Simultaneous Methods of Image Registration and Super-Resolution Using Analytical Combinational Jacobian Matrix

Hossein Rezayi*
Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
hrezayi2001@yahoo.com
Seyed Alireza Seyedin
Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
seyedin@um.ac.ir

Received: 13/Mar/2015 Revised: 05/Aug/2015 Accepted: 09/Aug/2015

Abstract

In this paper we propose two simultaneous image registration (IR) and super-resolution (SR) methods using a novel approach in calculating the Jacobian matrix. SR is the process of fusing several low resolution (LR) images to reconstruct a high resolution (HR) image; however, as an inverse problem, it consists of three principal operations of warping, blurring and down-sampling that should be applied sequentially to the desired HR image to produce the existing LR images. Unlike the previous methods, we neither calculate the Jacobian matrix numerically nor derive it by treating the three principal operations separately. We develop a new approach to derive the Jacobian matrix analytically from the combination of the three principal operations. In this approach, a Gaussian kernel (as it is more realistic in a wide range of applications) is considered for blurring, which can be adaptively resized for each LR image. The main intended method is established by applying the aforementioned ideas to the joint methods, a class of simultaneous iterative methods in which the incremental values for both registration parameters and HR image are obtained by solving one linear system of equations per iteration. Our second proposed method is formed by applying these ideas to the alternating minimization (AM) methods, a class of simultaneous iterative methods in which the incremental values of registration parameters are obtained after calculating the HR image at each iteration. The results show that our proposed joint and AM methods are superior to the recently proposed methods such as Tian's joint and Hardie's AM methods respectively.

Keywords: Super-Resolution; Image Registration; Jacobian Matrix; Combinational Coefficient Matrix; Joint Methods.

1. Introduction

Super-resolution (SR) refers to a series of techniques that integrate the information of a low resolution (LR) image sequence captured from a scene to produce a high resolution (HR) image with higher quality. Each LR image frame is required to have partial unique information or it will not have any positive effect on the final HR image. This partial unique information can be obtained in a number of ways such as camera/scene movement or zooming. There are various SR techniques which have been developed and reviewed in the literature including [1]-[4]. Generally SR techniques include three phases either implicitly or explicitly: 1) image registration (IR), 2) image interpolation, and 3) image deblurring and denoising [1]. In a small category of the techniques such as interpolation-based methods, these phases are performed separately [5]-[9]. To overcome the intensive presence of error propagation in these methods, a majority of other SR techniques attempt to perform the last two phases in an integrative phase called image reconstruction. However, an important source of error is the inaccuracy of registration parameters. Therefore, to further prevent the error propagation, a large group of SR techniques have been developed recently that deal with the inaccuracy of registration parameters. Some of these techniques utilize median estimator to reduce the artifacts caused by errors and outliers of registration parameters [10], [11]. Some others use Bayesian methods in which the unknowns (including registration parameters) are treated as stochastic variables [12]-[14]. In Tipping's method [12], marginalization is applied to HR image but in Pickup's method [13], it is applied to registration and blurring parameters. Nevertheless, the latter provides a wide range of various priors (regularizations) to select, but in both of them, IR and image reconstruction are implemented in relatively separate steps without persistent interaction. It seems that such values are not reliable enough. Babacan [14] extends the Pickup's method to consider hyperparameters (such as the regularization parameter) as stochastic variables, and as an alternating minimization (AM) method, establishes persistent interaction between the estimate of the reconstructed HR image, registration parameters and hyperparameters. AM methods are a class of iterative SR techniques in which the HR image and registration parameters are improved in two consecutive steps at each iteration [15]. A group of AM methods use expectation-maximization (EM) to estimate the HR image (in the expectation phase) and registration parameters (in the maximization phase) iteratively

* Corresponding Author

[16],[17]. The AM methods, nevertheless, may lead to suboptimal solutions [18].

There is another category of SR techniques which are also iterative similar to AM methods, but in this category, HR image and registration parameters are not calculated separately at each iteration [19]-[22]. A nonlinear cost function was used by Chung et al. [19] to estimate HR image and registration parameters. Using Euler–Lagrange necessary conditions for the cost function, they derived a nonlinear system of equations, proposing three methods for its solution. Their first method (called decoupled) resembled an AM method, but their second method (called partially coupled) was a kind of variable projection (VP) method [18]. A similar method was also proposed by Robinson et al. [20] where they used a similar nonlinear cost function to derive the maximum likelihood (ML)/maximum a posteriori (MAP) solution for the HR image. After substituting this solution of HR image into the cost function, the reduced cost function [19] was minimized with respect to the reminder of unknowns, i.e. registration parameters. Finally, these registration parameters were used to obtain the final HR image. Chung et al. [19] attempted to solve the nonlinear system of equations through Gauss-Newton algorithm in their third method. This led to the development of a new class of methods called fully coupled. In these techniques, which are referred to as joint methods in this paper, the incremental values of the HR image and registration parameters are jointly calculated in only one system of equations. He et al. [21] used a similar cost function by linearizing it at existing current values for HR image and registration parameters using Taylor series approximation. In this linearization, they obtained the Jacobian matrix analytically (in contrast to some methods like [12], [13] where it is calculated numerically). Finally, this linear system of equations (in terms of incremental values of HR image and registration parameters) has been solved through a conjugate gradient (CG) optimization algorithm. They have used Euclidean motion model and it has been extended to the similarity motion model by Tian et al. [22].

In all of these joint methods [19]-[22], the blurring is assumed to be the same for all LR images, which is impractical in many applications. In reality, when the motion model is more complex than Euclidean motion model, the size of blurring function is no longer the same for all LR images [23]. This has not been considered in the [19] and [22] methods, which have not restricted their motion model to the Euclidean motion model. However, in the methods proposed in [21] and [22], the convergence to the global solution is more probable (especially when the initial values are close enough to optimal values) but the derivation of Jacobian matrix is based on bilinear interpolation of warped pixels [19], [21], [22]. This may introduce the restriction in which only four neighboring pixels are effective in determining the values of warped pixels [21].

As an inverse problem, SR dictates three principal operations that should be applied sequentially to the original HR image to produce the corresponding LR images. These operations are image warping, image blurring and image down-sampling (for brevity they are referred to as principal operations). In many approaches, the principal operations are treated separately and in others they are combined into a unit operation [2], [12], [13], [23]. The advantages of this combination include the possibility of incorporating the pixels in any arbitrary neighboring radius to derive the warped pixels without changing the framework of the problem and employing a new interpolation method. Another advantage of combining the principal operations is error propagation reduction because all three operations are performed in one stage. Moreover, it allows having an adaptive kernel for blurring (blurring is treated as a function of zooming, which can be different for each LR image).

This paper focuses on the joint methods [21], [22] and treats the three principal operations in the inverse problem in a combinational form as [12] and [13]. Then, a new joint method based on this combinational form is proposed. In contrast to [12] and [13], the Jacobian matrix is not calculated numerically and unlike common joint methods [21], [22], the Jacobian matrix is not derived by treating the three principal operations separately. In the proposed method, the bilinear interpolation has not been used in the warping operation and its derivative. Moreover, the same blurring for all LR images [19]-[22] has not been considered. We developed a new approach to derive the Jacobian matrix analytically based on the combinational form of the three principal operations. In this regard, a Gaussian kernel blur (as is more realistic) was adopted the radius of which is adaptive to each LR image. We also used a bilateral total variation (BTV) regularization [11], which incorporates the eight directions of each pixel in the cost function. Although the main goal of this paper is to develop a new joint method of IR and SR using the combinational coefficient matrix and analytical combinational Jacobian matrix, the application of these concepts to the framework of AM methods gives rise to new method which will be discussed later. In this paper, the similarity motion model is used (which consists of translation, rotation and zooming) similar to [22].

The rest of this paper is organized as follows. In Section 2, the problem formulation including notation of SR problem, Gaussian kernel blur and combinational coefficient matrix are introduced. The proposed iterative joint and AM methods are developed in Section 3. In Section 4, experimental results on simulated and real images are presented. Conclusion and future works are discussed in Section 5.

