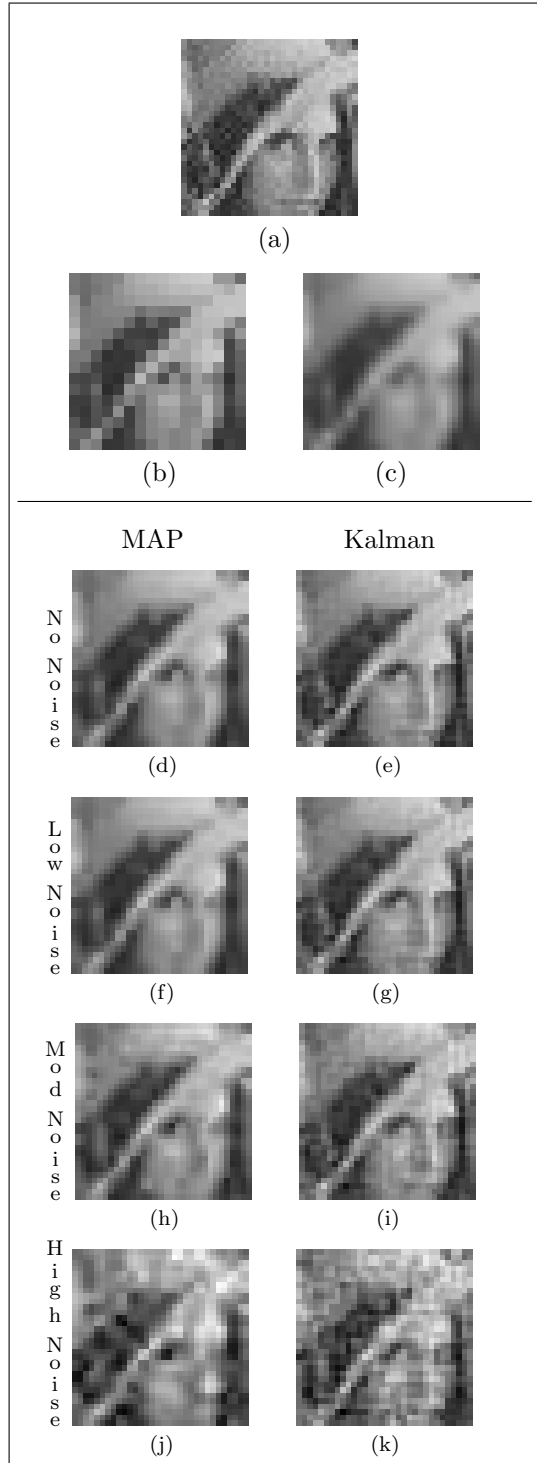


Figure 2: Images resulting from super-resolution simulations under different noise conditions; (a) Model image cropped to span 3rd frame; (b) 3rd Low-resolution frame; (c) Bilinear interpolation of 3rd low-resolution frame; No noise: MAP (d) and Kalman (e) SR images; Low noise: MAP (f) and Kalman (g); Moderate noise: MAP (h) and Kalman (i); High noise: MAP (j) and Kalman (k).

Visually the MAP estimate provides more image detail than either interpolation technique whilst maintaining a generalised smoothing effect. However, the Kalman filter results further halve the MSE of the MAP estimate and creates a sharper image with significantly more fine detail (for example, around the eyes and lips in figure 2e when compared to the model image of 2a).

Table 2: MSE for noiseless case

Method	MSE
ZOH	0.1211
Bilinear	0.1008
MAP	0.0120
Kalman	0.0059

The super-resolution images created by both the MAP and Kalman filter for each of the noise cases are shown in figures 2d-k, with the actual MSE values provided in table 3. Even as increasing noise levels degrade the super-resolution

image estimates, the Kalman filter continues to outperform the MAP estimator in MSE values. Visually, at low and moderate noise levels, the Kalman filter continues to produce images with finer details than the MAP estimator. At the ‘high noise’ level, the super-resolution images are understandably inaccurate, however it is worth noting that the two approaches differ visually in their treatment of the noise. The MAP algorithm has a smoothing effect that visually reduces the effects of the noise at the detriment of fine details, yet a degree of sharp edges caused by the noise itself are preserved. The Kalman filter breaks up the noise to create an image with a ‘noisier’ speckled appearance, a higher degree of accuracy in the detailed features, and a lower overall MSE.

Table 3: MSE for various noise levels

Applied Noise	MAP	Kalman
No Noise	0.0120	0.0059
Low Noise	0.0124	0.0060
Moderate Noise	0.0137	0.0090
High Noise	0.0316	0.0234

7 Conclusion

This paper has proposed a time invariant steady-state Kalman filter for image super-resolution for the specific case of a sensor moving with a constant translational motion. This filter has been shown to outperform the MAP approach from [7] in comparisons of MSE and restored visual detail. Further investigations will be needed to determine the suitability of this algorithm to be adapted to more generalised motions, and to quantify its computational efficiency.

References

- [1] S. Borman and R. Stevenson, “Spatial resolution enhancement of low-resolution image sequences a comprehensive review with directions for future research,” tech. rep., Laboratory for Image and Signal Analysis (LISA), University of Notre Dame, 8 July, 1998–1998.
- [2] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, “Advances and challenges in super-resolution,” *submitted to the International Journal of Imaging Systems and Technology*, 2004.
- [3] M. Elad and A. Feuer, “Super-resolution reconstruction of continuous image sequences,” in *1999 International Conference on Image Processing (ICIP 99), Proceedings of the*, vol. 3, (Kobe, Japan), pp. 459–463, 1999.
- [4] M. Elad and A. Feuer, “Superresolution restoration of an image sequence: adaptive filtering approach,” *Image Processing, IEEE Transactions on*, vol. 8, no. 3, pp. 387–395, 1999.
- [5] J. Woods and V. Ingle, “Kalman filtering in two dimensions: Further results,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 29, no. 2, pp. 188–196, 1981.
- [6] A. Patti, A. Tekalp, and M. Sezan, “A new motion-compensated reduced-order model kalman filter for space-varying restoration of progressive and interlaced video,” *Image Processing, IEEE Transactions on*, vol. 7, no. 4, pp. 543–554, 1998.
- [7] R. Schultz and R. Stevenson, “Extraction of high-resolution frames from video sequences,” *Image Processing, IEEE Transactions on*, vol. 5, no. 6, pp. 996–1011, 1996.
- [8] J.-P. Drécourt, “Kalman filtering in hydrological modelling,” DAIHM Technical Report 2003-1, DHI Water & Environment, Hørsholm, Denmark, May 2003.
- [9] R. Kalman, “A new approach to linear filtering and prediction problems,” *ASME - Journal of Basic Engineering*, vol. 82 (Series D), pp. 35–45, 1960.
- [10] R. Schultz and R. Stevenson, “A bayesian approach to image expansion for improved definition,” *Image Processing, IEEE Transactions on*, vol. 3, no. 3, pp. 233–242, 1994.