Video Super-Resolution through Image Registration based on Structural Similarity

M. Amintoosi, M. Fathy and N. Mozayani

Computer Engineering Department,
Iran University of Science and Technology
Narmak, Tehran, Iran
{*mAmintoosi,mahFathy,Mozayani*}*@iust.ac.ir*

Abstract

It is commonly known that the Mean Square Error (MSE) does not accurately reflect the subjective image quality for most video enhancement tasks. Among the various image quality metrics, Structural Similarity metric provides remarkably good prediction of the subjective scores. In this paper a new registration method based on contribution of structural similarity measurement to the well known Lucas-Kanade (LK) algorithm has been proposed. The core of the proposed method is contributing the SSIM in the sum of squared difference of images along with the Levenberg-Marquardt optimization approach in LK algorithm. Mathematical derivation of the proposed method, based on the unified framework of Baker *et. al.*is given. The proposed registration algorithm is applied to a video super resolution problem, successfully. Various objective and subjective comparisons show the superior performance of the proposed method.

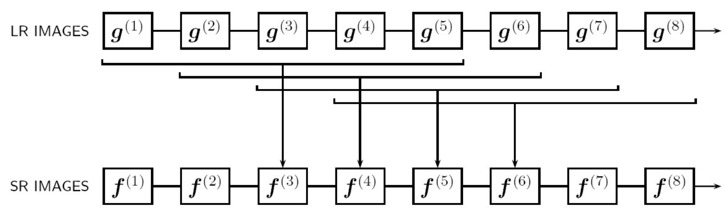


Figure 1: "sliding window" technique for video super resolution [3].

Keywords: Video Super-Resolution, Registration, Structural Similarity, Synthesis, Levenberg-Marquardt.

---

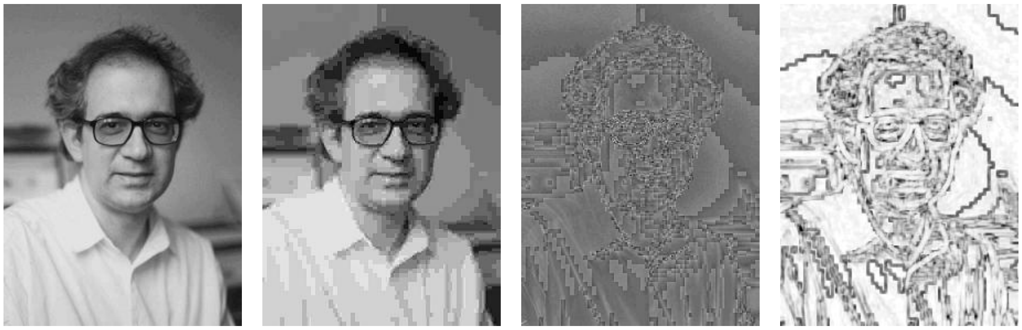
1. Introduction

Nowadays digital cameras are very popular and taking films and movies became usual tasks. Many of these devices – such as some mobile phones – can take High-Resolution (HR) photos and low-resolution (LR) videos. Enhancement of these LR videos using HR photos is related to Super-Resolution (SR) context. Video Super-resolution algorithms reconstruct a high resolution video from a low resolution video. The vast majority of the super-resolution restoration algorithms – named as reconstruction methods – use a short sequence of low-resolution input frames to produce a single super-resolved high-resolution output frame [1, 2]. These techniques have been applied to video restoration by using a shifting window of processed frames as illustrated in figure 1. For a given super-resolution frame, a "sliding window" determines the subset of LR frames to be processed to produce a given super-resolution output frame. The window is moved forward in time to produce successive super-resolved frames in the output sequence [3].

Some of the video resolution enhancement methods, map the whole of a training image onto each frame coordinates and fuse the result with the LR video frame [4, 5]. These methods require advanced motion-compensated signal processing. More precise mapping leads to a better synthesized result; hence any fruitful consideration of the mapping problem promises significant returns. In [6] a feature based registration approach using SIFT[1] key points [7] has been used. This approach followed by Lucas-Kanade registration method (LK-Algorithm)[8] in [5]. The well known LK-algorithm is a famous area based registration method and many variations of it has been introduced by researchers for several years [9]. The core part of this algorithm is finding the registration parameters with minimization of the square error between the reference image and a motion compensated of the other image.

Perhaps the mean square error is the most common objective criterion for measuring the differences in the image and video domains for several years. According to [10] automatic optimization based on a reliable subjective metric, seems a challenging target for future video enhancement research. Recently Amintoosi *et. al.*[11] proposed a new version of LK-algorithm which has higher performance relative to the its original form, in paticular when the LR image is very noisy. They used the Structural SIMilarity (SSIM) error measurement [12] as a weighting term to the objective function of LK algorithm. The chief idea of this approach is based on the fact that the contrast

[1]Scale Invariant Feature Transform



(a) Original Image (b) JPEG Compressed (c) MSE map (d) SSIM map

Figure 2: Comparing the error map of two images based on MSE and SSIM. The images are takes from [13].

inverted form of SSIM highlights the structural differences of two images, much better than the absolute error map, inparticular when one image is distorted. Figure 2 shows a reference image, its JPEG compressed version; the MSE map and the SSIM map between the original and its distorted version. As can be seen the structural differences are more clear in the SSIM image map.

In this paper another version of the image registration algorithm introduced in [11] has been proposed and applied to video super-resolution. Experimental results show the better performance of the new version of the LK-algorithm with respect to some others for the image registration purpose. Also the algorithm is applied in video super resolution problem and produced efficient results.

The reminder of this paper is organized as follows: in section 2 the proposed method and in section 3 experimental results are provided. The last section describes concluding remarks.

2. The Proposed Method

This section is categorized into three parts. Since the proposed method is based on the LK-algorithm and SSIM criterion, at first we will have a quick review to these concepts. In the second part we will discuss the mathematical derivation of the LK-algorithm based on SSIM and Levenberg-Marquardt optimization method. Then the application of this method for video resolution enhancement is explained.