2. Problem Formulation

2.1 Super Resolution Notations

Let us consider a series of K discrete LR images g_k of size $M_g \times N_g$ where $1 \leq k \leq K$. The lexicographically

ordered LR images are denoted by column vectors g_k and all these vectors are stacked in one column vector, i.e. $g = [g_1^T, \dots, g_K^T]^T$. The purpose of the SR technique is to reconstruct the original HR image f of the size $M_f \times N_f$ using existing LR images as well as some prior information about the original HR image. The lexicographically ordered original HR image is presented by column vector f. Here, it is assumed that the decimation factor is the same in both vertical and horizontal directions ($\rho = M_f/M_g = N_f/N_g$). Each LR image g_k is obtained by applying the three principal operations to the original HR image f as follows (generative model):

$$g_k = DH_kB(a_k)f + n_k \quad (1)$$

Where n_k is the column vector of additive white Gaussian noise, D is the down-sampling operator (which is realized as a $M_gN_g \times M_fN_f$ matrix), H_k is the blurring operator (which is realized as a $M_fN_f \times M_fN_f$ matrix) and $B(a_k)$ is the warping operator (which is realized as a $M_fN_f \times M_fN_f$ matrix) [14], [21], [22]. a_k is the vector of unknown registration parameters used for warping the grid of original HR image f (called reference grid) onto the up-scaled grid of *k* th LR image. Practically one of the existing LR images (the first LR image in this paper) is selected as the reference image and the up-scaled grid of the reference image is considered as the reference grid. Generally, in all simultaneous IR and SR methods like AM and joint methods, it is assumed that initial and imprecision values of registration parameters can be provided by some IR techniques. In this paper, enhanced correlation coefficient (ECC) method [24] is used for the IR.

The combination of the three principal operations can be considered as a unit combinational operation, which is realized by a matrix $W_k(a_k)$ of the size $M_gN_g \times M_fN_f$ as follows:

$$g_k = W_k(a_k)f + n_k \quad (2)$$

This paper adopts the similarity motion model with four degrees of freedom (zooming, rotation, vertical and horizontal translation). Hence, $a_k = [h_{1k}, \dots, h_{4k}]$ where $h_{1k}, \dots, h_{4k}$ are the variable elements of the kth 3x3 homogenous matrix [25] with the current motion model, as follows:

$$\mathbf{M}_k = \mathbf{M}(\boldsymbol{\alpha}_k) = \begin{bmatrix} h_{1k} & -h_{2k} & h_{3k} \\ h_{2k} & h_{1k} & h_{4k} \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

It is worth noting that the above homogenous matrix can be decomposed into pure translation, rotation and zooming matrices respectively as $M_k = M^T(u_k)M^R(\theta_k)M^Z(z_k)$ [25] in which $u_k = [u_{xk} \dots u_{yk}]^T = [h_{3k} \dots h_{4k}]^T$, $\theta_k = tan^{-1}(h_{2k}/h_{1k})$ and $z_k = (h_{1k}^2 + h_{2k}^2)^{0.5}$.

Equation (2) shows a series of relations between the original HR image and each LR image. These relations can be written as one equation as follows:

$$g = Wf + n \quad (4)$$

where $W = [W_1(a_1)^T, \dots, W_k(a_k)^T]^T$ is called combinational coefficient matrix and $n = [n_1^T, \dots, n_K^T]^T$.

2.2 Blur Considerations

Consistent with the literature, in this paper Gaussian kernel, which is more realistic, has been used to model the blur caused by the atmosphere turbulence and camera lens, and the motion blur has been excluded. Usually, the blur is assumed to be isotropic in the imaging plane [23]. When the motion model is similarity, the kernel of back-projected blur into the scene plane will be isotropic too. However, the greater the distance of a scene plane from the image plane (or less zoom is applied) the more extensive is the area encompassed in the scene to contribute to the blurring. Therefore, when the LR images are registered to the reference image, they have isotropic Gaussian blur, but possibly with different radiuses in the reference image [23].

2.3 Combinational Coefficient Matrix

Three methods of computing the combinational coefficient matrix has been addressed by Capel [23]. The simplest one is the directly computing the warping, blurring and down-sampling matrices separately and then calculating their multiplication (separately applying the principal operations) as (2) has been derived from (1). However this method involves large amount of memory usage. Additionally, since the operations are implemented separately the error propagation can occur. Although in the Capel's second and third methods the principal operations are not implemented separately, they have complex implementations. Another method has been proposed in [12] which is theoretically very similar to the Capel's third method but it has simpler implementation. This method which has also been used in some other papers (e.g. [13]) will be described in the reminder of this section and applied in this paper.

The elements of *i*th row of the $W_k(a_k)$ are the coefficients of linear combination of f required to generate the gray scale value of *i*th pixel of g_k. Thus, the sum of these coefficients should be equal to one. As such, the elements of $W_k(a_k)$ are calculated as:

$$W_{ijk}(\boldsymbol{\alpha}_k) = \frac{\widetilde{W}_{ijk}(\boldsymbol{\alpha}_k)}{\sum_{j'=1}^{M_fN_f} \widetilde{W}_{ij'k}(\boldsymbol{\alpha}_k)} \quad (5)$$

where the elements $\widetilde{W}_{ijk}(a_k)$ are obtained as [12]:

$$\widetilde{W}_{ijk}(a_k) = \exp\left\{-\frac{1}{2\sigma^2|M_k|}(v_j - s_k(u_i)^T(v_j - s_k(u_i))\right\} \quad (6)$$

Here $v_j = [v_{xj} \;\; v_{yj}]^T$ is the position of *j*th pixel of the original HR image in the reference grid, $u_i = [u_{xi} \;\; u_{yi}]^T$ is the position of *i*th pixel of *k*th LR image and $s_k(u_i) = [s_{xk}(u_i) \;\; s_{yk}(u_i)]^T$ is its transformation through the motion model, which is characterized by a_k with respect to the reference grid:

$$s_{xk}(\mathbf{u}_i) = h_{1k}u_{xi} - h_{2k}u_{yi} + h_{k3} \quad (7)$$

$$s_{yk}(u_i) = h_{2k}u_{xi} + h_{1k}u_{yi} + h_{4k} \quad (8)$$

Actually, $s_k(u_i)$ is the center of isotropic Gaussian kernel after projection to the reference grid. σ is the standard deviation of isotropic Gaussian kernel of the blur and $|\bullet|$ is the determinant operator.

It is worth noting that in the combinational generative models used by Capel [23] and Tipping et.al [12], Gaussian interpolation method [26] has been used implicitly instead of bilinear interpolation method, which is explicitly common in several SR techniques, e.g. [14], [19], [21] and [22]. This is shown in Fig.1 in which both bilinear and Gaussian interpolation methods have been displayed from a new aspect. As shown in this schematic view, the reverse mapping is adopted in the image warping implementation. Hence, instead of warping the HR pixels, the LR pixels are warped in the reference grid. Then, the neighboring HR pixels/samples around a warped LR pixel (the center of kernel in Fig.1) are assigned some weights equal to the height of the point in which the kernel and the neighboring sample intersect. As shown in Fig.1 (a), in the bilinear interpolation, at most four samples can be included in the domain of the square pyramid kernel, but there is not such limitation in the Gaussian interpolation (Fig.1 (b)).

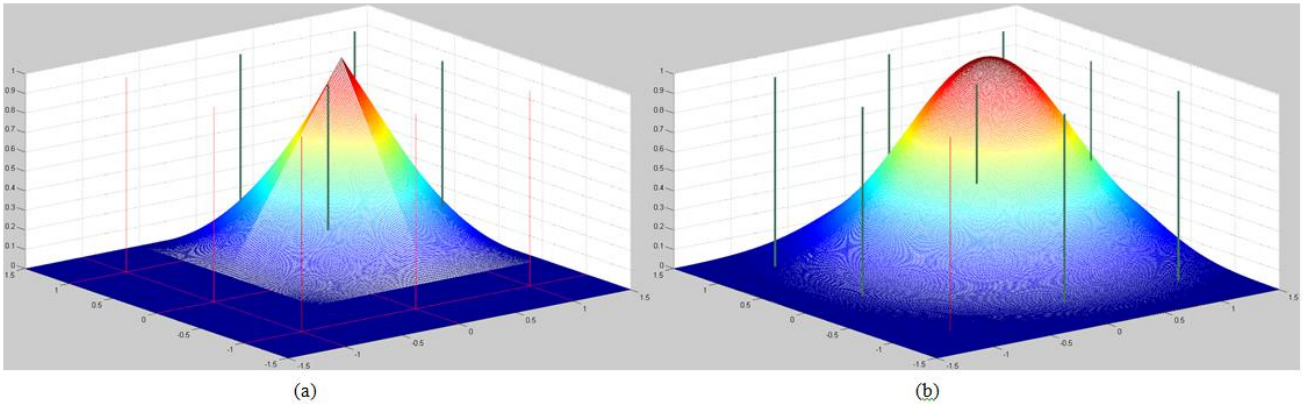

Fig. 1. Two kernels for interpolation; thick vertical lines show the samples located in the domain of the kernel. (a) Square pyramid kernel used in the bilinear interpolation, (b) Gaussian kernel used in the Gaussian interpolation.

