Video enhancement 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 16846-13114, Iran

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 (SSIM) 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 optimisation 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 enhancement successfully. Various objective and subjective comparisons show the superior performance of the proposed method.

Keywords: video enhancement, registration, super-resolution, structural similarity, synthesis, Levenberg–Marquardt

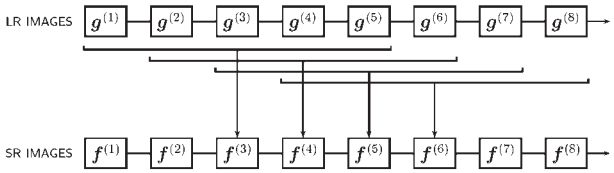
1 INTRODUCTION

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 SR algorithms reconstruct a HR video from a LR video. The vast majority of the SR restoration algorithms — named as reconstruction methods — use a short sequence of LR input frames to produce a single super-resolved HR output frame.[13,31] These techniques have been applied to video restoration by using a shifting window of processed frames as illustrated in Fig. 1. For a given SR frame, a 'sliding window' determines the subset of LR frames to be processed to produce a given SR output frame. The window is moved forward in time to produce successive super-resolved frames in the output sequence.[7]

The analysis performed by Lin and Shum[16] indicates that to achieve super resolution at large magnification factors, reconstruction-based algorithms are not favorable and one should try other kinds of super resolution algorithms, such as recognition-based algorithms. Hence, recent advances in SR techniques show trends towards methods which consider some prior knowledge or models, in addition to LR images as the input of the SR algorithm.[5,21] This can be considered as a special class of SR methods, named as learning-based methods.[11] These model-based approaches differ from the reconstruction-based approach in the final step where high-frequency details are recovered from the reconstructed (but possibly blurry) HR image followed by applying a fusion technique. Instead of deconvolution, the model-based approach imports plausible high-frequency textures from an image database into the HR image. These

The MS was accepted for publication on 16 October 2010.
* *Corresponding author: Mahmood Amintoosi, Computer Engineering Department, Iran University of Science and Technology, Narmak, Tehran 16846-13114, Iran; email: mamintoosi@iust.ac.ir*

 IMAG 40 © RPS 2011 DOI: 10.1179/1743131X10Y.0000000014



1 'Sliding window' technique for video super resolution[7]

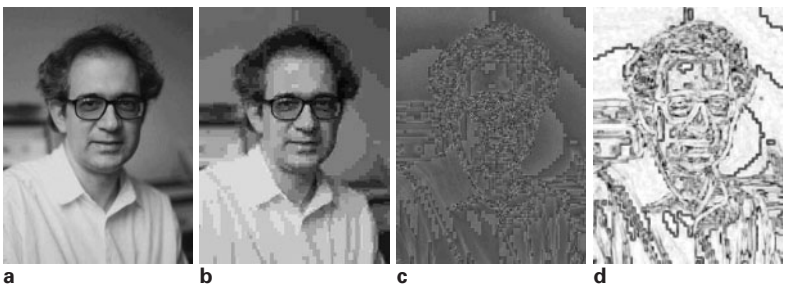
methods has gained significant interests in recent years because it promises to overcome the limit of reconstruction-based SR.[21] Freeman *et al.*[11] used a set of HR images as training dataset. For each patch of LR image, they searched the training set for finding a match. The corresponding high-frequency patch of the best match has been selected for enhancing the resolution of the LR patch. The output of Chang *et al.*[9] is not significantly different from that applying median filtering.

Although the mentioned methods has already shown an impressive performance, there is still room for improvement if we do not restrict ourself to small patches. Recently, 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.[2,28] 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 Amintoosi *et al.*,[1] a feature-based registration approach using scale invariant feature transform (SIFT) key points[17] has been used for mapping the HR image onto LR frame coordinates. Our approach in Ref. 2 — which is also used here — is based on using a feature based registration method followed by applying an area-based registration method. The estimated motion parameters by feature based method, are tunned by Lucas–Kanade (LK) algorithm as a powerful registration method. 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.[6] The core part of this algorithm is to find the registration parameters with minimisation of the square error between the reference image and a motion compensated of another distorted 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 Zhao,[30] automatic optimisation based on a reliable subjective metric seems a challenging target for future video enhancement research. In this paper, a new version of LK algorithm is introduced and applied to video enhancement. In the proposed approach, structural similarity (SSIM) error measurement[26] is contributed to the LK algorithm as a weighting term of its objective function. The chief idea of this approach is based on the fact that the contrast inverted form of SSIM highlights the structural differences of two images, much better than the absolute error map, in particular 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.

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 superior results.

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



2 Comparing the error maps of two images based on MSE and SSIM. The images are taken from Ref. 8: (a) original image; (b) JPEG compressed; (c) MSE map; (d) SSIM map

2 THE PROPOSED METHOD

This section is categorised 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 optimisation method. Then the application of this method for video resolution enhancement is explained.

