Super Resolution of speed signs in video sequences

Faisal Mufti, Robert Mahony, Jonghyuk Kim
Department of Engineering,
Australian National University,
ACT, 0200, AUSTRALIA.
faisal.mufti@anu.edu.au, mahony@ieee.org, jonghyuk.kim@anu.edu.au

Abstract

This paper considers the problem of generating a high resolution (super resolved) image of an object that is visible in a sequence of low resolution video frames. We are motivated by applications where the object is moving in the video sequence, due either to movement of the camera or of the object itself. In such cases, accurate sub-pixel image registration can be significantly improved by stochastic filtering for the estimates of inter-frame motion displacement. We use the expectation-maximization (EM) framework to formulate the coupled image registration filtering and super resolution problem. The expectation step is solved as a Bayesian smoothing algorithm based on a motion displacement model correlated to successive frames image registration with the previous super resolved image. The super resolved image is updated as a maximum likelihood estimate based on the new image registration parameters. Results obtained show excellent convergence, robustness and demonstrate quantitative improvement in image quality for real world sequences of speed signs obtained from a moving vehicle.

1 Introduction

Imaging devices have a limited capability to produce high resolution images. Spatial resolution is restricted by the spatial density of charged-coupled device (CCD) array. This causes blur and constrains the size of the spatial feature that can be otherwise visually detected in a scene [18]. One method to obtain high resolution images from real world data, is to fuse information from a number of low quality images to obtain a high resolution estimate of the desired scene; a process known as Super resolution (SR). There are many applications where only part of the image, associated with an object, needs to be super resolved, for example; recognizing speed signs, licence plates or resolving a face in video streams derived from vehicle or surveillance cameras.

Early work on Super resolution was done by Tsai and Huang [24] in frequency domain but was limited due to non-global motion estimation and lack of spatial domain knowledge. Stark and Oskoui [21] proposed Projection onto Convex Sets (POCS), later extended by Tekalp et al. [22] using successive projection as prior information of high resolution image. Stochastic approaches have included a Bayesian framework [20] with Gaussian prior and edge preserving Hubber Markov random field prior [17]. Later joint formulation of image registration and restoration was considered by Tom et al. [23] using Maximum likelihood (ML) and Hardie et al. [9] via Maximum a-posteriori (MAP) method. Iterative methods have been proposed for SR [11] as well as for extending the field of view by proper registration of low resolution (LR) images [3]. More recently, Woods et al. [26] presented an Expectation Maximization (EM) framework in frequency domain for simultaneous estimation of registration, noise and image statistic parameters with high resolution image.

Sub-pixel registration of the observed data sequence is a key requirement for any SR reconstruction algorithm [13]. To overcome error in image registration due to poor resolution, Wang et al. [25] proposed an algorithm that performs registration in high resolution domain using regularized mean curvature upsampled images [1]. Accurate registration of low resolution (LR) images is especially important when the sequence of LR images is drawn from a moving video sequence. Work related to SR in video sequences

Figure 1. Multi-scale super resolution model
has been considered in [18], [17], [25]. There are several issues that are not well addressed in existing literature of SR for videos. Existing methods do not deal with variation in scale of an image (expansion/decrease) that occurs when the image of an observed object changes significantly. When camera moves forward past an object (a speed sign in our case), the image of (the sign) expands frame by frame in the image sequence. Thus, early images in the image sequence will contain less pixels of information than later images (see Figure 1). In these spatio-temporal shifted multi-scale frames a natural approach is to incorporate a motion model of the object for image registration. Methods have been proposed to track and super resolve an object [6], [4]. However, to the authors’ knowledge there is no significant work on unified EM framework for filtering and super resolution.