2.1. Related Concepts

The goal of the Lucas-Kanade algorithm is to align a template image T to an input image I, by minimizing the following Sum of Squared Differences (SSD) between two images:

$$SSD = \sum_{x} [I(\mathbf{W}(\mathbf{x}; \mathbf{p})) - T(\mathbf{x})]^2 \tag{1}$$

where $\mathbf{x} = (x, y)^T$ is a column vector containing the pixel coordinates, $\mathbf{p} = (p_1, \ldots, p_n)^T$ is a vector of parameters; $\mathbf{W}(\mathbf{x}; \mathbf{p})$ denotes the parameterized set of allowed warps and $I(\mathbf{W}(\mathbf{x}; \mathbf{p}))$ is image I warped back onto the coordinates frame of the template T. The warp $\mathbf{W}(\mathbf{x}; \mathbf{p})$ takes the pixel $\mathbf{x}$ in the coordinate frame of the image I and maps it to the sub-pixel location $\mathbf{W}(\mathbf{x}; \mathbf{p})$ in the coordinates frame of the template T [9]. The warp model may be any transformation model such as affine, homography or optical flow.

But in this paper we concentrate on homography model. The minimization of the expression in equation (1) is performed with respect to $\mathbf{p}$ and the sum is performed over all of the pixels $\mathbf{x}$ in the template image T.

The Lucas-Kanade algorithm assumes that a current estimate of $\mathbf{p}$ is known and then iteratively solves for increments to the parameters $\triangle\mathbf{p}$; i.e. the following expression is minimized with respect to $\triangle\mathbf{p}$, and then the parameters are updated:

$$\sum_{x}[I(\mathbf{W}(\mathbf{x};\mathbf{p}+\triangle\mathbf{p}))-T(\mathbf{x})]^2 \tag{2}$$

$$\mathbf{p}\leftarrow\mathbf{p}+\triangle\mathbf{p} \tag{3}$$

These two steps are iterated until the estimates of the parameters converge. $\triangle\mathbf{p}$ is calculated as follows:

$$\triangle\mathbf{p}=H^{-1}\sum_{x}[\nabla I\frac{\partial\mathbf{W}}{\partial\mathbf{p}}]^T[T(\mathbf{x})-I(\mathbf{W}(\mathbf{x};\mathbf{p}))] \tag{4}$$

where H is the approximate Hessian matrix:

$$H=\sum_{x}[\nabla I\frac{\partial\mathbf{W}}{\partial\mathbf{p}}]^T[\nabla I\frac{\partial\mathbf{W}}{\partial\mathbf{p}}] \tag{5}$$

and $\nabla I=(\frac{\partial I}{\partial x},\frac{\partial I}{\partial y})$ is the gradient of image I evaluated at $\mathbf{W}(\mathbf{x};\mathbf{p})$; $\frac{\partial\mathbf{W}}{\partial\mathbf{p}}$ is the *Jacobian of the warp* and $\nabla I\frac{\partial\mathbf{W}}{\partial\mathbf{p}}$ is the *steepest descent images* [9, 11].

In [12] the Mean Structural SIMilarity ($MSSIM$) is defined for structural error measurement of two images as follows:

$$MSSIM(X,Y) = \frac{1}{M}\sum_{j=1}^{M} SSIM(x_j, y_j) \tag{6}$$

Where X and Y are the reference and the distorted images, respectively; x_j and y_j are the image contents at the j^{th} local window; M is the number of local windows of the image and the $SSIM$ is defined as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{7}$$

where C_1 and C_2 are some constants for avoiding instability, μ_x, σ_x and σ_{xy} are estimates of local statistics defined in [12]. Higher values of $MSSIM$ mean more structural similarity of X and Y.

Based on this measurement criterion, the $SDIS$ is defined in [11] as a measurement of Structural Dissimilarity:

$$SDIS(x,y) = -SSIM(x,y) \tag{8}$$

More structural difference leads to a higher value of $SDIS$. The error map of two images $I(\mathbf{W}(\mathbf{x};\mathbf{p}))$ and $T(\mathbf{x})$ based on $SDIS$ was called E_{SDIS}.

2.2. New Variation of the LK Algorithm based on SDIS and the Levenberg-Marquardt optimization

Here our goal is the optimization of the following function:

$$\sum_x E_{SDIS}.[I(\mathbf{W}(\mathbf{x};\mathbf{p})) - T(\mathbf{x})]^2 \tag{9}$$

where dot denotes the element by element multiplication as '.*' operator in MATLAB. For optimizing (9) in an iterative manner similar to (2), we have to optimize the following function:

$$\sum_x E_{SDIS}.[I(\mathbf{W}(\mathbf{x};\mathbf{p} + \triangle\mathbf{p})) - T(\mathbf{x})]^2 \tag{10}$$

where E_{SDIS} is evaluated at $\mathbf{W}(\mathbf{x};\mathbf{p})$, so it is independent to $\triangle p$[2]. Performing a first order Taylor expansion on $I(\mathbf{W}(\mathbf{x};\mathbf{p} + \triangle\mathbf{p}))$ gives:

$$SSD = \sum_x E_{SDIS}.[I(\mathbf{W}(\mathbf{x};\mathbf{p})) + \nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \triangle\mathbf{p} - T(\mathbf{x})]^2 \tag{11}$$

Finding the optimum value of $\triangle\mathbf{p}$ can be done by differentiating (11) with respect to $\triangle\mathbf{p}$, setting the result to equal zero and solving it:

$$\frac{\partial SSD}{\partial \triangle\mathbf{p}} = 2\sum_x E_{SDIS}.[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T [I(\mathbf{W}(\mathbf{x};\mathbf{p})) + \nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \triangle\mathbf{p} - T(\mathbf{x})] \tag{12}$$