2.1 Related concepts

2.1.1 LK algorithm

The goal of the LK algorithm is to align a template image T to an input image I, by minimising 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, $\mathbf{p}=(p_1, p_2, \ldots, p_n)^T$ 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.[6] 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 minimisation 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 LK algorithm assumes that a current estimate of $\mathbf{p}$ is known and then iteratively solves for increments to the parameters $\Delta\mathbf{p}$, i.e. the following expression is minimised with respect to $\Delta\mathbf{p}$, and then the parameters are updated:

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

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

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

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

where H is the approximate Hessian matrix:

$$H = \sum_{x} \left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}\right)^T \left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}\right) \tag{5}$$

and $\nabla I = (\partial I/\partial x, \partial I/\partial y)$ is the gradient of image I evaluated at $\mathbf{W}(\mathbf{x};\mathbf{p})$; $\partial\mathbf{W}/\partial\mathbf{p}$ is the *Jacobian of the warp* and $\nabla I(\partial\mathbf{W}/\partial\mathbf{p})$ is the *steepest descent images*.[6]

2.1.2 SSIM error measurement

In Ref. 26, 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 jth 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, and μ_x, σ_x and σ_{xy} are estimates of local statistics defined in Ref. 26. Higher values of *MSSIM* mean more structural similarity of X and Y.

2.2 New variation of the LK algorithm based on SSIM and the Levenberg–Marquardt optimisation

As can be seen in Fig. 2, SSIM captures structural errors better than absolute error. Hence one can expect that contributing the SSIM onto the LK algorithm's minimisation function promises better result than its original form which is based on usual image difference. Among the various inverted form of SSIM, such as '1/*SSIM*', '1–*SSIM*' and '$-$*SSIM*', we choose its negative form, and called it *SDIS* as structural dissimilarity measurement:

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

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}.

With contributing E_{SDIS} into the objective function of LK algorithm, the goal will be the optimisation of the following function:[3]

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

where dot denotes the element by element multiplication as '.*' operator in MATLAB. For

optimising equation (9) in an iterative manner similar to equation (2), we have to optimise the following function:

$$\sum_{\mathrm{x}} E_{\mathrm{SDIS}} \cdot [I(\mathbf{W}(\mathbf{x};\mathbf{p}+\Delta\mathbf{p})) - T(\mathbf{x})]^2 \tag{10}$$

where E_{SDIS} is evaluated at $\mathbf{W}(\mathbf{x};\mathbf{p})$, so it is independent to Δp [in Appendix 1, it is explained why E_{SDIS} is not evaluated at $\mathbf{W}(\mathbf{x};\mathbf{p}=\Delta\mathbf{p})$]. Performing a first order Taylor expansion on $I(\mathbf{W}(\mathbf{x};\mathbf{p}+\Delta\mathbf{p}))$ gives:

$$SSD = \sum_{\mathrm{x}} E_{\mathrm{SDIS}} \cdot \left[I(\mathbf{W}(\mathbf{x};\mathbf{p})) + \nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \Delta\mathbf{p} - T(\mathbf{x}) \right]^2 \tag{11}$$

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

$$\frac{\partial SSD}{\partial \Delta\mathbf{p}} = 2 \sum_{\mathrm{x}} E_{\mathrm{SDIS}} \cdot \left[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right]^{\mathrm{T}} \left[I(\mathbf{W}(\mathbf{x};\mathbf{p})) + \nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \Delta\mathbf{p} - T(\mathbf{x}) \right] \tag{12}$$

$$\frac{\partial SSD}{\partial \Delta\mathbf{p}} = 0 \Rightarrow \sum_{\mathrm{x}} E_{\mathrm{SDIS}} \cdot \left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right)^{\mathrm{T}} \nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \Delta\mathbf{p} + \sum_{\mathrm{x}} E_{\mathrm{SDIS}} \cdot \left[\left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right)^{\mathrm{T}} [I(\mathbf{W}(\mathbf{x};\mathbf{p})) - T(\mathbf{x})] \right] = 0 \tag{13}$$

Hence we have:

$$\Delta\mathbf{p} = H^{-1} \sum_{\mathrm{x}} E_{\mathrm{SDIS}} \cdot \left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right)^{\mathrm{T}} [T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p}))] \tag{14}$$

where H is:

$$H = \sum_{\mathrm{x}} E_{\mathrm{SDIS}} \cdot \left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right)^{\mathrm{T}} \left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right) \tag{15}$$

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

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

where H_{Diag} is defined as follows:

$$H_{\mathrm{Diag}} = \sum_{\mathrm{x}} \begin{pmatrix} \left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}_1} \right)^2 & 0 & \cdots & 0 \\ 0 & \left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}_2} \right)^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}_8} \right)^2 \end{pmatrix} \tag{17}$$

Hence, the approximate Hessian matrix for the Levenberg–Marquardt optimisation is computed as follows:

$$H_{\mathrm{LM}} = \sum_{\mathrm{x}} E_{\mathrm{SDIS}} \cdot \left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right)^{\mathrm{T}} \left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right) + \delta H_{\mathrm{Diag}} \tag{18}$$

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

$$\Delta\mathbf{p} = H_{\mathrm{LM}}^{-1} \sum_{\mathrm{x}} E_{\mathrm{SDIS}} \cdot \left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right)^{\mathrm{T}} [T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p}))] \tag{19}$$

The modified LK algorithm based on *SDIS* and Levenberg–Marquardt optimisation is illustrated in Algorithm 1. In the original form of LK algorithm, $\Delta\mathbf{p}$ 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 optimisation (LK–SSIM–LM)

Input: The reference image I, template image T and approximate estimation of the registration parameters $\mathbf{p}=(p_1, p_2, \ldots, p_{\mathrm{n}})^{\mathrm{T}}$ as the warp model $\mathbf{W}(\mathbf{x};\mathbf{p})$.

