A novel approach for enhancing quality of video by MMSE method

Joshi Khyati[1], *M N Charel*[2], *Ketan Bhavsar*[3]

ME, EC Dept., Dr. S & *S S Ghandhy Govt. Engineering College, Surat, Gujarat, India*[1]
HOD, EC Dept., Govt Polytechnic, Ambavadi, Ahmedabad, Gujarat, India[2]
Prof., EC Dept, L J Institute of Engineering, Ahmedabad, Gujarat, India[3]

Abstract: *Super-resolution (SR) is the term used to define the process of estimating a high resolution (HR) image or a set of HR images from a set of low resolution (LR) observations. In this paper we propose a SR algorithm, based on the maximum a posteriori (MAP) framework. These algorithms utilize a new multichannel image prior model, along with the state-of-the art image prior and observation models. Numerical experiments comparing the proposed algorithms, demonstrate the advantages of the adopted multichannel approach.*

Terms—*Maximum a posteriori (MAP) framework, multichannel prior, observation model, parameter estimation, super-resolution, video applications.*

I. INTRODUCTION

Resolution enhancement of an image or of a video sequence based on multiple LR observed frames. It has importance in signal processing applications, such as remote sensing, portable global positioning system, video surveillance, medical imaging, high definition television etc [1]-[2]. The super-resolution problem is an inverse problem that requires a regularized solution. The Bayesian framework offers many advantages. In most of the Bayesian formulations single channel image prior is used. It is based on a Gaussian stationary assumption for the residuals of the local image differences [3].

In this paper, we address the video SR problem utilizing a MAP approach. One of the main contributions of this work is the introduction of a new multichannel prior that incorporates registration information between frames. Accurate registration of each pixel of adjacent frames is critical for the video SR reconstruction. An additional novelty factor of our work, consists in using the non stationary image prior which was introduced in [3] , in the SR problem. A new single frame MAP SR algorithm is derived. Furthermore, this prior was also incorporated in the video resolution enhancement algorithm, resulting in a new proposed MAP SR technique for uncompressed video.

Moreover, we also introduce a non stationary version of the aforementioned new multichannel prior within the MAP framework. In particular, we propose a novel SR algorithm that utilizes a hierarchical Gaussian type prior which assumes that the residuals of first order differences within each frame in two directions are Gaussian random variables (RVs) with zero mean and variance that is spatially varying [2]. The same modeling assumption is made for each motion compensation error between every two frames which are used in the proposed prior.
In order to deal with the resulting over parameterization, the spatially varying variances (at each pixel location) are considered RVs and a Gamma hyper prior is imposed on each of them.

This paper is organized as follows. Section 2 describes the appropriate mathematical background on image prior and observation models used. Section 3 introduces a MAP problem formulation for the SR of uncompressed video for each one of the proposed models, along with the realizations of the corresponding algorithms. In section 4 describes the proposed method. Section 5 shows the experimental results. Finally, section 6 presents the conclusion.

II. MATHEMATICAL BACKGROUND

A. Observation Model

It is the relation between the MN×1 LR observation frame g_i and its LMLN×1 (L denotes the resolution enhancement factor) HR counterpart f_i is given in matrix vector form by

g_i = DHf_i+n_i for i = 1,2……P (1)

Where D is the MN×LMLN down sampling matrix, H is the LMLN×LMLN known blurring matrix, n_i of size MN×1, represents additive white Gaussian noise (AWGN) term, and P represents the total number of frames used.[1]

Equation (1) can be written as

$$\bar{\boldsymbol{g}} = \bar{D}\bar{H}\bar{f} + \bar{n} \quad (2)$$

where

$$\bar{g} = [g_{k-m}^T, \ldots\ldots, g_k^T, \ldots\ldots, g_{k+n}^T]^T$$
$$\bar{f} = [f_{k-m}^T, \ldots\ldots, f_k^T, \ldots\ldots, f_{k+n}^T]^T$$
$$\bar{n} = [n_{k-m}^T, \ldots\ldots, n_k^T, \ldots\ldots, n_{k+n}^T]^T \quad (3)$$

n indicate the number of frames used in the forward temporal direction and m indicate the number of frames used in the backward direction with respect to the kth frame (n+m+1 = P), T denotes the transpose of a matrix or vector, and

$\bar{H}$ = diag { H,…,H,…,H}, $\bar{D}$ = diag {D,…,D,…,D} (4)

are respectively of dimensions PLMLN×PLMLN and PMN×PLMPN.