[2] In appendix A it is explained why E_{SDIS} is not evaluated at $\mathbf{W}(\mathbf{x};\mathbf{p} + \triangle\mathbf{p})$

$$\frac{\partial SSD}{\partial \triangle \mathbf{p}} = 0 \Rightarrow$$

$$\sum_x E_{SDIS}.[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T \nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \triangle \mathbf{p} + \sum_x E_{SDIS}.[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T [I(\mathbf{W}(\mathbf{x};\mathbf{p})) - T(\mathbf{x})] = 0 \tag{13}$$

Hence we have:

$$\triangle \mathbf{p} = H^{-1} \sum_x E_{SDIS}.[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T [T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p}))] \tag{14}$$

where H is:

$$H = \sum_x E_{SDIS}.[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T [\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}] \tag{15}$$

In (15), H is the approximate Hessian Matrix in the Gauss-Newton method.The Levenberg-Marquardt optimization method, as an extension of Gauss-Newton method, uses the following approximation form of the Hessian matrix:

$$H_{LM} = H + \delta H_{Diag} \tag{16}$$

Where H_{Diag} is defined as follows:

$$H_{Diag} = \sum_x \begin{pmatrix} (\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p_1}})^2 & 0 & \cdots & 0 \\ 0 & (\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p_2}})^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p_8}})^2 \end{pmatrix} \tag{17}$$

Hence the approximate Hessian matrix for the Levenberg-Marquardt optimization is computed as follows:

$$H_{LM} = \sum_x E_{SDIS}.[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T [\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}] + \delta H_{Diag} \tag{18}$$

If we replace H in (14) with H_{LM} we have:

$$\triangle \mathbf{p} = H_{LM}^{-1} \sum_x E_{SDIS}.[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T [T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p}))] \tag{19}$$

The modified Lucas-Kanade algorithm based on $SDIS$ and Levenberg-Marquardt optimization is illustrated in Algorithm 1. In the original form of LK algorithm, $\triangle \mathbf{p}$ and the Hessian matrix were computed by equations (4) and (5); but in the proposed method, they are computed based on equations (14) and (15), respectively. For consistency with the unified framework, in Algorithm 1 shown below, we have not explicitly described the computation of E_{SDIS} required in equations (14) and (15). The initial approximation of warp model $\mathbf{W}(\mathbf{x};\mathbf{p})$ is computed with a feature-based registration method.

Algorithm 1 The proposed registration algorithm based on $SDIS$ and Levenberg-Marquardt optimization (LK-SSIM-LM)

Input: The reference image I, template image T and approximate estimation of the registration parameters $\mathbf{p} = (p_1, \dots, p_n)^T$ as the warp model $\mathbf{W}(\mathbf{x}; \mathbf{p})$.

Output: The tuned warp model $\mathbf{W}(\mathbf{x}; \mathbf{p})$.

1: Initialize $\delta = 0.01$.
2: Compute the gradient ∇I of $I(\mathbf{x})$.
3: Warp I with $\mathbf{W}(\mathbf{x}; \mathbf{p})$ to compute $I(\mathbf{W}(\mathbf{x}; \mathbf{p}))$.
4: Compute the error $e = \sum_x [T(x) - I(\mathbf{W}(\mathbf{x}; \mathbf{p}))]^2$
5: **repeat**
6: Compute the $SDIS$ map error image of $T(x)$ and $I(\mathbf{W}(\mathbf{x}; \mathbf{p}))$, based on (7), (8).
7: Warp the gradient ∇I with $\mathbf{W}(\mathbf{x}; \mathbf{p})$.
8: Evaluate the Jacobian $\frac{\partial \mathbf{W}}{\partial \mathbf{p}}$ at $(\mathbf{x}; \mathbf{p})$.
9: Compute the steepest descent images $\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}$.
10: Compute H_{LM} matrix using Equation (18).
11: Compute $\triangle \mathbf{p}$ using Equation (19).
12: Update the parameters $\mathbf{p} \leftarrow \mathbf{p} + \triangle \mathbf{p}$
13: Re-compute $I(\mathbf{W}(\mathbf{x}; \mathbf{p}))$.
14: Compute the new error e^*: $e^* = \sum_x [T(x) - I(\mathbf{W}(\mathbf{x}; \mathbf{p}))]^2$
15: **If** $e < e^*$ then $\delta \leftarrow \delta \times 10$, undo Steps 12–14;
else $\delta \leftarrow \delta / 10$, $e \leftarrow e^*$.
16: **until** $||\triangle \mathbf{p}|| \leq \epsilon$ or Reaching to Maximum Iteration allowed

2.3. Video Resolution Enhancement

The proposed method shown in algorithm 2 has been introduced in [5], but instead of algorithm 1, the original LK-algorithm has been used in line 6 of it. The warp model may be any transformation model such as affine, homography or optical flow. But in this paper we concentrated on the planar projective model.

An estimation of the warping model $\mathbf{W}(\mathbf{x}; \mathbf{p})$ for mapping training image T into coordinate frame of LR frame $g^{(i)}$ is found by a feature based registration model in lines 3-5. This estimation is tuned by an area-based registration method in line 6. Then the compensated form of training image T is fused with the resized form of LR frame. Mask M in line 8 is used for dealing the uncommon parts of LR frame $g^{(i)}$ and image T, which is explained below.

2.3.1. Handling Uncommon Parts

The fusion process must be done on the common parts of two images. The main source of these parts is due to moving objects in LR frames, and the objects which are visible in HR image, but not in the video frames. The usual methods for background and foreground detection which are based on background modeling and subtraction, may lead unacceptable results, due to illumination changes and camera movement. Here, we used a simple subtraction method between each LR frame ($g^{(i)}$) and the registered HR training image ($T(\mathbf{W}(\mathbf{x};\mathbf{p}))$). In line 8 of algorithm 2, mask M which illustrates the uncommon parts, is built by thresholding the subtraction image.

2.3.2. Fusion

For fusion stage of registered HR image $T(\mathbf{W}(\mathbf{x};\mathbf{p}))$ and LR frame ($g^{(i)}$), we used a version of multi-band blending approach [14] as a powerful image fusion technique. With this fusion method one can determine which regions of each image contributed in the final composite image by a mask. We produce the final HR frame $f^{(i)}$ by compositing the common parts of the registered

Algorithm 2 Video Enhancement using HR images with the proposed registration method in Algorithm 1.