In this paper we consider the problem of obtaining a SR image of a speed sign obtained from a sequence of image captured from a moving vehicle. We use the EM framework to formulate a coupled image registration filtering and super resolution problem for moving objects in a video sequence. We characterize image homographies of successive frames in terms of camera pose along with the depth and orientation of the object with respect to the super resolved image plane. The super resolved image plane of the image sequence data may be chosen in any orientation or point and is not tied to an existing LR image as done in classical methods [16], [22], [3], [25], [9], [26], [10]. The homography parameters play the role of the hidden variables in the expectation maximization framework. The expected value of these parameters is given by the output of a Bayesian smoothing filter for homography. The measurement data is provided by the expectation constant velocity model. The super resolved image of the object for image registration. Methods have been proposed to track and super resolve an object [6], [4]. However, to the authors’ knowledge there is no significant work on unified EM framework for filtering and super resolution.

The iterative optimization is computationally efficient and achieve fast convergence for the speed sign SR application considered. The paper is organized as follows: Section II describes a multi-scale SR observation model, section III presents EM framework of homography for Super resolution in spatial domain, measurement process and the formulation of Bayesian smoothing filter for homography, section IV describes the implementation of the algorithm with experimental results followed by conclusion in section V.

2 Multi-scale Super resolution Model

Super resolution (SR) algorithms are based on a motion model and an image model. The motion model represents motion of the real world with respect to camera or vice versa. The image model represents the quantization of the real world through camera sensor operation. In multi-scale SR, the dynamic scene has scale varying image model associated with the re-sampling of moving video sequences.

2.1 The Motion Model

Our motion model represents the movement of the camera with respect to an object in the scene or vice versa. This motion is independent of the image resolution. The image \( f \) represents the object as observed from the desired pose. Then the dynamic motion model for \( k \)th frame in a sequence is given by

\[
f_k = A_k f \quad 1 \leq k \leq n
\]

Here \( f_k \) is the SR version of \( f \) as would be observed by camera at time \( k \) and \( A_k \) is the motion matrix. In reality the true motion is never known and an estimate of this motion is used that accounts for the unknown motion displacement. Each \( k \)th motion matrix represents inter frame motion and each row of motion matrix contains non zero values for observable pixels. The constant moving video sequences have slight shift within frames, hence the sparsity of this matrix depends upon the amount of temporal variations between frames. The pose estimation of the camera with successive perspective projection on the SR image is represented by each \( A_k \) matrix of size \( q_k^2 M_i M_j \times q_k^2 M_i M_j \). The scale factor \( q_k \) depends upon the ratio between each observation frame and the SR image size.

In practice, the motion matrix is calculated from low resolution frames and assumed that this motion describes pixel motion in high resolution [17, 11]. On the other hand the precise estimation of displaced pixels is an important factor for SR. Therefore, we compute the \( A_k \) matrix using pixels in high resolution [25], which is a close realization and better estimate of the motion model (1).

2.2 Multi-scale Imaging Model

The goal of this model is to form a basis for mapping a high resolution image \( f \) of an object in an underlying scene observed into a sequence of LR images \( g_k \) [17]. The image acquisition process of the camera is implemented as a discrete matrix operator termed camera matrix \( S \)

\[
g_k = S_k f_k + u_k \quad 1 \leq k \leq n
\]
The vector $g_k$ represents the $k$th low resolution image of size $q_k M_i \times q_k M_j$ reordered as $q_k^2 M_i M_j \times 1$ vector and the high resolution image $f$ is a SR image of size $q_n M_i \times q_n M_j$ reordered as $q_n^2 M_i M_j \times 1$. The vector $v_k$ is the CCD sensor noise associated with the samples of this noisy process and is assumed to be independent and identically (i.i.d) Gaussian noise. The camera matrix $S_k$ is the product of a down-sampling and a blurring operation. The down-sampling matrix comprises a scale variant down-sampling operation and the scale invariant blurring (blur matrix) due to point spread function of the camera lens

$$S_k = D_k B_k \quad 1 \leq k \leq n$$

where for each vectorized image $q_k^2 M_i M_j \times 1$, $D_k$ is $k$th down-sampling matrix of size $M_i M_j \times q_k^2 M_i M_j$. The blurring matrix $B_k$ is of size $q_k^2 M_i M_j \times q_k^2 M_i M_j$. Hence, each row of this block diagonal $S_k$ matrix of size $M_i M_j \times q_k^2 M_i M_j$ forms a discrete image acquisition function that sequentially maps a higher resolution image to produce a single pixel of scale variant low resolution image.