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

1. Initialise $\delta = 0.01$.
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 $e = \sum_{\mathrm{x}} [T(x) - I(\mathbf{W}(\mathbf{x};\mathbf{p}))]^2$
5. **repeat**.
6. Compute the *SDIS* map error image of $T(x)$ and $I(\mathbf{W}(\mathbf{x};\mathbf{p}))$, based on equations (7) and (8).
7. Warp the gradient ∇I with $\mathbf{W}(\mathbf{x};\mathbf{p})$.
8. Evaluate the Jacobian $\partial\mathbf{W}/\partial\mathbf{p}$ at $(\mathbf{x};\mathbf{p})$.
9. Compute the steepest descent images $\nabla \mathbf{I}(\partial\mathbf{W}/\partial\mathbf{p})$.
10. Compute H_{LM} matrix using equation (18).
11. Compute $\Delta\mathbf{p}$ using equation (19).
12. Update the parameters $\mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}$
13. Re-compute $I(\mathbf{W}(\mathbf{x};\mathbf{p}))$.
14. Compute the new error e^*: $e^* = \sum_{\mathrm{x}} [T(x) - I(\mathbf{W}(\mathbf{x};\mathbf{p}))]^2$
15. **If** $e < e^*$ then $\delta \leftarrow \delta \times 10$, undo Steps 12–14; **else** $\delta \rightarrow \delta/10$, $e \leftarrow e^*$.
16. **until** $\|\Delta\mathbf{p}\| \leqslant \varepsilon$ or reaching to maximum iteration allowed.

2.3 Video resolution enhancement

The proposed method shown in Algorithm 2 has been introduced by us in Ref. 2, 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 the uncommon 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 modelling and subtraction, may lead to 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.

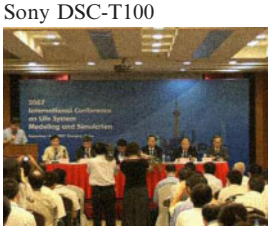
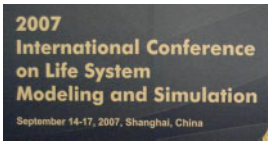
Algorithm 2 Video enhancement using HR images with the proposed registration method in Algorithm 1.

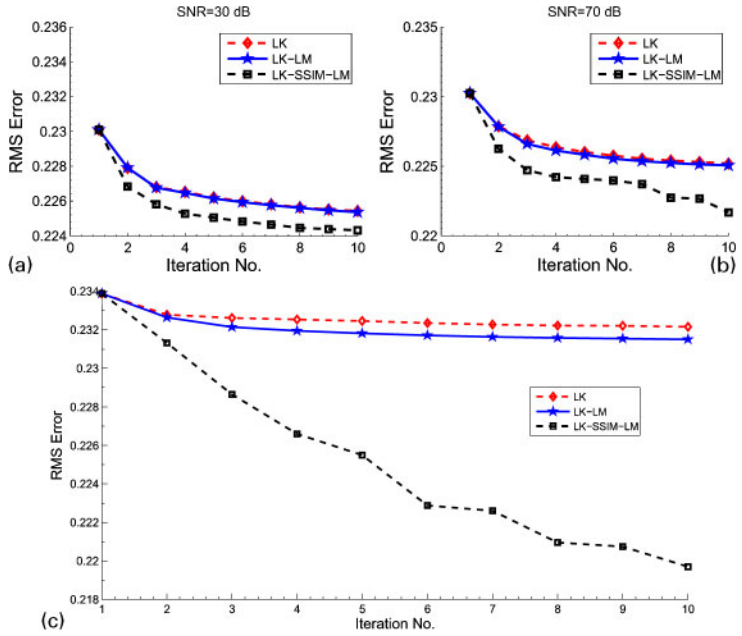
Input: LR video frames $g^{(1)}$, $g^{(2)}$, …, $g^{(n)}$, HR training image T, magnification factor r.

Output: HR video frames $f^{(1)}$, $f^{(2)}$, …, $f^{(n)}$.

1. Find the SIFT key points of HR training image.
2. **for** i=1–n **do**.
3. Resize $g^{(i)}$, with magnification factor r, for producing an LR image with desired number of pixels.

Table 1 Description of test sequences (available at: http://webpages.iust.ac.ir/mamintoosi/Research/Video_Enhancement/)

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



3 Convergence comparison of the proposed LK–SSIM–LM algorithm and two other variations of LK algorithm for 'Tokyo' sequence: (a) LR images: SNR=30 db; (b) LR images: SNR=70 db; (c) mean RMS over all experiments for SNRs=20, 30, 50 and 70 dB

4. Find SIFT key points of this resized LR image.
5. Remove outliers and estimate the transformation model ($\mathbf{W}(\mathbf{x};\mathbf{p})$).
6. Tune the warp model by Algorithm 1.
7. Warp T based on $\mathbf{W}(\mathbf{x};\mathbf{p})$ onto coordinate frame of $g^{(i)}$.
8. Create mask M by thresholding of subtraction of $g^{(i)}$ and $T(\mathbf{W}(\mathbf{x};\mathbf{p}))$ for dealing uncommon parts.
9. Produce $f^{(i)}$ by fusion of $g^{(i)}$ and $T(\mathbf{W}(\mathbf{x};\mathbf{p}))$ according to inversion of M with a version of multi-band blending approach.[4]
10. **end for**.