Input: LR video frames $g^{(1)}, \ldots, g^{(n)}$, HR training image T, magnification factor r.

Output: HR video frames $f^{(1)}, \ldots, f^{(n)}$.

Find the SIFT key-points of HR training image.

for $i = 1$ to n **do**

  Resize $g^{(i)}$, with magnification factor r, for producing an LR image with desired number of pixels,

  Find SIFT key-points of this resized LR image,

  Remove outliers and estimate the transformation model $(\mathbf{W}(\mathbf{x}; \mathbf{p}))$,

  Tune the warp model by Algorithm 1.

  Warp T based on $\mathbf{W}(\mathbf{x}; \mathbf{p})$ onto coordinate frame of $g^{(i)}$,

  Create mask M by thresholding of subtraction of $g^{(i)}$ and $T(\mathbf{W}(\mathbf{x}; \mathbf{p}))$ for dealing uncommon parts.

  Produce $f^{(i)}$ by fusion of $g^{(i)}$ and $T(\mathbf{W}(\mathbf{x}; \mathbf{p}))$ according to inversion of M with multi-band blending approach [14].

end for

HR image and LR frame $g^{(i)}$. The multi-band blending approach guaranties the smoothness of the transition between these parts, so we have a seam-less result.

In the next section we will mention the experimental results of the proposed algorithms for image registration and its application to video enhancement.

3. Experimental Results

To demonstrate the effectiveness of the proposed video enhancement method, we have applied it to a broad variety of low-quality videos, including those corrupted by impulse noise, indoor and outdoor video sequences. Because of our assumptions in the proposed algorithms, we have to use special videos and HR training image such that (i) HR image can be transformed to each frame using planar projective model, (ii) for super-resolution comparison purposes the frames must have some displacements against each other and (iii) the moving objects must not be so large to affect the registration procedure. These restrictions prohibited us from using some common LR videos in SR context, so we used our own collected data. Table 1 shows the description of the used video sequences. The different resolution between LR video frames and HR training images could be shown by zooming. Two separate sources of motions were present in each sequence. The first kind of motion was created by moving the camera for each individual frame. The second motion was due to the changing the positions of people or waterfall (see table 1). The videos are captured with different devices.

3.1. Comparing different registration methods in algorithm 2

We ran the proposed video enhancement algorithm (Algorithm 2) using different variations of the LK-algorithm over the mentioned sequences. In line 6 of the mentioned algorithm, we tried the LK algorithm [8, 9], the LK algorithm with the Levenberg-Marquardt optimization approach [15] (LK-LM), the LK algorithm with SSIM weighting term [11] (LK-SSIM) and the proposed registration method in algorithm 1 (LK-SSIM-LM).

The Mean Square Error (MSE) between 60 synthesized frames of 'Tokyo' sequence is shown in figure 3. The mean value of each criterion was displayed

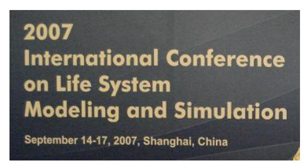
Sequence Name:	Tehran Park	LSMS Opening	Tokyo	Shanghai Garden
Frames First Original Frame Resolution Device:	60  720×576 Panasonic NV-GS75	150  320×240 Sony DSC-T100	60  640×480 Sony HDR-SR12E	86  160×112 Sony DSC-W30
First LR Frame Resolution	 360×288	 320×240	 320×240	 160×112
Training Image Resolution HR Training is: Device:	 720×576 From Seq. Panasonic NV-GS75	 740×380 Not in seq Sony DSC-W30	 640×480 Not in seq. Sony HDR-SR12E	 816×612 Not in seq. Sony DSC-W30

Table 1: Description of test sequences

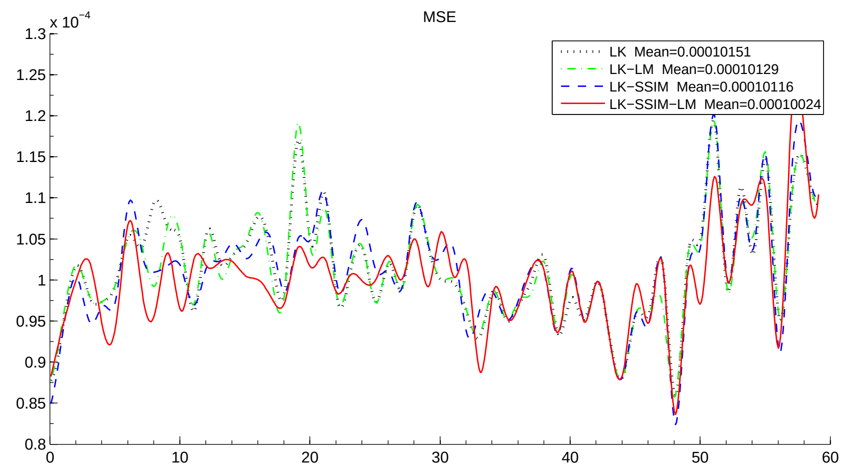


Figure 3: MSE comparison of the proposed video enhancement algorithm (Algorithm 2) using different variations of LK-algorithm for 'Tokyo' sequence .

Table 2: MSE comparison of the proposed video enhancement algorithm (Algorithm 2) using different variations of LK-algorithm over different sequences. The first and the second minimum scores are highlighted with **Bold** and *italic* letters, respectively.

MSE(10^{-6})	LK	LK-LM	LK-SSIM	LK-SSIM-LM
Tehran Park	108.31	*108.23*	108.35	**107.96**
LSMS Opening	70.86	70.86	**70.65**	*70.77*
Tokyo	101.51	101.29	101.16	**100.24**
Shanghai Garden	*203.17*	203.67	**202.87**	204.76

along with its legend. The mean error of the LK-SSIM-LM is lower than the others.