B. Image Prior Model

This model is based on the assumption that the local image differences follow the same statistics all over the image. This assumption induces a prior for the image which has been used extensively in image recovery problems. It is very effective because it contains only one parameter resulting easy calculation. Because of one parameter it cannot capture the local properties of image. One way to bypass this difficulty is to introduce a prior with a specially varying precision parameter. In order to avoid over parameterization problem, all of them are assumed to originate from the same conjugate Gamma probability density function (pdf)[1].

In this work, we extend these ideas to the video SR problem. We propose a new multichannel image prior model. The model takes into account both within frame smoothness and between frame smoothness. In this case the regularization parameters are assumed to be spatially varying RVs and a Gamma hyper prior is imposed on each of them.

We introduce a hierarchical Gaussian prior to model both its within channel and between channel FODs as $\varepsilon_j^d = Q^d f_j$, where d = 1, 2 denotes the directions of 0^o *and* 90^o, respectively, and Q^d is the LMLN×LMLN directional difference operators for each frame.

Assuming that the errors at each direction and at each pixel locations are independent we obtain the following improper pdf

$$\mathrm{P}(f_j|\alpha_j) \propto \prod_{d-1}^{2} \left(\prod_{\mu=1}^{LMLN} (\alpha_{j,\mu}^d)^{\frac{1}{4}}\right) \times \exp\left(-\tfrac{1}{2}((Q^d f_j)^T A_j^d Q^d f_j\right) \quad (5)$$

The independency assumption is based on the fact that at each pixel location an edge can occur in any direction independently of what happens in adjacent pixels and it also makes subsequent calculations tractable[3].

A Gamma hyper prior is placed on each inverse variance $\alpha_{j,\mu}^d$, according to

$$\mathrm{P}(\alpha_{j,\mu}^d;\ v_j^d, l_j^d) \propto (\alpha_{j,\mu}^d)^{\frac{l_j^d}{2}} \exp\left(-v_j^d(l_j^d - 2)\ \alpha_{j,\mu}^d\right) \quad (6)$$
(20)

Where $\mathrm{E}[\alpha_{j,\mu}^d] = l_j^d(2v_j^d(l_j^d - 2))^{-1}$,$\mathrm{Var}[\alpha_{j,\mu}^d] = l_j^d(2v_j^d)^2(l_j^d - 2)^2)^{-1}$ and v_j^d, l_j^d are the scale and shape parameters, respectively, per frame and per direction [37], [39]. The rationale for using a Gamma prior on each inverse variance is threefold.

It produces compact equations for the inverse variances. Assuming also that the variances of the Gaussians modeling the residual errors are statistically independent, we have from (7)

$$P(\alpha_{j,\mu}^{d};\ v_j^d, l_j^d) = \prod_{d=1}^{2} \prod_{\mu=1}^{LMLN} (\alpha_{j,\mu}^{d};\ v_j^d, l_j^d) \quad (7) \qquad (21)$$

Where $v_j = [v_j^{d=1}, v_j^{d=2}]^T$ and $[l_j^{d=1}, l_j^{d=2}]^T$. This assumption allows these variances to scale down the differences of adjacent pixels not only in smooth areas, but also in regions of image discontinuities (e.g., edges). Therefore, combining (5) and (7), we obtain

$$P(f_j, \alpha_j; v_j, l_j) = p(f_j|\alpha_j)p(\alpha_j; v_j, l_j) \quad (8) \qquad (22)$$

Representing the proposed within channel non-stationary prior.