### 2.3 Total Observation Model

The total SR observation model without generalization can be expressed as

$$g_k = S_k A_k f + v_k \quad 1 \leq k \leq n$$

where $v_k$ the additive noise vector of size $q_k^2 M_i M_j \times 1$ accounts for this noisy process.

### 3 EM framework for Super resolution

Image registration is parameterized as a state comprising camera pose (rotation and translation) and structure (depth and normal to the image plane). The state contains sufficient information to reconstruct the homography $A_k$ that relates the image observed at time $k$ with SR image. The general formulation allows us to consider an arbitrary SR image plane, for example an image plane that is fronto parallel to the observed object. The state formulation also enables us to formulate a physically motivated motion model that is used to condition the estimation of image registration parameters for the observed image sequence. This process is achieved through a Bayesian smoothing filter implemented as a linearization of the system trajectory. The expected value of the estimated state is combined with the desired Maximum Likelihood (ML) estimate for SR image estimation using a modified EM framework.

#### 3.1 Measurements

As shown in Fig. 2 the coordinates of the point $p \in P$ in the reference frame $n$ is $X_n \in \mathbb{R}^3$. The rigid body coordinate transformation between frame $k$ and $n$ is

$$X_k = \frac{k}{n} R_k X_n + \frac{k}{n} T_k \quad 1 \leq k \leq n$$

where $\frac{k}{n} R \in SO(3)$ and $\frac{k}{n} T \in \mathbb{R}^3$ are the rotation and translation components that transformed $n$ frame to $k$ frame. If the normal to the plane with respect to $n$ frame is given by $N_n = [n_x, n_y, n_z]$ and $d_n$ is the distance between the plane to the origin of frame $n$ then the planar homography [15]

$$\frac{k}{n} H_k := \frac{k}{n} R_k + \frac{1}{d_n \frac{k}{n} T_k} N_n^T$$

transforms $X_n \in \mathbb{R}^3$ to $X_k \in \mathbb{R}^3$ as

$$X_k = \frac{k}{n} H_k X_n.$$  

Let $x_i$ denote the homogenous image coordinates of the point $X_i$ in image $i$ then

$$\lambda_k x_i = \lambda_n (\frac{k}{n} H_k x_n)$$

where $\lambda_k$ and $\lambda_n$ are the scale factors, implied by the fact that we can recover the 3D object in a 2D image up to a scale factor $\frac{1}{d_n}$. The homography mapping in (8) assumed identity intrinsic camera calibration matrix $K_c$. However, in order to recover the 3D Euclidean structure, knowledge of camera calibration matrix $K_c$ is essential. From projective geometry we know that

$$\lambda_n x_n = K_c X_n$$

also

$$\lambda_k x_i = K_c (\frac{k}{n} R_k + \frac{1}{d_n \frac{k}{n} T_k} N_n^T) X_n$$

Using (9) and (10) we may write

$$x_k = \lambda K_c (\frac{k}{n} H) K_c^{-1} x_n$$

where $\lambda = \frac{\lambda_n}{\lambda_k}$. Using (8) and (11) we have

$$\lambda (\frac{k}{n} H) = \lambda K_c (\frac{k}{n} R_k + \frac{1}{d_n \frac{k}{n} T_k} N_n^T) K_c^{-1}$$

$$\lambda (\frac{k}{n} \hat{H}) = \lambda K_c (\frac{k}{n} H) K_c^{-1}$$

where $(\frac{k}{n} \hat{H})$ is the calibrated homography [27] used to recover pose of the SR image in 3D ($\mathbb{R}^3$). With a given a homography between two images, we decompose the homography matrix and find physically two possible solutions by imposing positive depth constraint for the camera $d > 0$ [15]. The unique solution is obtained by assuming rotation close to identity since there is a very little rotation of the vehicle along the three axis during the sequence.