Table 2 shows the MSE results over test sequences described in table 1. As can be seen the video enhancement with the proposed LK-SSIM-LM algorithm has the highest performance with achieving the first rank in two cases and the second rank in one sequence.

When the ground-truth HR image was not available (sequences 'LSMS Opening', 'Shanghai Garden') the resized version of the LR frame (without noise) was used as the reference image. The initial approximation of warp model $\mathbf{W}(\mathbf{x};\mathbf{p})$ in algorithm 1 is computed with a feature-based registration method using SIFT key-points [5]. Finding the homography matrix has been done using the RANSAC[3] method [16]. Since RANSAC is a random nature method, for each pair of images, the initial warp model has been found once and the resulting homography was used as the initial warp model for each of LK algorithm's variations. Thus the comparisons are not affected by random nature of the RANSAC method.

3.2. Comparing with different supper-resolution methods

We applied our proposed method in algorithm 2 on aforementioned test sequences and compared its performance with some other super-resolution algorithms. We used the 'sliding-window' techniques with the Interpolation (IN), Iterated Back-projection(BP)[1] and Robust Super-resolution(RS)[2] as reconstruction methods. Computing the motion parameters between frames has been done using the registration method of Keren *et. al.*[17] . The magnification factor r and the window size were set to 2 and 4 respectively.

Table 3 shows quantitative comparisons of the mentioned methods based on Mean Absolute Error (MAE), Power Signal to Noise Ratio (PSNR) and SSIM for the test sequences, in which:

[3] RANdom SAmple Consensus (RANSAC)

Table 3: MAE, PSNR and SSIM comparisons of the proposed video enhancement algorithm (Algorithm 2) and some super-resolution reconstruction methods over different sequences. The best score is highlighted with **Bold** letters in each row.

Method:	BP	IN	RS	Algorithm 2
MAE($\times 10^{-3}$)				
Tehran Park	102.46	54.16	82.57	**47.53**
LSMS Opening	61.99	34.90	52.02	**28.48**
Tokyo	133.43	55.94	96.70	**38.46**
Shanghai Garden	84.92	47.46	67.90	**35.08**
PSNR				
Tehran Park	17.58	22.44	19.30	**23.19**
LSMS Opening	22.21	25.99	23.09	**28.15**
Tokyo	15.44	22.10	17.97	**25.14**
Shanghai Garden	19.56	23.73	21.26	**25.28**
SSIM				
Tehran Park	0.24	0.50	0.31	**0.63**
LSMS Opening	0.45	0.62	0.47	**0.69**
Tokyo	0.13	0.39	0.20	**0.60**
Shanghai Garden	0.31	0.45	0.36	**0.70**

$$MAE = \frac{\Sigma_{i=1}^{N}\Sigma_{j=1}^{M}\Sigma_{q=1}^{Q}|F^q(i,j) - \hat{F}^q(i,j)|}{N.M.Q} \tag{20}$$

and

$$PSNR = 10 \times log\Big(\frac{255^2}{\frac{1}{N.M.Q}\Sigma_{i=1}^{N}\Sigma_{j=1}^{M}\Sigma_{q=1}^{Q}\big(F^q(i,j) - \hat{F}^q(i,j)\big)^2}\Big) \tag{21}$$

where M, N are the image dimensions, Q is the number of channels of the image ($Q = 3$ for color image), and $F^q(i,j)$ and $\hat{F}^q(i,j)$ denote the qth component of the original image vector and the distorted image, at pixel position (i,j), respectively. In these experiments, the mentioned criteria has been computed over gray scale version of images (Q=1). The best score is highlighted with **Bold** letters for each sequence in table 3. As can be seen the proposed method has the highest performance for all MAE, PSNR and SSIM criteria.

The 'Tokyo' sequence which its high-resolution version is available has been used here for further comparisons. Figure 4 shows MAE, PSNR and SSIM of the proposed video enhancement algorithm (Algorithm 2) and some super-resolution reconstruction methods over 60 frames of 'Tokyo' sequence. The superior performance of the proposed method is obvious.

Figure 5 shows close-up demonstrations of an instance frame, produced by some different methods. The original HR frame, the nearest and the bicubic resized versions of that frame has been shown for comparison purposes. Note that the windows of the rear building in the frame, is almost completely unrecognizable in the LR video frames and in the other super-resolution methods, except that of the proposed method (5(g)). The resolution is clearly enhanced and the mentioned windows are now visible.

Figure 6 shows SDIS map image

4. Concluding Remarks

In this paper a new version of the popular Lucas-Kanade image registration algorithm has been proposed and applied to video enhancement. Our goal is the enhancement of low resolution video frames, by fusion a motion compensated form of a high resolution image. The high resolution image is