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[4] 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 HR image and LR frame $g^{(i)}$. The multi-band blending approach guarantees 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 range of low-quality videos, including those corrupted by impulse noise and video sequences taken in indoor and outdoor environment. Because of our assumptions in the proposed algorithms, we have to use special videos and HR training image such that: (1) HR image can be transformed to each frame using planar projective model; (2) for SR comparison purposes, the frames must have some displacements against each other; and (3) 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 can be seen 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 (Table 1). The videos are captured with different devices.

For comparison, the following criteria are used: mean square error (MSE), root mean square error (RMS), mean absolute error (MAE), power-signal-to-noise ratio (PSNR) and SSIM, in which:

$$MSE = \frac{\Sigma_{i=1}^{N}\Sigma_{j=1}^{M}\Sigma_{q=1}^{Q}\left[F^{q}(i,j) - \hat{F}^{q}(i,j)\right]^{2}}{N \cdot M \cdot Q} \tag{20}$$

$$RMS = (MSE)^{1/2} \tag{21}$$

$$MAE = \frac{\Sigma_{i=1}^{N}\Sigma_{j=1}^{M}\Sigma_{q=1}^{Q}|F^{q}(i,j) - \hat{F}^{q}(i,j)|}{N \cdot M \cdot Q} \tag{22}$$

and

$$PSNR = 10\log\left[\frac{255^{2}}{(1/N.M.Q)\Sigma_{i=1}^{N}\Sigma_{j=1}^{M}\Sigma_{q=1}^{Q}\left[F^{q}(i,j) - \hat{F}^{q}(i,j)\right]^{2}}\right] \tag{23}$$

where M and N are the image dimensions, Q is the number of channels of the image ($Q=3$ for colour 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 have been computed over gray-scale version of images ($Q=1$).

The *SSIM* criterion was first introduced for still images[26] and then extended for video sequences.[27] Extending of *SSIM* for video frames is a weighted version of *SSIM* over local windows and video frames. As indicated by Wang,[27] 'the proposed method without any weighting adjustment provides reasonably good results compared with the other approaches'. In the following experiments, the *SSIM* without any weighting adjustment has been used for video sequences (The modified SSIM function, some of the main MATLAB files and the dataset used in this paper can be downloaded from the following address: http://webpages.iust.ac.ir/mamintoosi/Research/Video_Enhancement/).

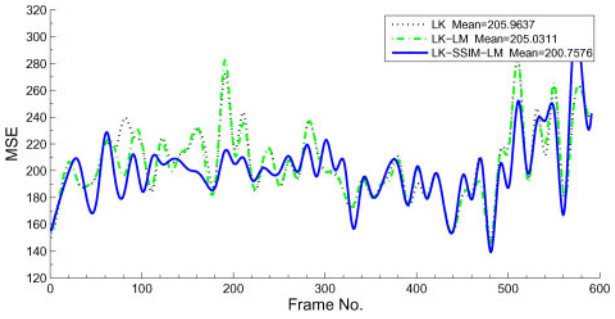
In the first part of this section, we will mention the experimental results for comparing the image registration algorithm (Algorithm 1). In the second part, we will use the proposed method in Algorithm 2 on an image SR application.

3.1 Registration comparison

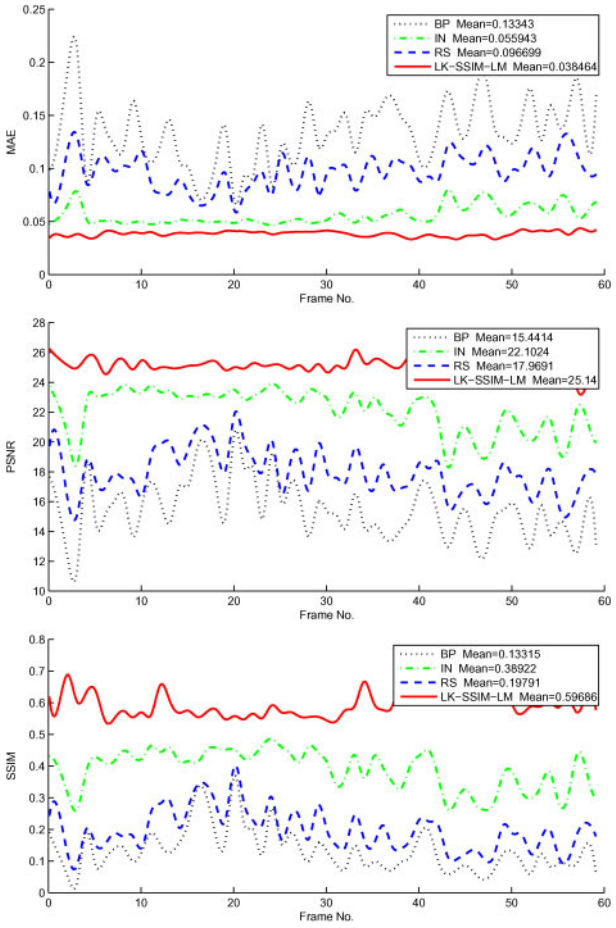
In this section, the proposed LK–SSIM–LM algorithm with two variations of LK algorithm, the original LK algorithm and the LK–LM algorithm are compared in various forms. At first, the convergences of three algorithms are compared; then we will compare the resulting SR image, when in the registration stage of Algorithm 2, each of the aforementioned algorithms is used. The better performance of the proposed method will be clear with the experimental results in the following sections.