3. Development of New Iterative Simultaneous Methods

3.1 Cost Function of the New Joint Method

As super resolution is an ill-posed problem, there are infinite or instable solutions which can satisfy Equation (4). Therefore, to make the solution unique and stable, prior information is necessary. In the joint methods, both original HR image and registration parameters are unknown. TV of the HR image is an important regularization which is often used in SR techniques as the prior information. For the registration parameters, a simple Tikhonov regularization, i.e. the minimum energy, has been used. Given the generative model of LR images expressed in (4) and considering the mentioned regularizations, the framework of the cost function for the joint method is [19], [21], [22]

$$E(f,a) = \|g - Wf\|^2 + \lambda T_V(f) + \beta R(a) \quad (9)$$

where $a = [a_1^T, \ldots, a_K^T]^T$, λ and β are regularization parameters, $T_V(f)$ is TV of the HR image. Similar to Farsiu's [11] and Tian's [22] methods, BTV is used in our cost function, but here it encompasses all eight neighboring pixels to reduce edge penalization in any directions

$$T_V(f) = \sum_{\substack{m=-1 \\ |m|+|n|\neq 0}}^{1} \sum_{n=-1}^{1} \sum_{ij} \frac{|f(i+m, j+n) - f(i,j)|}{\sqrt{m^2+n^2}} \quad (10)$$

Because $T_V(f)$ is a nonlinear function of f using half-quadratic scheme [27] and fixed-point techniques [28], it can be written as:

$$T_V(f) = \sum_{\substack{m=-1 \\ |m|+|n|\neq 0}}^{1} \sum_{n=-1}^{1} \sum_{ij} \frac{|f(i+m, j+n) - f(i,j)|^2}{\mu_{m,n}} \quad (11)$$

Where $\mu_{m,n}$ is calculated in the previous iteration as follows:

$$\mu_{m,n} = \sqrt{(m^2+n^2)\{(f(i+m, j+n) - f(i,j))^2 + \gamma} \quad (12)$$

and γ is the small positive value to ensure $\mu_{m,n}$ is nonzero. $T_V(f)$ can be expressed in a matrix-vector form as [22]:

$$T_V(f) = f^{\mathrm{T}} L^{\mathrm{T}} L f = f^{\mathrm{T}} T f = L f^2 \quad (13)$$

$R(a) = \|a - \bar{a}\|^2$ is the Tikhonov regularization for registration parameters and $\bar{a}$ is a vector containing the average values of registration parameters during all previous iterations. This method of $\bar{a}$ selection can help reduce

resonance around the final values of registration parameters. Hence, the cost function (9) can be expressed as:

$$E(f,a)=\left\|[r(a,f)^{\mathrm{T}} \quad \sqrt{\lambda}(Lf)^{\mathrm{T}} \quad \sqrt{\beta}(a-\bar{a})^{\mathrm{T}}]^{\mathrm{T}}\right\|^2 \qquad (14)$$

where $r(a,f)=g-Wf$ is called residual vector.

To estimate the unknown HR image and registration parameters, the cost function should be minimized. Although this optimization problem is convex with respect to f, it is nonconvex in terms of a because the term $r(a,f)$ is nonlinear with respect to a. To alleviate this problem, linear approximation has been used for $r(a,f)$. Linear approximation requires initial values for unknowns. As mentioned earlier, initial values for registration parameters may be obtained using registration techniques such as ECC algorithm [24]. Given these initial values for registration parameters, the initial value for HR image can be obtained using the simple interpolation-based method [5], the approach used by [21] or a simple up-scaling of the reference image. If Δf and Δa are incremental values for f and a then $r(a+\Delta a, f+\Delta f)$ can be approximately linearized with respect to Δf and Δa as follows:

$$r(a+\Delta a,\mathrm{f}+\Delta f)\approx r(a,\mathrm{f})+\begin{bmatrix}\frac{\partial r(a,\mathrm{f})}{\partial a} & \frac{\partial r(a,\mathrm{f})}{\partial f}\end{bmatrix}\begin{bmatrix}\Delta a\\ \Delta f\end{bmatrix} \qquad (15)$$

The two derivative terms in the above equation are calculated in the two following relations:

$$\frac{\partial r(a,\mathrm{f})}{\partial a}=\frac{\partial(g-Wf)}{\partial a}=-\frac{\partial Wf}{\partial a}=-J(a,f) \qquad (16)$$

$$\frac{\partial r(a,\mathrm{f})}{\partial f}=\frac{\partial(g-Wf)}{\partial f}=-W \qquad (17)$$

Substituting (16) and (17) in (15) yields

$$r(a+\Delta a,\mathrm{f}+\Delta f)\approx r(a,\mathrm{f})-J(a,\mathrm{f})\Delta a-W\Delta f \qquad (18)$$

Where $J(a,f)$ is the Jacobian matrix, the calculation of which presents a challenging issue, as will be discussed in the next section (3.2.).

Substituting (18) in (14) and rewriting the content of the norm as a linear combination of incremental values of unknowns leads to

$$E(f+\Delta f,a+\Delta a)=\left\|\begin{matrix}-r(a,f)+J(a,f)\Delta a+W\Delta f\\ \sqrt{\lambda}Lf+\sqrt{\lambda}L\Delta f\\ \sqrt{\beta}(a-\bar{a})+\sqrt{\beta}(\Delta a)\end{matrix}\right\|^2 \qquad (19)$$

$$=\left\|\begin{pmatrix}J(a,f) & W\\ 0 & \sqrt{\lambda}L\\ \sqrt{\beta}I & 0\end{pmatrix}\begin{pmatrix}\Delta a\\ \Delta f\end{pmatrix}+\begin{pmatrix}-r(a,f)\\ \sqrt{\lambda}Lf\\ \sqrt{\beta}(a-\bar{a})\end{pmatrix}\right\|^2$$

Now, instead of minimizing the nonlinear cost function in (14) with respect to the main unknowns, i.e. f and a, the linear cost function in (19) is minimized with respect to their incremental values, i.e. Δf and Δa respectively [19],[21]. The incremental values are used to update the main unknowns. Obtaining the optimal solution for the cost function in (14) is not guaranteed, but starting from initial values close to optimal values and continuing the optimization of the linear cost function in (19) may yield the optimal solution [21].

3.2 The proposed Method for Derivation of Jacobian Matrix $J(a,f)$

Jacobian matrix in the existing joint methods is derived from the three principal operations separately [19], [21], [22]. In this paper, however, we propose the derivation of Jacobian matrix using the combinational operation (5), and in contrast to [13], it is calculated analytically rather than numerically. As mentioned in (16) it can be written as follows:

$$J(a,f)=\frac{\partial Wf}{\partial a}=\frac{\partial Wf}{\partial[h_{11},...,h_{41},h_{12},...,h_{4K}]} \qquad (20)$$

$J(a,f)$ is a $KM_gN_g\times 4K$ block diagonal matrix [21] of matrices $J(a,f)$ where

$$J(a_k,f)=\frac{\partial W_k(a_k)f}{\partial a_k}=\frac{\partial W_k(a_k)f}{\partial[h_{1k},...,h_{4k}]} \qquad (21)$$

Naturally $J(a_k,f)$ is a $M_gN_g\times 4$ matrix and the nth column of which can be obtained as follows:

$$J_n(a_k,f)=\frac{\partial W_k(a_k)f}{\partial h_{nk}}=\frac{\partial W_k(a_k)}{\partial h_{nk}}f \qquad (22)$$

where $1\le n\le 4$. Using equations (5) and (6), each entry of matrix $\partial W_k(a_k)/\partial h_{nk}$ is calculated as follows:

$$\frac{\partial W_{ijk}(a_k)}{\partial h_{nk}}=\frac{\sum_{j'=1}^{M_fN_f}(\nabla_n Q_{ijk}-\nabla_n Q_{ij'k})\tilde{W}_{ij'k}(a_k)}{\sum_{j'=1}^{M_fN_f}\tilde{W}_{ij'k}(a_k)}W_{ijk}(a_k) \qquad (23)$$

where the operator ∇_n is derivative with respect to h_{nk} and

$$Q_{ijk}=-\frac{1}{2\sigma^2|M_k|}(v_j-S_k(u_i))^{\mathrm{T}}(v_j-S_k(u_i)) \qquad (24)$$

therefore

$$\nabla_n Q_{ijk}=\frac{1}{\sigma^2|M_k|}\nabla_n S_k(u_i)^{\mathrm{T}}(v_j-S_k(u_i))-\frac{\nabla_n|M_k|^{-1}}{2\sigma^2}(v_j-S_k(u_i))^{\mathrm{T}}(v_j-S_k(u_i)) \qquad (25)$$

Where $\nabla_n|M_k|^{-1}$ and $\nabla_n S_k(u_i)$ can be calculated simply using Equations (3), (7) and (8).