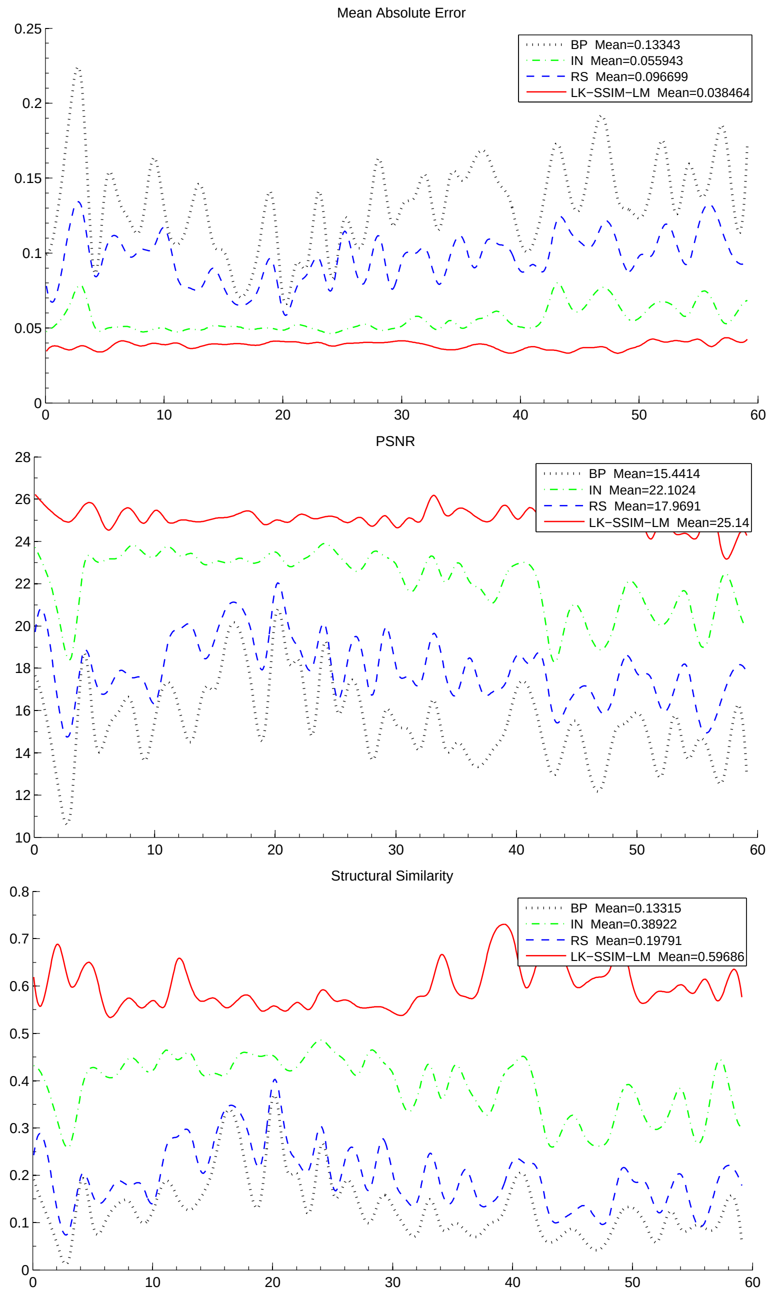
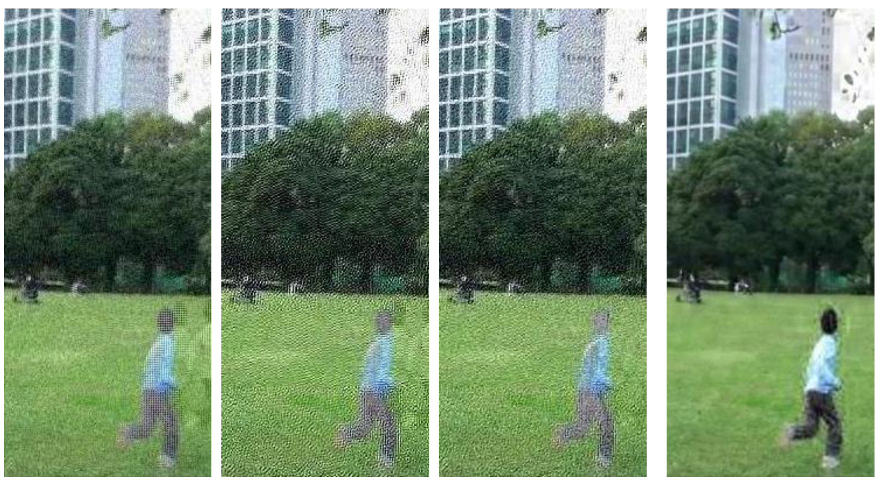


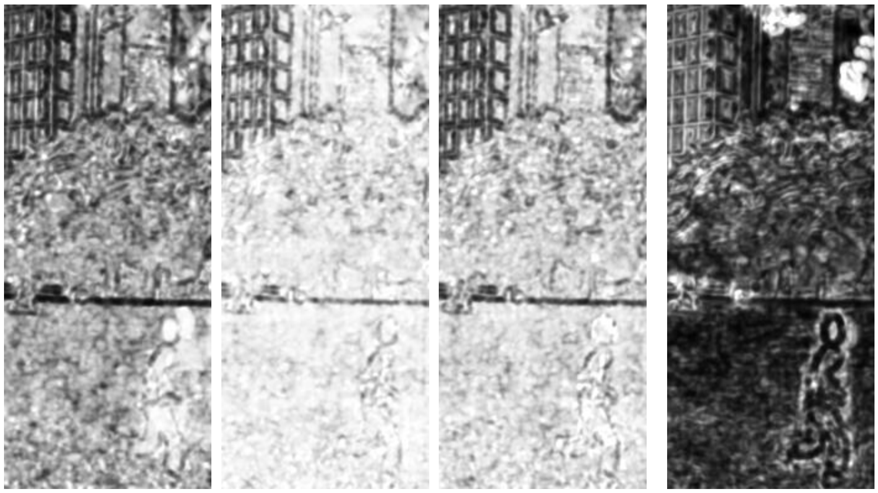
Figure 4: MAE, PSNR and SSIM comparison of the proposed video enhancement algorithm (Algorithm 2) and some super-resolution reconstruction methods for 'Tokyo' sequence .

(a) Original HR frame (b) LR frame (Nearest) (c) LR frame (Bicubic)



(d) Interpolation (e) Iterated Back-projection (f) Robust Super-resolution (g) This paper

Figure 5: Close-up views of the original HR image, replication (nearest) and bicubic resizing methods, super-resolution reconstruction methods: Interpolation, Iterated Back-projection[1] and Robust Super-resolution [2] and the proposed method in Algorithm 2 on 'Tokyo' sequence.



(a) Interpolation (b) Iterated Back-projection (c) Robust Super-resolution (d) This paper

Figure 6: SDIS map image (E_{SDIS}) between figures 5(d)-5(g) with related HR image (5(a)). Brighter pixel means higher error.

from the same scene of the video but perhaps with a different resolution, different illumination and color and slightly different capturing view. The precise mapping of this image onto each video frame has been done with the proposed registration method. In the registration stage structural similarity metric used as a weighting term of the objective function of LK algorithm. The SSIM criterion exhibited very good consistency with a qualitative visual appearance and when the signal to noise of each video frame is low it reflects the structural differences of two images much batter than absolute error map. The mathematical derivation of the proposed approach using the Levenberg-Marquardt optimization method, based on the unified framework of Baker *et. al.*[9] was given. The accuracy of the proposed registration method is compared with some variations of LK-algorithm. The experimental results over video super-resolution using the mentioned registration algorithm, showed the superior performance of the proposed method against some other methods in terms of final perceived quality and objective comparisons.