The constant motion of the feature from $k$ to $n$ frame are the filter measurements $u_k$, where

$$u_k = \frac{k}{n} R_k T_k$$

$$\frac{d}{d_n} \frac{k}{n} R_k$$

$$\frac{k}{n} T_k \quad \frac{k}{n} R_k$$

(14)
hence

\[
z_k^T = \begin{bmatrix} \phi_k & \theta_k & \psi_k & T_{n_k} & V_{n_k} & d_{n_k} & N_{n_k} \end{bmatrix}^T \tag{15}\]

The rotation matrix \( R_k \in \mathbb{R}^3 \) has been decomposed into Euler angles of roll (\( \phi \)), pitch (\( \theta \)) and yaw (\( \psi \)). Note that these parameters are not directly related with the homography because \( d_{n_k} \) is related to frame \( k \) and not with frame \( k \).

At each time stamp \( k \) a homography matrix \( H_k \) (see Fig. 3) is computed from image registration technique between the speed signs in the scene and reference frame of SR image and is transformed into measurement parameters of the filter.

![Figure 2. Structure and motion parameters of image frames with moving vehicle](image1)

**Figure 2. Structure and motion parameters of image frames with moving vehicle**

![Figure 3. Homography between speed sign in the scene with SR image](image2)

**Figure 3. Homography between speed sign in the scene with SR image**

### 3.2 Bayesian smoothing filter

The goal of the Bayesian smoothing filter is to estimate the state \( y_k \) as a function of homography in terms of motion \((R_k, \hat{T}_k)\) and structure parameters \((\hat{N}_k, \hat{d}_{n_k})\). The homography \( H_k \) that is used in Super resolution (SR) algorithm is computed from these estimates. The state estimation is modelled as a constant velocity model. The state variables of pose, depth and normal to the plane are estimated with Bayesian smoothing filter. The Kalman filter [12] is a discrete forward dynamic model of Bayesian estimation and a recursive solution to the discrete data filtering. While the smoothing filter runs in the backward direction in time sequence of the last measurement, computing the smooth estimate of the state from the intermediate results stored in the forward step. The accuracy of the smoother is thus superior to that of a forward filter because it uses more measurements for its estimation. The state vector for this constant velocity model is given as

\[
y_k^T = \begin{bmatrix} \phi_k & \theta_k & \psi_k & T_{n_k} & V_{n_k} & d_{n_k} & N_{n_k} \end{bmatrix}^T \tag{16}\]

The state (16) represents elements of the filter in terms of motion and structure parameters of the motion model and \( V \) is the velocity of the object. The non-linear measurement dynamics (15) are linearized with first degree of Taylor’s approximation and is implemented as Extended Kalman Filter (EKF) to optimally execute the filter. The measurement model is given as

\[
z_k = h(y_k, k) + w_k \tag{17}\]

where \( w_k \) is the measurement noise vector \((w_k \sim N(0, Q_k))\) and \( Q_k \) is the associated covariance matrix. Here \( h(\cdot) \) is the non linear function of the measurements. The linearization of measurements (17) around point \( y_k \) is given by

\[
z_k \approx C_k(y_k - \hat{y}_k) \tag{18}\]

where

\[
C_k = \frac{\partial h}{\partial y} \bigg|_{y_k = \hat{y}_k} \tag{19}\]

is the Jacobian matrix consists of partial derivative terms that relate how homography changes with small variation in pose of the object.

In the above estimation, the Kalman filter estimates the uncertainty of the state as pdf over the states. The pdf of the predict state of the transition from \((k - 1)\) to \( k \) step will be given by Chapman-Kolmogorov equation

\[
p(y_k|z_{k-1}) = \int p(y_k|y_{k-1})p(y_{k-1}|z_{k-1})dy_{k-1} \tag{20}\]