The proposed method can be implemented in the framework of an algorithm as described in the next subsection:

3.3 Algorithmic Steps of the Proposed Joint Method

Step 1: Calculating initial values for registration parameters ($\hat{a}$) using ECC algorithm.

Step 2: Calculating the initial value of HR image using Delaunay triangulation-based interpolation method of SR, given the LR images and the estimated $\hat{a}$.

Step 3: (At the iteration i) Calculating the combinational coefficient matrix $W(a)$, Jacobian Matrix $J(a,f)$, residual vector $r(a,f)$ and BTV matrix T as discussed earlier.

Step 4: Solving the linear system of equations (26) using a CG algorithm:

$$\begin{bmatrix} J^{\mathrm{T}}J+\beta I & J^{\mathrm{T}}W \\ W^{\mathrm{T}}J & W^{\mathrm{T}}W+\lambda T \end{bmatrix}\begin{pmatrix} \Delta a \\ \Delta f \end{pmatrix}=\begin{bmatrix} J^{\mathrm{T}}r-\beta(a-\bar{a}) \\ W^{\mathrm{T}}r-\lambda Tf \end{bmatrix} \tag{26}$$

This linear system of equations is the result of minimizing the cost function expressed in (19), which is calculated by taking derivative of the cost function with respect to Δf and Δa and then equating the result to zero.

Step 5: Updating the unknown variables using the estimated incremental values:

$$\begin{pmatrix} a \\ f \end{pmatrix}^{i+1}=\begin{pmatrix} \Delta a \\ \Delta f \end{pmatrix}+\begin{pmatrix} a \\ f \end{pmatrix}^{i} \tag{27}$$

Where the superscript shows the iteration number.

Step 6: Updating $\bar{a}$ according to the following relation:

$$\bar{a}^{i+1}=\frac{i\bar{a}^{i}+a^{i+1}}{i+1} \tag{28}$$

Step 7: If the following condition of HR image, (29), is satisfied for a specified threshold (Thr), which is assumed 10-6 here, or a maximum number of iteration is reached, Then stop Else go to step 3.

$$\frac{\left\|f^{i}-f^{i-1}\right\|^{2}}{\left\|f^{i-1}\right\|^{2}}<\mathrm{Thr} \tag{29}$$

3.4 Cost Function of the New AM Method

In the AM methods, the cost function is minimized in two separate phases. In the first phase, it is minimized with respect to f. Fortunately, the cost function (14) is quadratic with respect to f and its MAP estimation is obtained as follows [15]:

$$f(a)=\left(W^{\mathrm{T}}W+\lambda T\right)^{-1}W^{\mathrm{T}}g \tag{30}$$

In the second phase, the cost function, which is not quadratic with respect to a, is minimized as a function of a. For this reason, the cost function is linearized only with respect to incremental values of registration parameters Δa using (19)

$$E(f,a+\Delta a)=\left\|\begin{pmatrix} J(a,f) \\ 0 \\ \sqrt{\beta}I \end{pmatrix}\Delta a+\begin{pmatrix} -r(a,f) \\ \sqrt{\lambda}Lf \\ \sqrt{\beta}(a-\bar{a}) \end{pmatrix}\right\|^{2} \tag{31}$$

Where $J(a,f)$ is obtained as discussed in Section 3.2, using proposed method. Finally, the vector Δa is calculated by solving the following linear equation:

$$\left(J^{\mathrm{T}}J+\beta I\right)\Delta a=J^{\mathrm{T}}r-\beta(a-\bar{a}) \tag{32}$$

Other details, the same as the proposed iterative joint method in section 3.3, have not been included here.

This proposed AM method, like other AM methods, is more sensitive to the initial values compared to the joint method.

3.5 Algorithmic Steps of the Proposed AM Method

Step 1: Calculating initial values for registration parameters ($\hat{a}$) using ECC algorithm and then calculating the combinational coefficient matrix $W(a)$.

Step 2: Calculating the initial value of HR image using Delaunay triangulation-based interpolation method of SR based on the LR images and the estimated $\hat{a}$.

Step 3: (At the iteration i) Calculating the Jacobian Matrix $J(a,f)$ and the residual vector $r(a,f)$.

Step 4: Solving the linear system of equations (32) using a CG algorithm.

Step 5: Updating the registration parameters using the estimated incremental values:

$$(a)^{i+1}=(\Delta a)+(a)^{i} \tag{33}$$

Step 6: Updating $\bar{a}$ using Equation (27).

Step 7: Calculating the combinational coefficient matrix $W(a)$ and BTV matrix T.

Step 8: Solving the linear system of equations (30) to obtain a new update of the HR image f.

Step 9: If the condition (29) is satisfied or a maximum number of iteration is reached, then stop Else go to step 3.

3.6 Implementation Details

To realize the proposed methods and compare them with other methods, MATLAB R2012b is used. The ECC library provided by Evangelidis [29] is used to obtain initial values for registration parameters. The preconditioned conjugate gradient (PCG) routine "pcg" is used where a linear system of equations should be solved. We partially used (just subroutine "makeW.c") the codes provided by Pickup [30], making some modifications to produce the desired combinational coefficient matrix. The Gaussian kernel used in this subroutine is truncated after three standard deviations in terms of LR pixels. This has a significant impact on reducing the computational cost of Equations (5) and (23), as will be discussed in the next subsection 3.7. The above-mentioned truncation is also considered in the blur mask produced by the routine "fspecial", which is used to perform the blurring operation in Tian's and Hardie's methods. Other parameters like λ and β are taken as [22].

The original Hardie's method proposed for translational motion model and direct search in neighboring pixels was used in the optimization of registration parameters. To draw a fair comparison, we extended this method to include similarity motion model, and the nonlinear least square optimization method used in the proposed AM method was also applied to the Hardie's method because a motion model more complex than translational was not applicable to original Hardie's method.

In the methods that perform the warping and blurring operations separately [21], [22], after warping the HR image, some blank areas like corners usually appear. These blank regions may be filled by black pixels or nearest pixels [21]. When the motion is considerable, selecting each of these tricks may negatively affect the final results because the blur mask can significantly change the gray level of the pixels located in the borders of these blank regions. To overcome this problem, we used another trick. When using the reverse mapping model, the pixels that leave the frame are marked during the warping operation. This process does not require considerable additional computation since it is a part of warping operation. Then, the columns of blurring matrix H_k associated to these pixels are set to zero and the rows of the matrix are renormalized. These operations are all conducted through matrix calculations. This trick has been used in our implementations in the next section to increase the performance of Tian's and Hardie's methods.