Acknowledgment

The authors are indebted to Dr. Vandewalle [18] for his Super-Resolution package and Dr. D. Lowe for his SIFT key-points program[4]. They also thank to Dr. Peter Kovesi [5] Dr. Simon Baker and his co-workers [9] for providing many useful MATLAB functions and Dr. Gh. Mohajeri for 'Tokyo' sequence.

[1] M. Irani, S. Peleg, Improving resolution by image registration, CVGIP: Graph. Models Image Process. 53 (3) (1991) 231–239. 2, 17, 21

[2] A. Zomet, A. Rav-Acha, S. Peleg, Robust super resolution, in: Proceedings of the Int. Conf. on Computer Vision and Patern Recognition (CVPR), 2001, pp. 645–650. 2, 17, 21

[3] S. Borman, Topics in multiframe superresolution restoration, Ph.D. thesis, University of Notre Dame, Notre Dame, IN (May 2004). 2

[4] K. Watanabe, Y. Iwai, H. Nagahara, M. Yachida, T. Suzuki, Video synthesis with high spatio-temporal resolution using motion compensation and spectral fusion, IEICE - Trans. Inf. Syst. E89-D (7) (2006) 2186–2196. 3

[4] Available online at: http://www.cs.ubc.ca/∼lowe/keypoints/

[5] http://www.csse.uwa.edu.au/∼pk/research/matlabfns/

[5] M. Amintoosi, M. Fathy, N. Mozayani, Video resolution enhancement in the presence of moving objects, in: International Conference on Image Processing, Computer Vision, and Pattern Recognition, Las Vegas, USA, 2009. 3, 11, 17

[6] M. Amintoosi, M. Fathy, N. Mozayani, Reconstruction+synthesis: A hybrid method for multi-frame super-resolution, in: (MVIP08) 2008 Iranian Conference on Machine Vision and Image Processing, Tabriz University, Tabriz, Iran, 2008, pp. 179–184. 3

[7] D. G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2) (2004) 91–110. 3

[8] B. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: IJCAI81, 1981, pp. 674–679. 3, 14

[9] S. Baker, R. Gross, I. Matthews, Lucas-kanade 20 years on: A unifying framework, International Journal of Computer Vision 56 (2004) 221–255. 3, 5, 6, 14, 22, 23

[10] M. Zhao, Video enhancement using content-adaptive least mean square filters, Ph.D. thesis, Technische Universiteit Eindhoven (2006). 3

[11] M. Amintoosi, M. Fathy, N. Mozayani, Precise image registration with structural similarity error measurement applied to super-resolution, EURASIP Journal on Applied Signal Processing 2009 (2009) 7 pages, Article ID 305479. 3, 4, 6, 7, 14

[12] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: From

error visibility to structural similarity, IEEE Trans. Image Processing 13 (4) (2004) 600–612. 3, 7, 26

[13] A. Brooks, What makes an image look good?, presentation for the Image & Video Processing Lab (IVPL) at Northwestern University for ECE 510 Video Processing (March 17 2005).
URL http://dailyburrito.com/projects/ImageQuality_Brooks2005.pdf 4

[14] P. J. Burt, E. H. Adelson, A multiresolution spline with application to image mosaics, ACM Trans. Graph. 2 (4) (1983) 217–236. 12, 13

[15] R. Szeliski, Video mosaics for virtual environments, IEEE Computer Graphics and Applications 16 (2) (1996) 22–30. 14

[16] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395. 17

[17] D. Keren, S. Peleg, R. Brada, Image sequence enhancement using sub-pixel displacement, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 1988, pp. 742–746. 17

[18] P. Vandewalle, S. Süsstrunk, M. Vetterli, A Frequency Domain Approach to Registration of Aliased Images with Application to Super-Resolution, EURASIP Journal on Applied Signal Processing (special issue on Super-resolution) 2006 (2006) Article ID 71459, 14 pages. 23

Appendix A. On the derivation of the proposed algorithm based on $\boldsymbol{E_{SDIS}}(\mathbf{W}(\mathbf{x};\mathbf{p}))$

In equation (10) in section 2.2 we mentioned that "E_{SDIS} is evaluated at $\mathbf{W}(\mathbf{x};\mathbf{p})$", here we discuss why E_{SDIS} *is not evaluated at* $\mathbf{W}(\mathbf{x};\mathbf{p}+\triangle\mathbf{p})$. Suppose that E_{SDIS} is evaluated at $\mathbf{W}(\mathbf{x};\mathbf{p}+\triangle\mathbf{p})$, rewriting eq. (10) based on this assumption yields:

$$\sum_x E_{SDIS}(\mathbf{W}(\mathbf{x};\mathbf{p}+\triangle\mathbf{p})).[I(\mathbf{W}(\mathbf{x};\mathbf{p}+\triangle\mathbf{p}))-T(\mathbf{x})]^2 \tag{A.1}$$

Performing a first order Taylor expansion on $E_{SDIS}(\mathbf{W}(\mathbf{x};\mathbf{p}+\triangle\mathbf{p}))$ and $I(\mathbf{W}(\mathbf{x};\mathbf{p}+\triangle\mathbf{p}))$ gives:

$$SSD = \sum_x [E_{SDIS}(\mathbf{W}(\mathbf{x};\mathbf{p}))+\nabla E_{SDIS}\frac{\partial \mathbf{W}}{\partial \mathbf{p}}\triangle\mathbf{p}].[I(\mathbf{W}(\mathbf{x};\mathbf{p}))+\nabla I\frac{\partial \mathbf{W}}{\partial \mathbf{p}}\triangle\mathbf{p}-T(\mathbf{x})]^2 \tag{A.2}$$