Using Bayes rule we can calculate the update pdf for state \( y_k \) after taking all the measurement upto \( k \) into consideration as

\[
p(y_k|z_k) = \frac{p(z_k|y_k)p(y_{k-1}|z_{k-1})}{\int p(z_k|y_k)p(y_{k-1}|z_{k-1})dy_{k-1}} \tag{21}\]
The Kalman gain matrix $K$ and error covariance matrix $P$ are

$$K_k = P_{k|k-1} C_k T (C_k P_{k|k-1} C_k^T + R_k)^{-1}$$

$$P_{k|k} = (I - K_k C_k) P_{k|k-1}$$

In the next step, the smoothing filter [8] runs a second sweep in the backward direction in time sequence from time $t_n$ of the last measurement, computing the smooth estimate of the state from the intermediate results stored in the first step. With each iteration of backward sweep, the old filter values are updated to yield an improved and smooth estimate, which is based on all the measurements. The smooth estimate is given by [2]

$$\hat{y}_{k|n} = \hat{y}_{k|k} + G(k)[\hat{y}(k + 1|n) - \hat{y}(k + 1|k)]$$

where the smoothing gain matrix $G(k)$ is given by

$$G(k) = P_{k|k} \Phi T (k + 1, k) P_{k|k-1}^{-1}(k + 1, k), \quad k = n - 1, \cdots, 0.$$  (24)

and $\Phi$ is the constant velocity state model.

The output of this filter has smooth statistical distribution as $a_k \sim (\tilde{a}_k, \Sigma^2_{a_k})$ where $a_k \in \mathbb{R}^{M_i \times 1}$ and is given by

$$a_k \approx \frac{\partial f(y_k)}{\partial y_k} \bigg|_{\hat{y}_k} (y_k - \hat{y}_k)$$

The observed homography $a_k$ (implemented as matrix operator $A_k$) is a function of low resolution frames and the super resolved frame such as $A_k(\tilde{g}_k, f)$ and cannot be maximized directly. We therefore, use the Expectation-Maximization (EM) framework to iteratively maximize the super resolution parameter $f$.

### 3.3 EM framework

The EM framework allows for estimation of stochastic parameters given only partial measurements of the hidden variables. Let $g_k$ denote the observable data of low resolution (LR) frames $g_k = [g_{k_1}^T \cdots g_{k_5}^T]$ and $A_k = f(R_k, T_k, N_a, d_a)$ as the hidden variable then $g_k$ and $A_k$ together constitute the complete data. The conditional distribution of $A_k$ using Bayes rule will be computed as

$$p(A_k|g_k, f) = \frac{p(g_k|A_k, f)p(A_k|f)}{\int p(g_k|A_k, f)p(A_k|f)dA_k}$$  (26)

then the marginal likelihood of $p(g_k|f)$ becomes

$$p(g_k|f) = \int p(g_k|A_k, f)p(A_k|f)dA_k$$  (27)

Therefore, we may re-write equation (26) as

$$p(A_k|g_k, f) = \frac{p(g_k, A_k|f)}{p(g_k|f)}$$  (28)

Since the super resolved image depends upon the hidden states, direct minimization is difficult to achieve. Each iteration of EM algorithm consist of two steps; The E-step and the M-step.

#### 3.3.1 The E-Step

In E-step of EM algorithm one computes the expected value of the complete data log likelihood $\log p(g_k, A_k|f)$ with respect to unknown data $A_k$ given the observed data $g_k$ and the current estimate $f^{(n)}$ of the unknown parameter.

$$Q(f, f^{(n)}) := E \left[ \log p(g_k, A_k|f) | g_k, f^{(n)} \right]$$  (29)

The conditional expectation $E[.]$ provides a likelihood function $Q(f, f^{(n)})$ that depends on the previous estimate $f^{(n)}$ of the SR image and average over all the possible realizations of $A_k$. We expand this conditional expectation as

$$Q(f, f^{(n)}) = E \left[ \log p(g_k, A_k|f) | g_k, f^{(n)} \right]$$

$$= E \left[ \sum_{k=1}^{n} (||g_k - A_k f||^2) | g_k, f^{(n)} \right]$$

$$= \sum_{k=1}^{n} g_k^T g_k - 2 \sum_{k=1}^{n} (g_k^T A_k f)$$

$$+ E \left[ \sum_{k=1}^{n} (a_k^T (F^T f)a_k) | g_k, f^{(n)} \right]$$

where

$$F a_k = A_k f \quad \text{and} \quad F \bar{J}_k Y_k = A_k f.$$  (33)