3.1.1 Registration convergence

In each iteration of LK algorithm, the RMS error between the registered image and the original image is calculated. In each experiment, the LR images are corrupted with noise so that the signal-to-noise ratio (SNR) of the resulting image be a predefined value (SNRs=20, 30, 50 and 70 dB are tested in the experiments here). The maximum iteration count in all variations of LK algorithm is 10. Figure 2 shows the average error over video frames in each iteration for 'Tokyo' sequence when the SNR equals 30 dB. Since 'Tokyo' sequence has 60 frames, each point in Fig. 2 is the average RMS error of 60 images in each cycle of each algorithm. The convergence result, when SNR=70 dB, illustrated in Fig. 2, is also similar to that in the previous case (SNR=30 dB). As can be seen, in both cases, the proposed LK–SSIM–LM algorithm converges faster. Figure 2 shows the mean value of RMS over all frames of 'Tokyo' sequence and SNRs=20, 30, 50 and 70 dB. Each point in this figure is the average RMS error of 240 images in each cycle of each algorithm over various amount of noise. Faster convergence of the



4 MSE comparison of the proposed video enhancement algorithm (Algorithm 2) using different variations of LK algorithm for 'Tokyo' sequence



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

proposed approach with respect to other algorithms is obvious.

3.1.2 *Running time comparison*

Since the computing of *SSIM* is not complicated, it does not increase the overall time significantly. For instance, in average each cycles of the LK–LM and LK–SSIM–LM algorithms for 'Tokyo' sequence, took 1.583 and 1.578 s; hence, the difference is negligible.

Table 2 MSE comparison of the proposed video enhancement algorithm (Algorithm 2) using different variations of LK algorithm over different sequences. The minimum scores are highlighted with bold letters

MSE	LK	LK–LM	LK–SSIM–LM
Tehran Park	317.87	317.47	**316.16**
LSMS Opening	99.92	99.93	**99.33**
Tokyo	205.96	205.03	**200.76**
Shanghai Garden	**152.29**	153.2	155.76

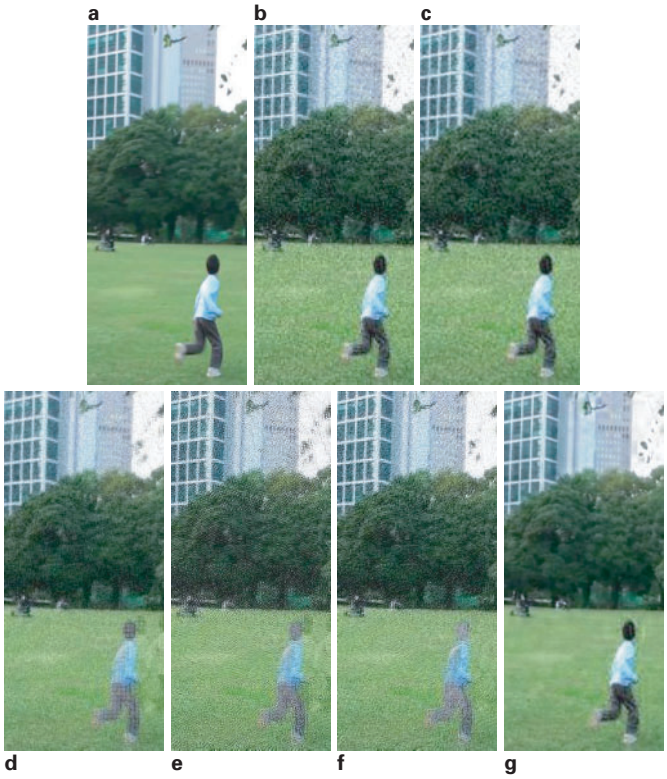
3.1.3 *Comparing different registration methods in Algorithm 2*

We ran the proposed video enhancement algorithm (Algorithm 2) using different variations of the LK algorithm on the mentioned video sequences. In line 6 of algorithm 2, we tried the LK algorithm,[6,18] the LK algorithm with the Levenberg–Marquardt optimisation approach[23] (LK–LM) and the proposed registration method in algorithm 1 (LK–SSIM–LM).

In the experiments of the rest of the paper, the SNR value of LR video frames is 20 dB. 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.[2] Finding the homography matrix has been done using the random sample consensus (RANSAC) method.[10] 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.