3.7 Computational Complexity

In this subsection, the computational complexity of the proposed methods is calculated and compared with two other simultaneous methods (Tian's [22] and Hardie's [15] methods). The overall computational cost of the proposed methods (as well as Tian's and Hardie's methods) is affected by several factors including the dimensions of LR images, the number LR images, the increasing factor, the size of blur mask and the termination criteria of algorithm (the number of iterations allowed in the algorithm and the pcg routine). Similar to [31] and [22] it is assumed that the addition and multiplication operations are the same. In all methods, the sparsity of matrices is considered in the computational complexity. There are three processes that their complexity dominates others: the calculation of combinational coefficient matrix $W(a)$, the calculation of Jacobian Matrix $J(a,f)$ and solving the linear system of equations (26) or (30). Since the frameworks of the methods are the same, the complexity is calculated only for one iteratio n. In Tian's and Hardie's methods $W(a)$ is calculated in accordance with (1) [22] by separately calculating its component matrices, i.e. D, H_k and b(a_k). Because the motion model is similarity and the bilinear interpolation is used, the computational cost of $B = blkdiag\{B(a_1), \dots, B(a_K)\}$ is $O(8kM_fN_f)$. The blur matrix $H = blkdiag\{H_1, \dots, H_K\}$ has M_hN_h non-zero elements in each row and the multiplication of HB can be considered as the convolution of M_hN_h blur mask (since the blur mask is truncated) with a 2×2 interpolation mask. Hence, the computational cost of this multiplication is O(4($M_h + 1)(N_h + 1)KM_fN_f)$. Finally since $O(8KM_fN_f) < O(4(M_h + 1)(N_h + 1)KM_fN_f)$ the computational cost of $W(a)$ is $O\big(4KM_hN_hM_fN_f\big) = O(K\rho^2M_hN_hM_gN_g)$. In these two methods $J(a,f) = blkdiag\{J(a_1,f), \dots, J(a_{K-1},f)\}$ is derived by separate calculation of its component matrices, i.e. $J(a_k,f) = \mathrm{DH}_k(\partial B(a_k)/\partial a_k) = \mathrm{DH}_k\mathrm{E}_kC$ where C is a $2M_fN_f \times 4$ matrix that requires no considerable calculation and E_k is a $M_fN_f \times 2M_fN_f$ matrix that consists of two adjacent diagonal matrices (see [21] or [22] for further details). The computational cost of E_k is $O(4M_fN_f)$ and its multiplication with C yields the computational cost of $O(8M_fN_f)$. Since C is not sparse, the new matrix E_KC will not be sparse too, hence the computational cost of H_KE_KC will be $O(4M_hN_hM_fN_f)$. Finally, the computational cost of $J(a,f)$ will be $O(K\rho^2M_hN_hM_gN_g)$. For Hardie's AM method, the HR image is obtained in each iteration by solving the linear system of equation (30) using the iterative PCG algorithm. The computational cost of the PCG for one iteration is equal to the number non-zero elements of the coefficient matrix of the linear system of equations [32]. Although the dimensions and the number of non-zero elements of the sparse matrix $W(a)$ are dependent on the number of LR images and the increasing factor, this is not the case for the coefficient matrix of the linear system of equations (30), i.e. $W^TW + \lambda T$. The number of non-zero elements of this sparse matrix is approximately equal to $(M_f(2M_h + 1) - M_h^2 - M_h)(N_f(2N_h + 1) - N_h^2 - N_h)$. Hence, if the iteration number allowed for PCG is L, the computational cost of PCG algorithm will be $O(L\rho^2M_hN_hM_gN_g)$. For the Tian's joint method, the larger linear system of equation (26) should be solved. As mentioned earlier, since $J(a_k,f)$ is not sparse, W^TJ will not be sparse too. Hence, the computational cost of PCG algorithm for solving linear system of equations (26) is $O((M_hN_h + 2K)L\rho^2M_gN_g)$. In the proposed methods, the combinational coefficient matrix $W(a)$ is calculated in accordance with relations (5)-(8). Each row of this matrix, has $(M_h + 1) \times (N_h + 1)$ elements (as before) and requires $3(M_h + 1)(N_h + 1) + 7$ multiplications and $(M_h + 1)(N_h + 1)$ exponentiations. Since, at most three standard deviations of a Gaussian is preserved, its argument will have a maximum absolute value of 3^2/2=4.5. Experimentally, it was verified that such an exponent has a computational cost which is at most 15 times greater than the computational cost of a multiplication. Hence, the computational cost of $W(a)$ is $O\Big(\big(3(M_h + 1)(N_h + 1) + 7 + 15(M_h + 1)(N_h + 1)\big)KM_gN_g\Big) = O(KM_hN_hM_gN_g)$. To calculate the computational complexity of the combinational Jacobian matrix $J(a,f)$ in the proposed methods, we found that the computational cost of $J(a_k,f)$ was 4 times greater than $J_n(a_k,f)$ because the motion model is similarity. According to (22) $J_n(a_k,f)$ is the multiplication of matrix $\partial W_k(a_k)/\partial h_{nk}$ with vector f. Each row of $\partial W_k(a_k)/\partial h_{nk}$ has the same number of non-zero elements as $W_k(a_k)$, i.e. $(M_h + 1)(N_h + 1)$, hence this matrix-vector product will require $(M_h + 1)(N_h + 1)M_gN_g$ multiplications. According to (23)-(25), each element of $\partial W_k(a_k)/\partial h_{nk}$ requires $(M_h + 1)(N_h + 1) + 8$ multiplications. Finally, the computational cost of $J(a,f)$ is $O\big(4(K-1)(M_h + 1)(N_h + 1)M_gN_g + 4(K-1)(M_h + 1)(N_h + 1) + \big((M_h + 1)(N_h + 1) + 8\big)M_gN_g\big) = O(KM_h^2N_h^2M_gN_{g)}$.

Now, the summery of computational complexity of the four methods are as follows:

A. Extended Hardie's AM method:

$Max\ (O(K\rho^2 M_h N_h M_g N_g)\ ,\ O(L\rho^2 M_h N_h M_g N_g)) = O(Max(K,L)\rho^2 M_h N_h M_g N_g)$

B. The proposed AM method:

$Max\ (O(KM_h^2 N_h^2 M_g N_g)\ ,\ O(L\rho^2 M_h N_h M_g N_g)) = O(Max(KM_h N_h, L\rho^2) M_h N_h M_g N_g)$

C. Tian's joint method:

$Max\ (O(K\rho^2 M_h N_h M_g N_g)\ , O((M_h N_h + 2K)L\rho^2 M_g N_g)) = O(Max(KM_h N_h, LM_h N_h + 2KL)\rho^2 M_g N_g)$

D. The proposed joint method:

Max $(O(KM_h^2 N_h^2 M_g N_g)\ , O((M_h N_h + 2K)L\rho^2 M_g N_g)) = O(Max(KM_h^2 N_h^2, (M_h N_h + 2K)L\rho^2) M_g N_g).$

Generally, it can be concluded that joint methods have higher computational cost than AM methods. Practically, when a higher increasing factor is selected, the larger blur mask is desirable so that $M_h N_h$ is approximately proportional to ρ^2 and then $O(KM_h^2 N_h^2 M_g N_g) \approx O(K\rho^2 M_h N_h M_g N_g)$. Hence, the proposed AM and joint methods will have the same computational cost as Hardie's AM and Tian's joint methods respectively. It should be noted that the computational complexity derived here is more detailed than the one used in [22].

4. Experimental Results

In this section, the performance of the proposed Joint and AM methods are discussed and compared with a recently joint method (Tian's method [22]) and the famous AM method (Hardie's method [15]). We have provided five experiments including three synthetic image sequences and two real-life images sequences. The synthetic sequences, produced by warping some test images, are used to evaluate the performance of methods according to the following metrics: normalized mean square error (NMSE) for the estimated registration parameters vector and the peak signal to noise ratio (PSNR) for the reconstructed HR image, which are defined as follows [21],[22]:

$$\mathrm{NMSE}(\tilde{a}) = 100\frac{\|a-\tilde{a}\|^2}{\|a\|^2} \tag{34}$$

$$\mathrm{PSNR}(\tilde{f}) = 10\log_{10}\left(\frac{M_f N_f}{\|f-\tilde{f}\|^2}\right) \tag{35}$$

where "~" denotes the currently estimated values of the unknowns. Lower values of NMSE and higher values of PSNR are preferred in a method. Apart from the two objective measures, the final HR images are used to compare the performance of the methods subjectively.

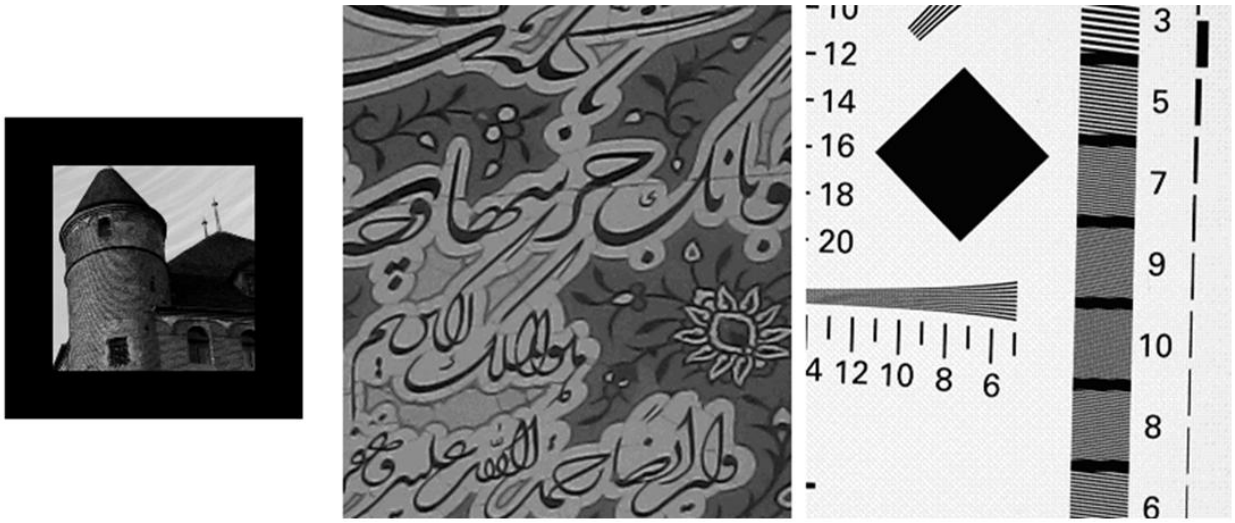

Fig. 2. Three test images, 'Castle', 'Khayam' and 'Chart' which are considered as the ground truth in the reconstruction.

Fig. 3. Four LR images of the sequence were created from the first test image.