After re-arranging we get

$$Q(f, f^{(n)}) = \sum_{k=1}^{n} ||g_k - \tilde{A}_k f||^2$$  (34)

$$+ \sum_{k=1}^{n} \left[ Tr(F \bar{J}_k \Sigma^2_{a_k} \bar{J}_k^T F^T) \right].$$  (35)

The right term in the above equation is the covariance that associate SR image with the filter dynamics in image space. Since the measurements are the parameters of homography there exist a relationship in particular to our application such that

$$a_k = f(z_k).$$  (36)

Using (18), (33) and (36) we may write

$$a_k \approx \tilde{J}_k Y_k$$

$$\approx \tilde{J}_k C_k Y_k$$

$$\approx \tilde{J}_k z_k$$  (37)
where $\mathbf{J}_k$ is the Jacobian for the measurement function. The empirical estimation of this Jacobian $\mathbf{J}_k$ in spatial domain is difficult and is solved through least square approximation $\hat{\mathbf{J}}_k$. Let

$$
\hat{y}_k = (y_k - \bar{y}_k) \quad \text{and} \quad \hat{z}_k = (z_k - \bar{z}_k)
$$

then

$$
\hat{\mathbf{J}}_k = \frac{1}{\|\hat{z}_k\|^2} \hat{\mathbf{f}}_k \hat{\mathbf{A}}_k^T.
$$

Using (35) and (38) we can write

$$
Q(\mathbf{f}, f^{(n)}) \approx \sum_{k=1}^{n} \|g_k - \hat{\mathbf{A}}_k \mathbf{f}\|^2
$$

$$
+ \sum_{k=1}^{n} \left[ 2 \Sigma_{\mathbf{x}}^2 \hat{z}_k \right] \left( \mathbf{f}^T \hat{\mathbf{A}}_k^T \hat{\mathbf{A}}_k \mathbf{f} \right) \tag{40}
$$

This variational approach helps in making E-step more tractable and simplifies our likelihood estimate in M-step.

### 3.3.2 The M-Step

In M-step we maximize the likelihood function $Q(\mathbf{f}, f^{(n)})$ computed in the E-step.

$$
f^{(n+1)} = \arg \max_f Q(\mathbf{f}, f^{(n)}) \tag{41}
$$

If we have some degree of prior knowledge then instead of maximizing the likelihood we maximize the expected posterior

$$
Q(\mathbf{f}, f^{(n)}) \propto E \left[ \log p(g_k, \hat{\mathbf{A}}_k \mathbf{f}| \mathbf{f}^{(k)}) \right] + \log p(\mathbf{f})
$$

Hence

$$
p(\hat{\mathbf{A}}_k|g_k, \mathbf{f}) = \frac{p(g_k, \hat{\mathbf{A}}_k \mathbf{f}| \mathbf{f})}{p(\mathbf{f})} \tag{42}
$$

where the second term in (42) is the prior probability $p(\mathbf{f})$ of SR image given as

$$
p(\mathbf{f}) = \frac{1}{Z_1} \exp \left( - \frac{\alpha \|\mathbf{f}\|^2}{2\sigma^2_{\eta_k}} \right) \tag{43}
$$

where $Z_1$ is constant, $\alpha$ is the weighting factor of the prior, $\mathbf{L}$ is a laplacian operator and $\eta_k$ is associated with a statistical distribution of regularized Laplacian prior. The objective of this prior is to compensate for the ill posed nature of SR problem by assuming some degree of smoothness of edges as a prior knowledge. Usually a 2-D Laplacian is used in a regularization framework which helps in converging to a stable solution [5]. As a result, using (39) and (43) our proposed cost function for SR can now be stated as