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

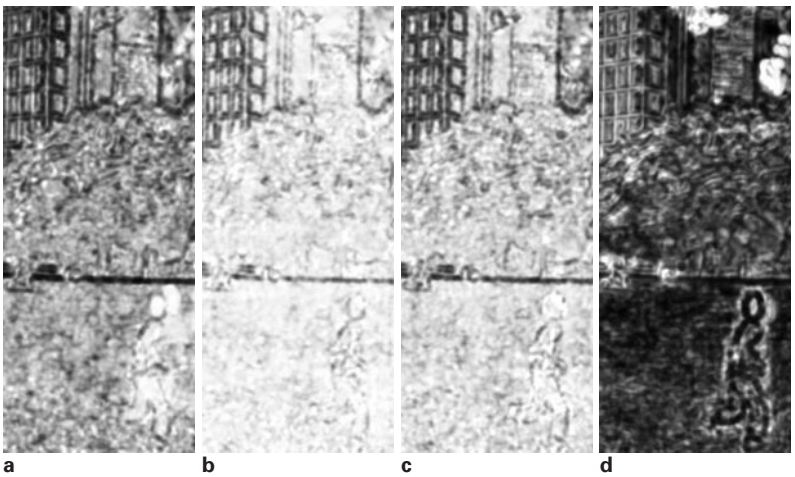
Method	BP	IN	RS	Algorithm 2
MAE ($\times 10^{-3}$)				
Tehran Park	102.46	54.16	82.57	**47.53**
LSMS Opening	61.99	34.90	52.02	**28.48**
Tokyo	133.43	55.94	96.70	**38.46**
Shanghai Garden	84.92	47.46	67.90	**35.08**
PSNR				
Tehran Park	17.58	22.44	19.30	**23.13**
LSMS Opening	22.21	25.99	23.09	**28.15**
Tokyo	15.44	22.10	17.97	**25.14**
Shanghai Garden	19.27	24.39	20.87	**26.22**
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**



6 Close-up views of the original HR image, replication (nearest) and bicubic resizing methods, and super-resolution reconstruction methods: interpolation, iterated back-projection[13] robust super-resolution[31] and the proposed method in Algorithm 2 on '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

The MSE between 60 synthesised frames of 'Tokyo' sequence is shown in Fig. 4. The mean value of each criterion was displayed 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



7 SDIS map image (E_{SDIS}) between Fig. 6d–g with related HR image (Fig. 6a). Brighter pixel means higher error: (a) interpolation; (b) interated back-projection; (c) robust super-resolution; (d) this paper

the first rank in three cases. The 'Shanghai Garden' sequence has a very poor quality and it has not good structures for computing SSIM; hence in this video, the result of LK–SSIM–LM is not better than the other two approaches (other registration approaches also can be used in Algorithm 2. In Appendix 2, some other registration methods are compared with the mentioned method).

3.2 Comparing with different SR methods

We applied our proposed method in Algorithm 2 on aforementioned test sequences and compared its performance with some other SR algorithms. We used the 'sliding-window' techniques with the interpolation (IN), iterated back-projection (BP)[13] and robust super-resolution (RS)[31] as reconstruction methods. Computing the motion parameters between frames has been done using the registration method of Keren *et al.*[14] 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 MAE, PSNR and SSIM for the test sequences. 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 where its HR version is available has been used here for further comparisons. Figure 5 shows MAE, PSNR and SSIM of the proposed video enhancement algorithm (Algorithm 2) and some SR reconstruction methods over 60 frames of 'Tokyo' sequence. The superior performance of the proposed method is obvious.

Figure 6 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 have been shown for comparison purposes. Note that the windows of the rear building in the frame, is almost completely unrecognisable in the LR video frames and in the other SR methods, except that of the proposed method (Fig. 6g). The resolution is clearly enhanced and the mentioned windows are now visible.

Figure 7 shows SDIS map image between Figure 6d–g with the related HR image (Fig. 6a). The brighter pixel value means the more structural difference between the produced image and the reference HR image. As can be seen, the proposed method has the lowest structural error.

4 CONCLUSION

In this paper, a new version of the popular LK image registration algorithm has been proposed and applied to video enhancement. Our goal is the enhancement of low resolution video frames, by fusing a motion compensated form of a high resolution image. The high resolution image is from the same scene of the video but perhaps with a different resolution, different illumination, colour 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, SSIM metric is used as a weighting term of the objective function of LK algorithm. The SSIM criterion exhibited very good consistency with a qualitative visual appearance. The mathematical derivation of the proposed approach using the Levenberg–Marquardt optimisation method, based on the unified framework of Baker *et al.*[6] was given. The accuracy of the proposed registration method is compared with some variations of LK algorithm. The experimental results over video SR 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.

ACKNOWLEDGEMENTS

The authors are indebted to anonymous referees for valuable comments. We would also like to thank Dr Vandewalle Vandewalle[25] for his SR package and Dr D. Lowe for his SIFT key point program (http://www.cs.ubc.ca/lowe/keypoints/). We also thank to Dr Peter Kovesi (http://www.csse.uwa.edu.au/pk/research/matlabfns/), Dr Simon Baker and his co-workers[6] for providing many useful MATLAB functions and Dr Gh. Mohajeri for 'Tokyo' sequence.

REFERENCES

1 Amintoosi, M., Fathy, M. and Mozayani, N. Reconstruction + synthesis: a hybrid method for multi-frame super-resolution, Proc. 5th Iranian Conf. on *Machine vision and image processing: MVIP08*, Tabriz, Iran, November 2008, Tabriz University, pp. 179–184.

2 Amintoosi, M., Fathy, M. and Mozayani, N. Video resolution enhancement in the presence of moving

objects, Proc. Int. Conf. on *Image Processing, Computer Vision, and Pattern Recognition: IPCV 2009*, Las Vegas, NV, USA, July 2009, IEEE Computer Society, pp. 456–460.