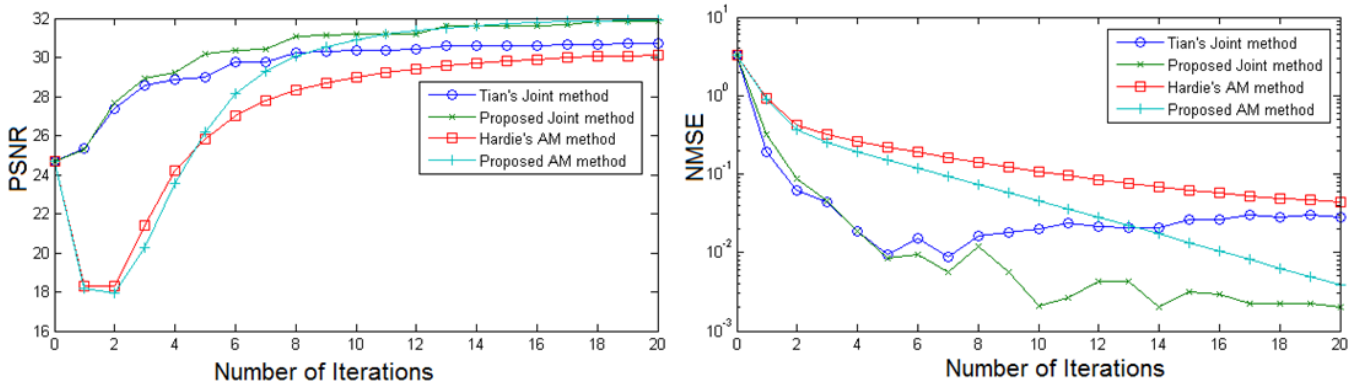

Fig. 4. PSNR of the reconstructed HR image and NMSE of the estimated registration parameters using the four methods for the 'Castle' image.

4.1 Experimental Results of Degraded Test Images

Three test images are used for the experiments in this subsection. The first one (188x186 pixels) is 'Caslte' image [33], which has been placed at the center of a frame surrounded by a thick strip of 30 black pixels (zero gray value). By doing so, the main part of the image is not even pushed out of the frame partially after warping. The second and third test images, namely 'Khayam' and 'Chart' images [33], are relatively large (580x640 and 512x640 pixels respectively). After warping these test images, the central part of images (320x280 pixels), which are still in the frame, are selected as warped HR images. Therefore, the original HR images (the ground truth) be reconstructed are shown in Fig. 2. In the first experiment, a sequence of six images were created by warping one of test images through the use of different homogenous matrices (M_k) each of which containing a zooming factor (z_k), a rotation angle (θ_k) and a translation in both vertical and horizontal directions (v_{xk}, v_{yk}), randomly chosen from the ranges $[0.95,1.05]$, $[-5°, 5°]$ and $[-3,3]$ respectively. Then, they were blurred by an isotropic Gaussian kernel with a variance of 0.333 LR pixel. These images, degraded by Additive White Gaussian Noise (AWGN) to have 30 dB SNR, were down-sampled by a decimation factor of 2. Four degraded LR images for the first test image are shown in Fig.3. Using four methods, i.e. Tian's method, Hardie's method and our two proposed methods, the HR image was reconstructed. To obtain initial values for registration parameters, no explicit IR techniques were used in the experiments of this subsection; instead, the inverse of each original homogeneous matrix M_k was multiplied by a homogeneous error matrix. This homogenous error matrix, similar to M_k, contains a zooming factor, a rotation angle and a translation in both vertical and horizontal direction randomly chosen from the ranges $[0.995,1.005]$, $[-0.5°, 0.5°]$ and $[-0.5,0.5]$ respectively. The PSNR of the reconstructed HR images and NMSE of estimated registration parameters using these four methods for the 'Castle' image are shown in Fig.4. Also, the HR images reconstructed by these methods for the test image are shown in Fig.5. As can be seen, the proposed methods have improved NMSE and PSNR more than Tian's and Hardie's methods. Additionally, the superior performance of the proposed methods is evident in the reconstructed HR images highlighted in Fig.5 (e-h). As can be seen, the PSNR is reduced in the AM methods at the first iteration, while the reconstructed HR image deviates from the desired solution, though it returns to the desired solution in the later iterations. We have also examined other initial

values of HR image such as up-scaled version of reference image and MAP estimation of the HR image using the initial values of registration parameters [21]. However, it had no significant impact on the performance of joint methods, but it could affect the performance of the AM methods. This experiment was repeated with a homogeneous error matrix that contained twice the values of zooming factor, rotation angle and translation in both vertical and horizontal directions. Nevertheless, it was observed that the joint methods converged during 20 iterations, with most of the attempts to perform the AM methods leading to divergence. In the second experiment, in which the second and third test images were used, we focused on the zooming part of the motion model. A sequence of six images with zooming factors $z_k = 1 + k/6$ for $k = 0,2,3,4,5,6$ and vertical translations $v_{xk} = -26, -39, -53, -66, -78$ and horizontal translations $v_{yk} = -23, -35, -46, -57, -67$ were created. Vertical and horizontal translations were selected in such a way that the center of the test image remained unchanged after warping. Other steps were the same as the first experiment. Four degraded LR images for the 'Chart' image are shown in Fig.6. The PSNR of the reconstructed HR images and the NMSE of the estimated registration parameters for the test image are shown in Fig.7 and HR images reconstructed by the four methods are illustrated in Fig.8. The results show that our proposed methods can reconstruct the edge regions with greater precision, especially when the relative zooming is considerable.

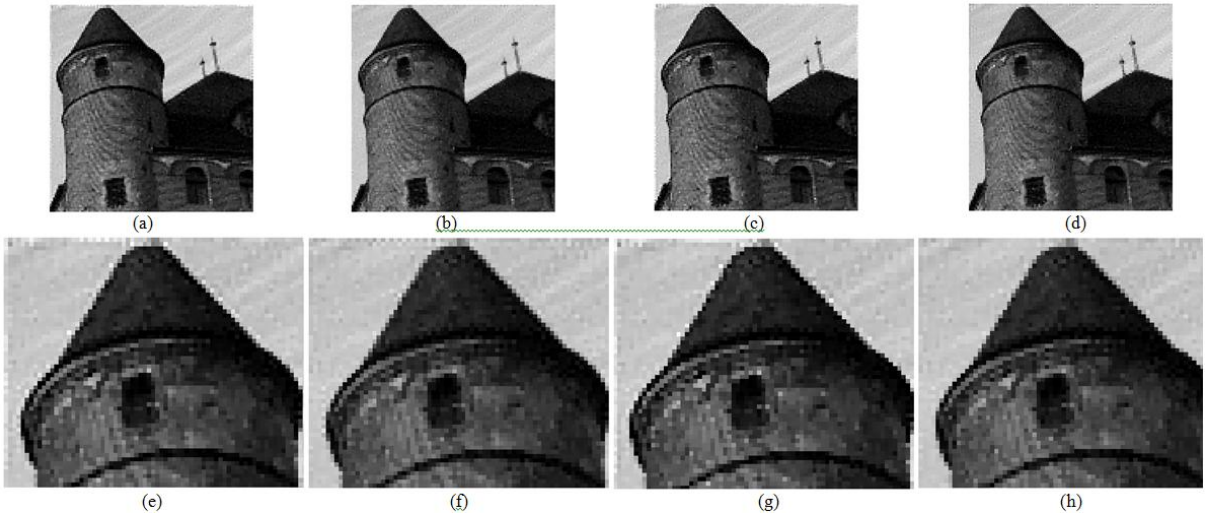

Fig. 5. the reconstructed HR images using Tian's Joint method (a), proposed Joint method (b), Hardie's AM method (c) and proposed AM method (d). A zoomed part of these images have also been shown in (e-h) respectively.

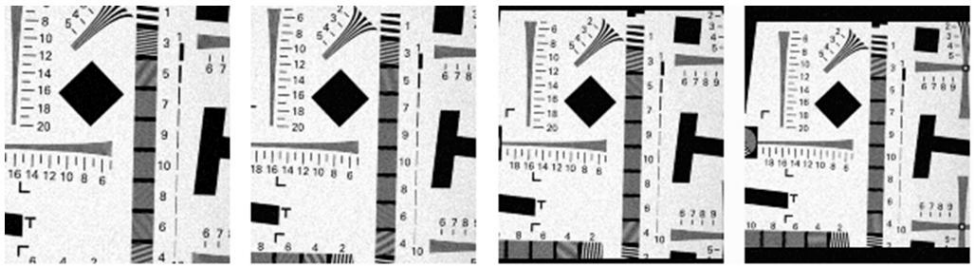

Fig. 6. A sequence of four degraded LR images was created from the third test image by emphasizing the zooming.