$$
\nabla \mathbf{f} = \arg \min_f \left[ \sum_{k=1}^{n} \left( \|g_k - S_k \hat{\mathbf{A}}_k \mathbf{f}\|^2 \right) / 2\sigma^2_{\eta_k} \right. + \left. \frac{2 \Sigma_{\mathbf{x}}^2}{\|\hat{z}_k\|^4} \left( \mathbf{f}^T \hat{\mathbf{A}}_k^T \hat{\mathbf{A}}_k \mathbf{f} \right) \right] \tag{44}
$$

$$
\hat{\mathbf{f}}_{\text{MAP}} = \arg \min_f \nabla \mathbf{f} \tag{45}
$$

where $\sigma^2_{\eta_k}$ is Gaussian noise variance.

### 4 Experiments and Results

The experiments were carried out using real world data. The video sequences were recorded at 30 fps for speed sign detection system under development at Seeingmachines. The SR image size is fixed (normally in the range of $20 \times 20$ to $32 \times 32$ due to the requirement of the speed sign recognition system. Hence the sampling matrix $S_k$ varies for each observation frame. The regularization parameter $\alpha$ in (43) is estimated after trial and error. This regularization varies the smoothness of the final image and an optimal value of $\alpha=0.67$ is chosen for most of the video sequences.

In the absence of ground truth we use relative residual ratio $\|\hat{z}_k\|$ as the measure of performance, where $r_n$ is the initial residual and $r_n$ is the current residual after nth iteration defined as [16]

$$
r_n = g_k - S_k \hat{\mathbf{A}}_k \mathbf{f}^{(n)} \tag{46}
$$

In each EM iteration, homography estimation is done in high resolution as mentioned in section III. For image registration we use forward compositional [19] approach for Lucas-Kanade Optic flow [14] to compute homography. The main reason for implementing optic flow for image registration is due to video observation data and computational efficiency over featurebased methods. The homography estimation is between the speed sign feature and not the image itself, therefore, care must be taken while selecting any homography estimation technique. The filter state graph in Fig. 4 shows consistency in x and y axis and major variation in the direction of motion along z axis. Due to quantization of image, smoothness in scale along z axis is difficult to achieve.

The convergence criteria for each EM iteration is defined as a threshold of the relative residual error $\epsilon$.

$$
\frac{\|\mathbf{f}^{(n+1)} - \mathbf{f}^{(n)}\|^2}{\|\mathbf{f}^{(n)}\|^2} < 10^{-4} = \epsilon \tag{47}
$$

The convergence plot shows a drop after 3-4 four iterations and then reaches a stability point as desired by an EM algorithm. The graph in Fig. 6 also shows that residual error possible with standard SR algorithm and improvement.

6
Figure 4. Filter states of motion parameters

Figure 5. (a) Low resolution frame (b) Bilinear Interpolation of last frame (c) SR frame after 10th EM iteration

Figure 6. EM convergence plot for speed sign video sequence

reflect that the proposed filtering frame work mitigates several issues related to multi-scale image registration. However, significant work is required in image registration for multi-scale with wide temporal varied images to make this process more resilient to changing input condition. The proposed framework is an effective measure for SR resolving images in particular, where observation data is not consistent e.g occlusion due to trees, shadows and other objects, glare of sunlight and limited dynamic range of the automotive application cameras.

5 Conclusion

In our work, we have tackled the issue of SR imaging that requires image enhancement in multi-scale and multi-resolution moving objects in video streams. The approach has addressed this issue by successive filtering of image registration parameters with EM in spatial domain. We believe that the improvement factor can be further increased with proper conditioning of the data. Although the experiments have been performed for speed sign data, the same approach is applicable to other domains that have moving objects and multi-scale video streams. The results shown, reflect that the proposed filtering frame work mitigates several issues related to multi-sale image registration. However, significant work is required in image registration for multi-scale with wide temporal varied images to make this process more resilient to changing input condition. The proposed framework is an effective measure for SR resolving images in particular, where observation data is not consistent e.g occlusion due to trees, shadows and other objects, glare of sunlight and limited dynamic range of the automotive application cameras.

References