3 Amintoosi, M., Fathy, M. and Mozayani, N. Image registration based on structural similarity applied to super resolution, Proc. 17th Iranian Conf. on *Electrical engineering*, Tehran, Iran, May 2009, Iran University of Science and Technology, pp. 401–406 (in Persian).

4 Amintoosi, M., Fathy, M. and Mozayani, N. Seamless image fusion using multiband blending and wavelet transform, Proc. 15th Annual National CSI Computer Conf.: *CSICC 2010*, Tehran, Iran, February 2010, MATN Co. (in Persian). p. 6.

5 Baker, S. and Kanade, T. Hallucinating faces, Proc. 4th IEEE Int. Conf. on *Automatic face and gesture recognition 2000: FG 2000*, Grenoble, France, March 2000, IEEE Computer Society, p. 83.

6 Baker, S., Gross, R. and Matthews, I. Lucas–Kanade 20 years on: a unifying framework. *Int. J. Comput. Vision*, 2004, **56**, 221–255.

7 Borman, S. 'Topics in multiframe superresolution restoration', PhD thesis, University of Notre Dame, Notre Dame, IN, 2004.

8 Brooks, A. What Makes an Image Look Good? [online], March 17 2005. Available at: ⟨http://dailyburrito.com/projects/ImageQuality_Brooks2005.pdf⟩, Presentation for the Image & Video Processing Lab (IVPL) at Northwestern University for ECE 510 Video Processing, November 19th, 2010.

9 Chang, H., Yeung, D.-Y. and Xiong, Y. M. Super-resolution through neighbor embedding, Proc. IEEE Computer Society Conf. on *Computer vision and pattern recognition: CVPR 2004*, Washington, DC, USA, June–July 2004, Vol. 1, pp. 275–282.

10 Fischler, M. A. and Bolles, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 1981, **24**, 381–395.

11 Freeman, W. T., Jones, T. R. and Pasztor, E. C. Example-based super-resolution. *IEEE Comput. Graph. Appl.*, 2002, **22**, 56–65.

12 Hartley, R. I. and Zisserman, A. Multiple View Geometry in Computer Vision, 2004, 2nd edition (Cambridge University Press, Cambridge).

13 Irani, M. and Peleg, S. Improving resolution by image registration. *CVGIP: Graph. Models Image Process.*, 1991, **53**, 231–239.

14 Keren, D., Peleg, S. and Brada, R. Image sequence enhancement using sub-pixel displacement, Proc. IEEE International Conf. on *Computer vision and pattern recognition: CVPR '88*, Ann Arbor, MI, June 1988, IEEE Computer Society, pp. 742–746.

15 Ko, S., Lee, S., Jeon, S. and Kang, E. Fast digital image stabilizer based on gray-coded bit-plane matching. *IEEE Trans. Consum. Electron.*, 1999, **45**, 598–603.

16 Lin, Z. and Shum, H. Y. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 2004, **26**, 83–97.

17 Lowe, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 2004, **60**, 91–110.

18 Lucas, B. D. and Kanade, T. An iterative image registration technique with an application to stereo vision, Proc. 7th Int. Joint Conf. on *Artificial intelligence: IJCAI '81*, Vancouver, BC, Canada, August 1981, IEEE Computer Soceity, pp. 674–679.

19 Lucchese, G. M. and Cortelazzo, L. A noise-robust frequency domain technique for estimating planar roto-translations. *IEEE Trans. Signal Process.*, 2000, **48**, 1769–1786.

20 Marcel, B., Briot, M. and Murrieta, R. Calcul de translatione trotation par la transformation de fourier. *Traite. Signal*, 1997, **14**, 135–149.

21 Pham, T. Q. 'Spatiotonal adaptivity in super-resolution of under-sampled image sequences', PhD thesis, aan de Technische Universiteit Delft, Delft, The Netherlands, 2006.

22 Sheikh, H. R., Wang, Z., Cormack, L. and Bovik, A. C. Live Image Quality Assessment Database Release 2 [online], 2005 (LIVE, Austin, TX). Available at: ⟨http://live.ece.utexas.edu/research/quality⟩, November 19th, 2010.

23 Szeliski, R. S. Video mosaics for virtual environments. *IEEE Comput. Graph. Appl.*, 1996, **16**, 22–30.

24 Talbi, H. and Batouche, M. C. Particle swarm optimization for image registration, Proc. 1st Int. Conf. on *Information and communication technologies: from theory to applications: ICTTA '04*, Damascus, Syria, April 2004, IEEE Computer Society, pp. 397–398.

25 Vandewalle, P., Süsstrunk, S. and Vetterli, M. A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP J. Appl. Signal Process.*, 2006, **2006**, 1–14.

26 Wang, Z., Bovik, A., Sheikh, H. and Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.*, 2004, **13**, 600–612.

27 Wang, Z., Lu, L. G. and Bovik, A. C. Video quality assessment based on structural distortion measurement. *Signal Process.: Image Commun.*, 2004, **19**, 121–132.

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

29 Yuan, Z., Yan, P. and Li, S. Super resolution based on scale invariant feature transform, Proc. Int. Conf. on *Audio, language and image processing*, Shanghai, China, July 2008, ICAD, pp. 1550–1554.