III PROPOSED ALGORITHM

In this paper the MAP estimate is utilized for recovering HR information from a sequence of low resolution observations. The MAP formulation for the SR problem of uncompressed video, is specified for each one of the proposed algorithms[1]. It is given by

$$\bar{f}, \hat{\alpha}^{total}, \hat{\beta}^{total} = \mathrm{argmax}_{\bar{f}, \hat{\alpha}^{total}, \hat{\beta}^{total}}\, p\,(\bar{f}, \hat{\alpha}^{total}, \hat{\beta}^{total} \mid \bar{g}|; \bar{\theta})$$
$$= \mathrm{argmax}_{\bar{f}, \hat{\alpha}^{total}, \hat{\beta}^{total}} \frac{p\,(\bar{g}, \bar{f}, \hat{\alpha}^{total}, \hat{\beta}^{total}; \bar{\theta})}{p(\bar{g})} \quad (9)$$

Where $\bar{\theta}$ is the column vector that contains all the parameters $v^{total}\, l^{total} \tau^{tatal} \xi^{tatal}$ along with the noise inverse variance parameters, whereas the RVs $\alpha^{total}\beta^{total}$ are jointly estimated with the HR frames $\bar{f}$ Instead of the maximization in (22) the negative of the logarithm of p($\bar{f}$, $\alpha^{total}\beta^{total}|\bar{g};\bar{\theta}$) with respect to $\bar{f}$, $\alpha^{total}\beta^{total}$ and to the noise parameter is typically minimized. We propose a formulaton of HR problem and derive the MAP algorithm[5].

In this case no MF is used. More specifically, each frame in the sequence is recovered independently from the other frames. The fidelity pdf then is defined as

$$P\,(g_i|f_{i;\gamma_i}) \propto \gamma_i^{\frac{MN}{2}} \exp\,(-\frac{\gamma_i}{2} \| g_i - DHf_i \|^2) \quad (10)$$

Where the parameter γ_i^{-1} is the acquisition noise variance and the prior model is defined by (9).
The objective functional to be minimized for obtaining a MAP estimate is given by

$$J_{MAP}\,(f_i, \alpha_i | g_i; \gamma_i, v_i, l_i) \propto -2 \log p\,(g_i\, f_i, \alpha_i; \gamma_i, v_i, l_i)$$
$$= -2 \log p\,(g_i\, f_i, \alpha_i; \gamma_i)$$
$$= -2 \log p\,(f_i, \alpha_i; v_i, l_i) \quad (11)$$

Taking its partial derivatives with respect to $f_i \alpha_i\ and\ \gamma_i$ and setting them equal to zero results in following linear system of equations

$$\alpha_{j,\mu}^{d} = \frac{(\frac{1}{4} + \frac{1}{2}(l_i^d - 2))}{(\frac{1}{2}(\varepsilon_{i,\mu}^{d})^2 + v_i^d(l_i^d - 2))} \quad (12)$$

$$\gamma_i = \frac{MN}{||g_i - DHf_i||^2} \quad (13)$$

$$(H^T D^T DH + \gamma_i^{-1} \textstyle\sum_{d=1}^{2} Q^{d^T} A_i^d Q^d)\hat{f}_i = H^T D^T g_i. \quad (14)$$

For stationary algorithm, (12) and (14) reduce respectively to the following expressions:

$$\alpha_i = \frac{(LMLN - 1)}{||Qf_i||^2} \quad (15)$$

And

$$(H^T D^T DH + \frac{\alpha_i}{\gamma_i} Q^T Q)\,\hat{f}_i = H^T D^T g_i \quad (16)$$

While (13) remains the same. These equations are obtained based on the previous analysis and more specifically on the fact that as $l_i^d \to \infty$, E[$\alpha_{i,\mu}^d$] $\to (2v_i^d)^{-1}$ and Var[$\alpha_{i,\mu}^d$] $\to 0$ and consequently E[$\alpha_{i,\mu}^d$] equals α_i (stationary prior model) and $A_i^d = \alpha_i$I.