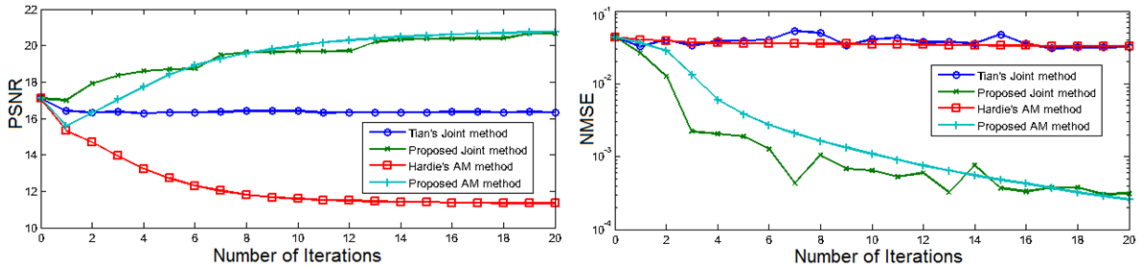

Fig. 7. PSNR of the reconstructed HR images and the NMSE of the estimated registration parameters using all four methods for the 'Chart' image.

4.2 Experimental Results of Real-life Images

In this subsection, two sequences were extracted from two videos of 'text' and 'car', which are accessible in [34]. Four frames of each video are shown in Fig. 9. In the first experiment, a sequence of 10 frames with a size of 57x49 pixels was extracted from the 'text' video. Also, σ was set at 0.55 LR pixel and the increasing factor was set at 4. Initial values for registration parameters were obtained by setting the homogeneous matrix M_k equal to the identity matrix for all frames, because the relative motion between successive frames was small in this experiment. This can be considered as a special application of these SR methods. The HR images reconstructed by the four methods are displayed in Fig.10. The results of this experiment also confirm the superior performance of our proposed methods in the edges. Also, it is observed that letters in reconstructed images are clearly distinguishable in the proposed method. Selecting a larger standard deviation increases the shadows around the letters, especially in Tian's and Hardie's methods. In the second experiment, a sequence of 16 frames with a size of 121x72 pixels was extracted from 'car' video. σ was set at 0.45 LR pixel and the increasing factor was set at 3. Initial values for registration parameters were obtained by adopting the ECC IR technique [24]. The HR images reconstructed by the four methods are shown in Fig. 11.

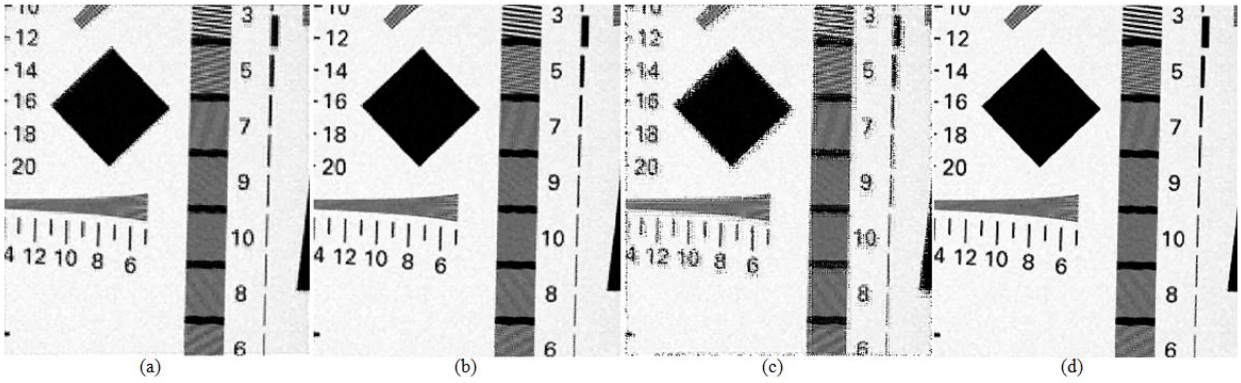

Fig. 8. Reconstructed HR images using Tian's joint method (a), our proposed joint method (b), Hardie's AM method (c) and our proposed AM method (d).

Fig. 9. Four LR images of the sequences were extracted from videos 'text' and 'car'

Our proposed joint method produced better results in this experiment. Although both license plates, namely '3PLK273' and the name of vehicle manufacturer 'SUBARU', are recognizable, the letters of license plate are more distinctive in the proposed method. Our proposed AM method detected the license plate letters precisely, though some shadows in edge regions could be observed in other parts of image.

We used a DELL/Vostro notebook with 4 GB RAM and a 2.5 GHz dual core processor in our experiments. The maximum iteration number for both pcg routine and the four discussed simultaneous methods was set at 20. At each iteration, however, the proposed AM and joint methods had the same computational complexity as Hardie's and Tian's methods respectively, the run time of the proposed joint and AM methods were less than Tian's and Hardie's methods respectively. For example, the run times of the first experiment for 6 LR images with a size of 94x93 pixels were 64s, 40.5s, 52.5s and 24s for Tian's method, the proposed joint, Hardie's method and the proposed AM methods respectively. This is due to three reasons. First of all, the proposed methods require fewer iterations for the convergence the two others, as shown in Fig.4. Also, the three principal operations are merged into one operation and hence fewer loops are employed in implementations. Although the use of loops instead of matrix calculation has no significant impact on complexity, it affects the run time. Finally, additional

operations are required in Tian's and Hardie's methods to avoid the blank regions discussed in subsection 3.6.

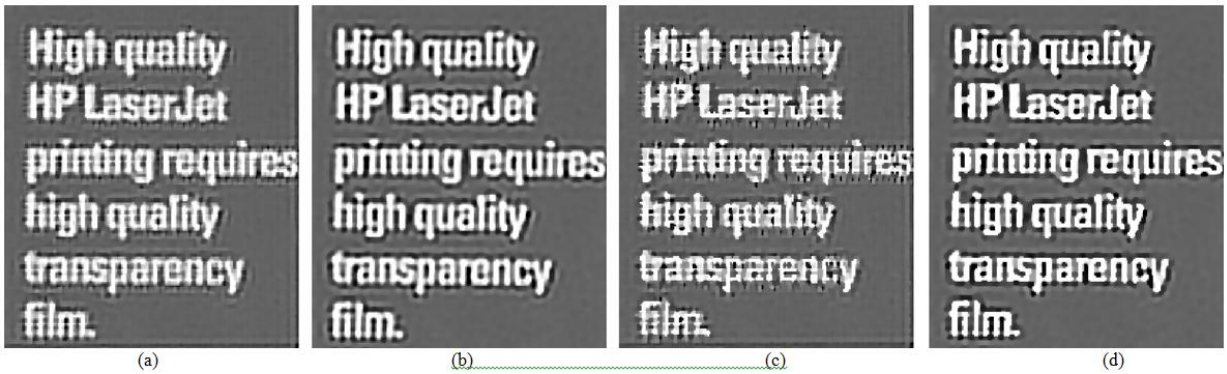

Fig. 10. Reconstructed HR images using Tian's Joint method (a), proposed Joint method (b), Hardie's AM method (c) and proposed AM method (d).

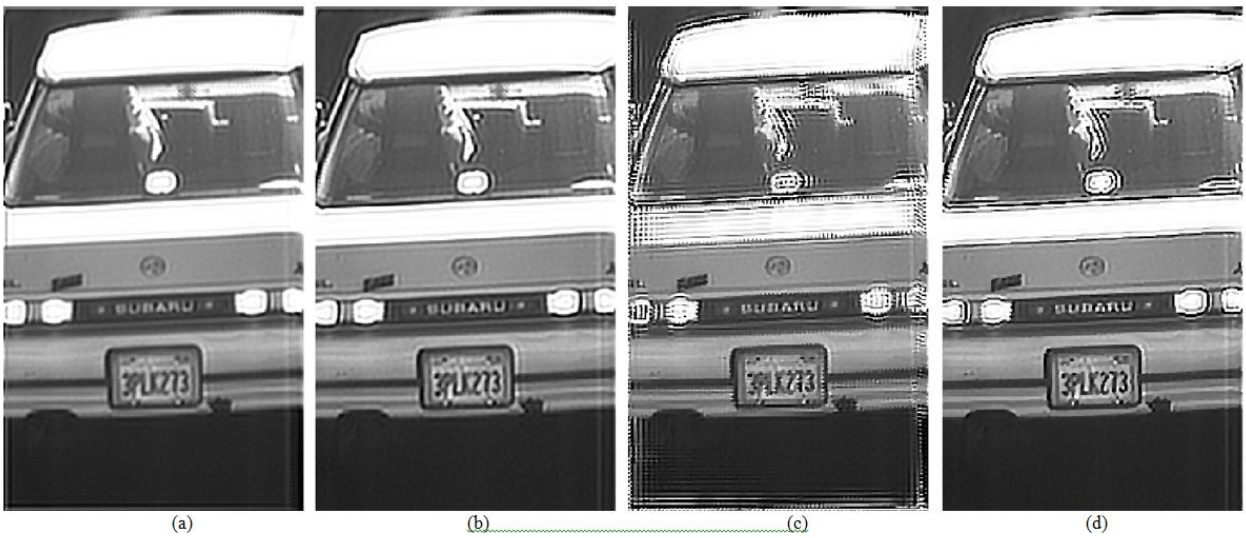

Fig. 11. Reconstructed HR images using Tian's Joint method (a), proposed Joint method (b), Hardie's AM method (c) and proposed AM method (d).