It should be mentioned that according to [12], E_{SDIS} is differentiable. Finding the optimum value of $\triangle\mathbf{p}$ can be done by differentiating (A.2) with respect to $\triangle\mathbf{p}$, setting the result to equal zero and solving it:

$$\begin{aligned}\frac{\partial SSD}{\partial \triangle\mathbf{p}} = \sum_x \Big(&[\nabla E_{SDIS}\frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T[I(\mathbf{W}(\mathbf{x};\mathbf{p}))+\nabla I\frac{\partial \mathbf{W}}{\partial \mathbf{p}}\triangle\mathbf{p}-T(\mathbf{x})]^2+\\ &2[\nabla I\frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T[I(\mathbf{W}(\mathbf{x};\mathbf{p}))+\nabla I\frac{\partial \mathbf{W}}{\partial \mathbf{p}}\triangle\mathbf{p}-T(\mathbf{x})][E_{SDIS}(\mathbf{W}(\mathbf{x};\mathbf{p}))+\\ &\nabla E_{SDIS}\frac{\partial \mathbf{W}}{\partial \mathbf{p}}\triangle\mathbf{p}]\Big)\end{aligned} \tag{A.3}$$

For simplicity of driving we define the following terms:

$$
\begin{aligned}
A &= [E_{SDIS}(\mathbf{W}(\mathbf{x};\mathbf{p})) + \nabla E_{SDIS}\frac{\partial \mathbf{W}}{\partial \mathbf{p}}\triangle\mathbf{p}] \\
B &= [I(\mathbf{W}(\mathbf{x};\mathbf{p})) + \nabla I\frac{\partial \mathbf{W}}{\partial \mathbf{p}}\triangle\mathbf{p} - T(\mathbf{x})] \\
I &= I(\mathbf{W}(\mathbf{x};\mathbf{p})) \\
T &= T(x)
\end{aligned}
$$

$$
\begin{aligned}
E &= E_{SDIS}(\mathbf{W}(\mathbf{x};\mathbf{p})) \\
e &= I(\mathbf{W}(\mathbf{x};\mathbf{p})) - T(\mathbf{x}) \\
S_E &= [\nabla E_{SDIS}\frac{\partial \mathbf{W}}{\partial \mathbf{p}}] \qquad ,\textbf{Steepest descent image of E} \\
S_I &= [\nabla I\frac{\partial \mathbf{W}}{\partial \mathbf{p}}] \qquad ,\textbf{Steepest descent image of I}
\end{aligned} \tag{A.4}
$$

Hence equation (A.3) can be simplified as follows:

$$
\begin{aligned}
\frac{\partial SSD}{\partial \triangle\mathbf{p}} &= \sum_x \left[S_E^T B^2 + 2BS_I^T A\right] \\
&= \sum_x \left[(S_E^T B + 2S_I^T A)B\right]
\end{aligned} \tag{A.5}
$$

The above factorization is legal; because the distribution of multiplication over addition is hold for '.' operator (which denotes '.*' operator in MATLAB) :

$$X.Z + Y.Z = (X + Y).Z \tag{A.6}$$

For simplicity we temporary drop the summation operator $\sum_x$, from equation (A.5); $B = 0$ or $(S_E^T B + 2AS_I^T) = 0$ are the sufficient conditions such that $\frac{\partial SSD}{\partial \triangle \mathbf{p}} = 0$.

If $B = 0$ then from our definitions in (A.4) and regarding the summation, we will have:

$$\triangle \mathbf{p} = -\frac{e}{S_I} = \frac{\sum_x [T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p}))]}{\sum_x [\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T} \tag{A.7}$$

which is non-acceptable, because the the size of denominator is $n \times 1$ and hence it is not invertible.

If $(S_E^T B + 2S_I^T A) = 0$, we will have:

$$\begin{aligned} S_E^T B + 2S_I^T A = 0 \quad &\Rightarrow \\ [\nabla E_{SDIS} \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T [I(\mathbf{W}(\mathbf{x};\mathbf{p})) + \nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \triangle \mathbf{p} - T(\mathbf{x})] & \\ +2[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T [E_{SDIS}(\mathbf{W}(\mathbf{x};\mathbf{p})) + \nabla E_{SDIS} \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \triangle \mathbf{p}] = 0 \quad &\Rightarrow \end{aligned}$$

$$\begin{aligned} S_E^T [I + S_I \triangle \mathbf{p} - T] + 2S_I^T [E + S_E \triangle \mathbf{p}] = 0 \quad &\Rightarrow \\ S_E^T I + S_E^T S_I \triangle \mathbf{p} - S_E^T T + 2S_I^T E + 2S_I^T S_E \triangle \mathbf{p} = 0 \quad &\Rightarrow \\ S_E^T I - S_E^T T + 2S_I^T E + (S_E^T S_I + 2S_I^T S_E) \triangle \mathbf{p} = 0 \quad &\Rightarrow \\ \triangle \mathbf{p} = -\frac{S_E^T (I - T) + 2S_I^T E}{S_E^T S_I + 2S_I^T S_E} = -\frac{S_E^T e + 2S_I^T E}{S_E^T S_I + 2S_I^T S_E} & \end{aligned} \tag{A.8}$$

Based on the definitions in (A.4) and regarding the summation, we will have:

$$\triangle \mathbf{p} = -H^{-1} \sum_{x} \big([\nabla E_{SDIS} \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T [I(\mathbf{W}(\mathbf{x};\mathbf{p})) - T(\mathbf{x})] + 2[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T E_{SDIS}\big) \tag{A.9}$$

where H is:

$$H = \sum_{x} \big([\nabla E_{SDIS} \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T [\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}] + 2[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^T [\nabla E_{SDIS} \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]\big) \tag{A.10}$$

But our implementation based on (A.10) did not produce satisfactory results. The reason may be due to non homogeneous nature of S_E and S_I (and also E and e) in (A.9). This makes the computations to be wrong and even affects the singularity of H in some examples.