The algorithm is illustrated step by step as follows.
Single frame SR

Given an initial estimate for each HR image
While convergence criterion is not met do

1) Estimate $\alpha_{i,\mu}^{d}$ using (12)- nonstationary prior – or α_i using (15)- stationary prior-
2) Estimate γ_i using (13)
3) Based on steps 1) and 2), estimate HR image using (14) non stationary prior or (16) stationary prior

IV PROPOSED METHOD

Minimum Mean Squared Error (MMSE) Methods

Given a model of the imaging process, any high resolution image can be passed through the model to produce a simulated low resolution output. In matrix notation, W (the imaging model) is known, so an estimate $\boldsymbol{y_e}$, can be produced from the high resolution image z (a column vector of data points) via:

$$\boldsymbol{y_e} = \mathrm{Wz} \qquad (17)$$

MMSE methods return the high resolution image for which the simulated output is closest to the actual observed data (in the MMSE sense). Mathematically, the output $\hat{\boldsymbol{z}}$ is given by

$$\hat{\boldsymbol{z}} = \boldsymbol{arg}\,\mathbf{min}_{\boldsymbol{z}} \quad (\boldsymbol{y}-\boldsymbol{y_e})^{\boldsymbol{T}}(\boldsymbol{y}-\boldsymbol{y_e})$$

$$= \boldsymbol{arg}\,\mathbf{min}_{\boldsymbol{z}}(\boldsymbol{y}-\boldsymbol{W_z})^{\boldsymbol{T}}(\boldsymbol{y}-\boldsymbol{W_z}) \qquad (18) \qquad\qquad (10)$$

Where, y is a column vector of the observed pixel values. This criterion is favourable, but it is also incomplete: the inverse imaging problem is ill-posed, so many high resolution images exist which minimize the error.

MMSE terms, along with a regularization term, constitute a composite cost function that is minimized by an iterative optimization. Since the regularization term is an explicit cost term, it can easily be understood and modified. The regularization term used below is a linear smoothness constraint. The composite cost function is thus linear, but is still minimized iteratively because of the huge number of variables. This Regularized MMSE (RMMSE) method was coded for the algorithm comparison, and a complete description is given below.

A. Technical Description:

RMMSE algorithm seeks the high resolution image z that minimizes a cost function, C(z):

$$\hat{\boldsymbol{z}} = \mathbf{min}_{\boldsymbol{z}}\,\boldsymbol{C(z)} \qquad (19)$$

The cost function is defined as:

$$\mathrm{C(z)} = \frac{1}{2}(\boldsymbol{y}-\boldsymbol{W_z})^{\boldsymbol{T}}(\boldsymbol{y}-\boldsymbol{W_z}) + \frac{\lambda}{2}(\boldsymbol{A_z})^{\boldsymbol{T}}(\boldsymbol{A_z}) \qquad (20)$$

Where, y is a column vector containing the pixel values of all the low resolution data. z is a column vector containing the pixel values of the high resolution estimate. The size of z is determined by the resolution increase. W is a matrix determined by the system transfer function. $\boldsymbol{\lambda}$ is a scalar parameter that determines the relative weight of the regularization term and typically has a value around 0.05-0.1. A is a matrix of weights that produce a regularization term. For instance, if $\boldsymbol{z_a}$ is a given pixel in the high resolution image, and $\boldsymbol{z_b}$, $\boldsymbol{z_c}$, $\boldsymbol{z_d}$ and $\boldsymbol{z_e}$ are the 4 cardinal neighbours of $\boldsymbol{z_a}$, then row a of A would read:

This regularization term has lower cost for a smooth image than a rapidly varying image. With the cost function thus defined, it is minimized with a standard iterative optimization. This particular gradient descent method operates in the following manner:

- The gradient of the cost function is calculated
- The line extending in the negative gradient direction is searched for lowest cost (line minimization method)
- z is updated to this point of lowest cost
- The next iteration begins

The initial value of z used to start the iteration was chosen to be the composite created by averaging all low resolution frames. The calculations are straightforward but algebraically intensive. The iterative update procedure is tersely written as

$$\hat{\mathbf{z}}^{n+1} = \hat{\mathbf{z}}^{n} - \boldsymbol{\varepsilon}^{n}\mathbf{g}^{n} \qquad (21)$$

Where the gradient, $\boldsymbol{g}^{\boldsymbol{n}}$, is given by

$$\boldsymbol{g}^{\boldsymbol{n}} = \boldsymbol{W}^{\boldsymbol{T}} (\boldsymbol{W}_{\boldsymbol{z}} - \mathrm{y}) + \lambda\, \boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}_{\boldsymbol{z}} \qquad (22)$$

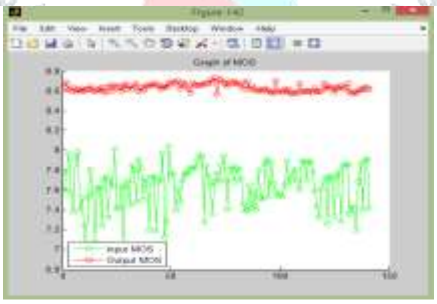
We calculate $\boldsymbol{\varepsilon}$ by minimizing C ($\boldsymbol{z}^{\boldsymbol{n}} - \boldsymbol{\varepsilon}^{\boldsymbol{n}}\boldsymbol{g}^{\boldsymbol{n}}$) with respect to $\boldsymbol{\varepsilon}$, yielding:

$$\boldsymbol{\varepsilon}^{\boldsymbol{n}} = \frac{(\mathbf{W}\boldsymbol{g}^{\boldsymbol{n}})^{\boldsymbol{T}}(\boldsymbol{W}_{\boldsymbol{z}} - \mathbf{y}) + \boldsymbol{\lambda}\,(\mathbf{A}\boldsymbol{g}^{\boldsymbol{n}})^{\boldsymbol{T}}\boldsymbol{A}_{\boldsymbol{z}}}{(\mathbf{W}\boldsymbol{g}^{\boldsymbol{n}})^{\boldsymbol{T}}(\mathbf{W}\boldsymbol{g}^{\boldsymbol{n}}) + \boldsymbol{\lambda}\,(\mathbf{A}\boldsymbol{g}^{\boldsymbol{n}})^{\boldsymbol{T}}(\mathbf{A}\boldsymbol{g}^{\boldsymbol{n}})} \qquad (23)$$

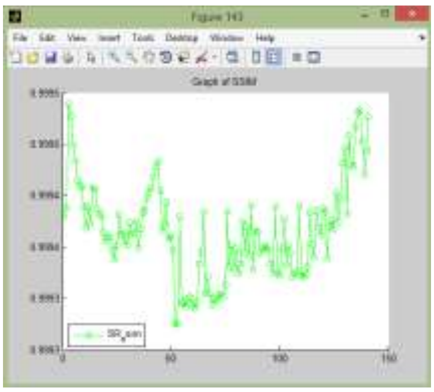
The above derivation is fairly general: the PSF need not be space-invariant, nor must it remain constant from frame to frame. For this study, we're assuming a space-invariant PSF, which allows some significant simplifications[6] and computational savings. These computational efficiencies can be found in the Matlab code.