5. Conclusion and Future Works

In this paper we proposed a new joint method that combined the three principal operations in one operation. The application of this combinational operation in the calculation of Jacobian matrix is one of the most important contributions of this paper. The proposed joint method reduced error propagation, was less likely to be trapped into suboptimal solutions, especially when the relative zooming between frames was considerable, and finally increased the quality of the reconstructed HR images. Applying this combinational operation to the framework of AM methods presented a new AM method. The proposed AM method was not as stable as the proposed joint method, but it was more reliable than existing AM methods such as Hardie's method. Similar to other AM methods, the convergence of the proposed AM method was highly dependent on initial values of HR image and the registration parameters but if the initial values were close to optimal values, this method provided fast convergence. In future works, we will extend the motion model to the affine and finally to the homography, and the size of blur kernel in the image plane will be refined during the iterations.

References

[1] S. Park, M. Park and M. Kang. "Super-Resolution Image Reconstruction: A Technical Overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp.21– 36, 2003.

[2] S. Borman and R.L. Stevenson. "Spatial resolution enhancement of low-resolution image sequences. A comprehensive review with directions for future research," Laboratory for Image and Signal Analysis (LISA), University of Notre Dame, Notre Dame, Ind, USA, Tech. Rep., 1998.

[3] K. Katsaggelos, R. Molina and J. Mateos. *Super Resolution of Images and Video*, Morgan and Claypool Publishers, San Rafael, 2007.

[4] P. Milanfar. *Super-Resolution Imaging*, CRC Press, 2011.

[5] S. Lertrattanapanich and N. Bose. "High Resolution Image Formation from Low Resolution Frames Using Delaunay Triangulation," *IEEE Transaction on Image Processing*, vol. 11, no. 12, pp. 1427–1441, 2002.

[6] F. Xu, H. Wang, L. Xu and C. Huang. "A new framework of normalized convolution for superresolution using robust certainty," *IEEE International Conference on Computer and Automation Engineering*, vol. 2, pp. 144-148, 2010.

[7] T.Q. Pham, L.J.V. Vliet and K. Schutte. "Robust Fusion of Irregularly Sampled Data Using Adaptive Normalized Convolution," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1-12, 2006.

[8] K. Zhang, G. Mu, Y. Yuan, X. Gao and D. Tao, "Video super-resolution with 3D adaptive normalized convolution", *Elsevier Journal of Neurocomputing*, vol. 94, pp. 140-151, 2012.

[9] A. Sánchez-Beato and G. Pajares, "Noniterative Interpolation-Based Super-Resolution Minimizing Aliasing in the Reconstructed Image", *IEEE Transaction on Image Processing*, VOL. 17, NO. 10, pp. 1817-1826, 2008.

[10] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution," *IEEE International Conference on Computer Vision and Pattern Recognition* , pp. 645–650, 2001.

[11] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[12] M.E. Tipping and C.M. Bishop, "Bayesian image super-resolution," *Conference on Advances in Neural Information Processing Systems 15 (NIPS)*. Cambridge: MIT Press, 2003.

[13] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman. "Bayesian Methods for Image Super-Resolution," *Computer Journal*, 2007

[14] S. D. Babacan, R. Molina and A. K. Katsaggelos. "Variational Bayesian Super Resolution," *IEEE Transaction on Image Processing*, vol. 20, no. 4, 2011.

[15] R. C. Hardie, K.J. Barnard and E.E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images, " *IEEE Transaction on Image Processing*, vol. 6, no. 12, pp. 1621-1633, 1997.

[16] N. A. Woods, N. P. Galatsanos and A. K. Katsaggelos, "Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images," *IEEE Transaction on Image Processing*, vol. 15, no. 1, 201–213. 2006.

[17] B. C. Tom, A. K. Katsaggelos and N. P. Galatsanos, "Reconstruction of a high resolution image from registration and restoration of low resolution images," *IEEE International Conference on Image Processing*, vol. 3, pp. 553–557, 1994.

[18] G. Golub and V. Pereyra, "Separable nonlinear least squares: The variable projection method and its applications," *Journal of Inverse. Problems*, vol. 19, no. 2, pp. 1–26, 2003.

[19] J. Chung, E. Haber, and J. Nagy, "Numerical methods for coupled super-resolution," *Journal of Inverse Problems*, vol. 22, no. 4, pp. 1261–1272, 2006.

[20] D. Robinson, S. Farsiu, and P. Milanfar, "Optimal registration of aliased images using variable projection with applications to super-resolution," *Computer. Journal*., vol. 52, no. 1, pp. 31–42, 2007.

[21] Y. He, K.-H. Yap, L. Chen and L.-P. Chau. "A Nonlinear Least Square Technique for Simultaneous Image Registration and Super-Resolution," *IEEE Transaction on Image Processing*, vol. 16, no. 11, pp. 2830–2841, 2007.

[22] Y. Tian and K.H. Yap. "Joint Image Registration and Super-Resolution from Low-Resolution Images with Zooming Motion," *IEEE Transaction on Circuits and System for Video Technology*, vol. 27, no. 3, pp. 1224-1234, 2013.

[23] Capel, D. Image Mosaicing and Super-resolution (Distinguished Dissertations). Springer, ISBN: 1852337710, 2004.

[24] G.D. Evangelidis, E.Z. Psarakis, "Parametric Image Alignment using Enhanced Correlation Coefficient", *IEEE Transaction on Pattern Analysis and Machine Inteligence*, vol. 30, no. 10, 2008.

[25] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, 2nd ed., Ed. Cambridge, United Kingdom: Cambridge University Press, 2003.

[26] T. Lehmann, C. Gonner, and K. Spetzer, "Survey: Interpolation Methods in Medical Image Processing, " *IEEE Transaction on Medical Imaging*, vol. 18, no. 11, pp. 1049-1067, 1999.

[27] F. Sroubek and J. Flusser, "Multichannel blind iterative image restoration," *IEEE Transaction on Image Processing*, vol. 12, no. 9, pp. 1094–1106, 2003.

[28] T. F. Chan and C. K. Wong, "Total variation blind deconvolution," *IEEE Transaction on Image Processing*, vol. 7, no. 3, pp. 370–375, 1998.

[29] [On-line], Available: http://xanthippi.ceid.upatras.gr/people /evangelidis/ecc/ [2014-09-29].

[30] [On-line], Available: http://www.robots.ox.ac.uk/~elle/SRcode/ [2014-09-29].

[31] M. V. W. Zibetti and J. Mayer, "A robust and computationally efficient simultaneous super-resolution scheme for image sequences," *IEEE Trans. Circuits Syst. Video Technol*., vol. 17, no. 10, pp. 1288-1300, 2007.

[32] L. Olikery, X. Liy, P. Husbandsy and R. Biswasz, "Effects of Ordering Strategies and Programming Paradigms on Sparse Matrix Computations." *Journal of SIAM Review*, Vol. 44, No. 3, pp. 373-393, 2002.

[33] [On-line], Available: http://lcav.epfl.ch/software/superresolution [2014-09-29].

[34] [On-line], Available: http://users.soe.ucsc.edu/~milanfar/ software /sr-datasets.html [2014-09-29].

Hossein Rezayi received his B.Sc. degree in Electrical Engineering in 2001 and his M.Sc. degree in Communication Engineering in 2004 from the Ferdowsi University of Mashhad, Iran, where he is currently a Ph.D. candidate in Communication Engineering. His current research interests include machine vision, image processing, pattern recognition and artificial intelligence algorithms.

Seyed Alireza Seyedin received his B.Sc. degree in Electronics Engineering from Isfahan University of Technology, Isfahan, Iran in 1986, his M.Sc. degree in Control and Guidance Engineering from Roorkee University, Roorkee, India in1992, and his Ph.D. degree from the University of New South Wales, Sydney, NSW, Australia in, and 1996.In 1996, he joined the Faculty of Engineering of Ferdowsi University of Mashhad as an assistant professor, where he has been teaching as an associate professor since 2007. He was the Head of the Department of the Electrical Engineering from 1998 to 2000. His current research interests include machine vision, DSP-based signal processing algorithms, and digital image processing techniques. Dr. Seyedin is a member of the Technical Committee of the Iranian Conference on Machine Vision and Image Processing.