V. EXPERIMENTAL RESULTS

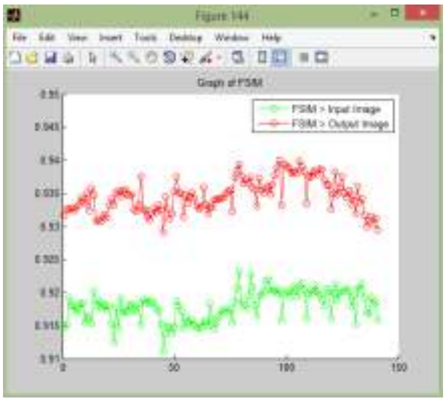
(a) Original LR image frame, (b) SR frame using proposed algorithm



(c)



(d)



(e)

(c) result of MOS parameter using proposed algorithm, (d) result of SSIM parameter using proposed algorithm, (e) result of FSIM parameter using proposed algorithm.

VI. CONCLUSION AND FUTURE RESEARCH

In this paper, we presented a MAP approach that utilizes a new multichannel image prior, and applied to the digital video SR problem. We tested our new algorithm for certain experimental cases and also using real video data.

The proposed algorithms were tested for different parameters PSNR, SSIM and MOS. The experimental result showed that the proposed algorithm performs better in terms of both SNR and visual quality.

Our algorithm provided a superior performance. In future we can further improve the proposed algorithm by introducing new model which can be applied for real time applications.

REFERENCE:

[1] Stefanos P. Belekos, Member IEEE, Nikolaos P. Galatsanos, Senier Member, IEEE, and Aggelos K Katsaggelos, Fellow, IEEE, " Maximum a Pestiriori Video Super-Resolution Using a New Multichannel Image Prior" IEEE Transactions on Image Processing, June 2010 Vol No 6.

[2] Sune Hogild Keller1, Fran, cois Lauze2 and Mads Nielsen 1,2, "Motion Compressed Video Super Resolution" IEEE Trans. On Image Processing, 608-618, 2010.

[3] Neil alldrin "Super Resolution" IEEE Transaction on Image Processing, 1167-1183, 2009

[4] V.Hari Kumar, Prakash P Gajjar, Manjunath V. Joshi , Mehul S. Raval, " Multi Resolution Image Fusion : Use of Compressive Sensing and Graph Cuts" , IEEE Journal of selecterd topics in applied earth observation, 2013.

[5] Harikumar V, Manjunath V. Joshi, Mehul S. Raval and Prakash P. Gajjar, "Multi resolution Image Fusion Using Compressive Sensing and Graph Cuts" , IEEE Journal of selected topics in applied earth observation, 2012.

[6] Hasan Demirel and Gholamreza AnbarJafari , " Discrite Wavelet Transform- Based Satellite Image resolution Enhancement ", IEEE Transactio on Geoscience and Remote Sensing , Vol 49, No.6 , June 2011

[7] Stefanos P. Belekos, Nikolaos P. Galatsanos, and Aggelos K. Katsaggelos, " Maximum a Posteriori Video Super-Resolution with a new Multichannel Image Prior " 16th European Signal Processing Conference, 2008

[8] : Sune Høgild Keller*, Franc¸ois Lauze and Mads Nielsen , " Video Super-Resolution using Simultaneous Motion and Intensity Calculation" , 2011

[9] Title: Anil Bhagat1, Balasaheb Deokate2, " Improve Image Quality at Low Bit rate using Wavelet based Fractal Image Coder.", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (ISO 3297: 2007 Certified Organization) Vol. 2, Issue 8, August 2013

[10] Cao Bui-Thu Tuan Do-Hong , Thuong Le-Tien , Hoang Nguyen-Duc "Scene-based video super-resolution with minimum mean square error estimation" IEEE advanced technologies for communication(ATC) , 2011 international conference, pg- 48 – 52

[11] Merugu Suresh, Dr Kamal Jain Department of Civil Engineering, Indian Institute of Technology Roorkee, "To Generate High Resolution Images (Conventional Subpixel) from Low Resolution Satellite Images",(introduction)

[12] Suma Chappidi, Sandeep Kumar Mekapothula, " A Novel Method for Enhancing Satellite & Land Survey Images Using Color Filter Array Interpolation Technique (CFA)", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 5, September-October 2014, pg-217-221 (intro)