30 Zhao, M. 'Video enhancement using content-adaptive least mean square filters', PhD thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 2006.

31 Zomet, A., Rav-Acha, A. and Peleg, S. Robust super resolution, Proceedings of the Int. Conf. on *Computer vision and patern recognition: CVPR 2001*, Kauai, HI, USA, December 2001, pp 645–650.

APPENDIX 1

A1.1 On the derivation of the proposed algorithm based on $E_{\mathrm{SDIS}}(\mathrm{W(x;p)}$

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

$$\sum_{\mathrm{x}} E_{\mathrm{SDIS}}(\mathbf{W}(\mathbf{x};\mathbf{p}+\Delta\mathbf{p}))\cdot[I(\mathbf{W}(\mathbf{x};\mathbf{p}+\Delta\mathbf{p}))-T(\mathbf{x})]^2 \quad (24)$$

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

$$SSD=\sum_{\mathrm{x}}\left[E_{\mathrm{SDIS}}(\mathbf{W}(\mathbf{x};\mathbf{p}))+\nabla E_{\mathrm{SDIS}}\frac{\partial\mathbf{W}}{\partial\mathbf{p}}\Delta\mathbf{p}\right]\cdot \left[I(\mathbf{W}(\mathbf{x};\mathbf{p}))+\nabla I\frac{\partial\mathbf{W}}{\partial\mathbf{p}}\Delta\mathbf{p}-T(\mathbf{x})\right]^2 \quad (25)$$

It should be mentioned that according to Ref. 26, E_{SDIS} is differentiable. Finding the optimum value of $\Delta\mathbf{p}$ can be done by differentiating equation (25) with respect to $\Delta\mathbf{p}$, setting the result to equal zero and solving it:

$$\frac{\partial SSD}{\partial\Delta\mathbf{p}}=\sum_{\mathrm{x}}\left\{\left(\nabla E_{\mathrm{SDIS}}\frac{\partial\mathbf{W}}{\partial\mathbf{p}}\right)^{\mathrm{T}}\left[I(\mathbf{W}(\mathbf{x};\mathbf{p}))+\nabla I\frac{\partial\mathbf{W}}{\partial\mathbf{p}}\Delta\mathbf{p}-T(\mathbf{x})\right]^2+ 2\left(\nabla I\frac{\partial\mathbf{W}}{\partial\mathbf{p}}\right)^{\mathrm{T}}\left[I(\mathbf{W}(\mathbf{x};\mathbf{p}))+\nabla I\frac{\partial\mathbf{W}}{\partial\mathbf{p}}\Delta\mathbf{p}-T(\mathbf{x})\right]\left[E_{\mathrm{SDIS}}(\mathbf{W}(\mathbf{x};\mathbf{p}))+\nabla E_{\mathrm{SDIS}}\frac{\partial\mathbf{W}}{\partial\mathbf{p}}\Delta\mathbf{p}\right]\right\} \quad (26)$$

For simplicity of driving, we define the following terms:

$$\begin{aligned}
&A=E_{\mathrm{SDIS}}(\mathbf{W}(\mathbf{x};\mathbf{p}))+\nabla E_{\mathrm{SDIS}}\frac{\partial\mathbf{W}}{\partial\mathbf{p}}\Delta\mathbf{p}\\
&B=I(\mathbf{W}(\mathbf{x};\mathbf{p}))+\nabla I\frac{\partial\mathbf{W}}{\partial\mathbf{p}}\Delta\mathbf{p}-T(\mathbf{x})\\
&I=I(\mathbf{W}(\mathbf{x};\mathbf{p}))\\
&T=T(x)\\
&E=E_{\mathrm{SDIS}}(\mathbf{W}(\mathbf{x};\mathbf{p}))\\
&B=I(\mathbf{W}(\mathbf{x};\mathbf{p}))+\nabla I\frac{\partial\mathbf{W}}{\partial\mathbf{p}}\Delta\mathbf{p}-T(\mathbf{x})\\
&I=I(\mathbf{W}(\mathbf{x};\mathbf{p}))\\
&T=T(x)\\
&E=E_{\mathrm{SDIS}}(\mathbf{W}(\mathbf{x};\mathbf{p}))\\
&e=I(\mathbf{W}(\mathbf{x};\mathbf{p}))-T(\mathbf{x})\\
&S_{\mathrm{E}}=\nabla E_{\mathrm{SDIS}}\frac{\partial\mathbf{W}}{\partial\mathbf{p}},\ \text{steepest descent image of E}\\
&S_{\mathrm{I}}=\nabla I\frac{\partial\mathbf{W}}{\partial\mathbf{p}},\ \text{steepest descent image of I}
\end{aligned} \quad (27)$$

Hence, equation (26) can be simplified as follows:

$$\begin{aligned}
\frac{\partial SSD}{\partial\Delta\mathbf{p}}&=\sum_{\mathrm{x}}[S_{\mathrm{E}}^{\mathrm{T}}B^2+2BS_{\mathrm{I}}^{\mathrm{T}}A]\\
&=\sum_{\mathrm{x}}[(S_{\mathrm{E}}^{\mathrm{T}}B+2S_{\mathrm{I}}^{\mathrm{T}}A)B]
\end{aligned} \quad (28)$$

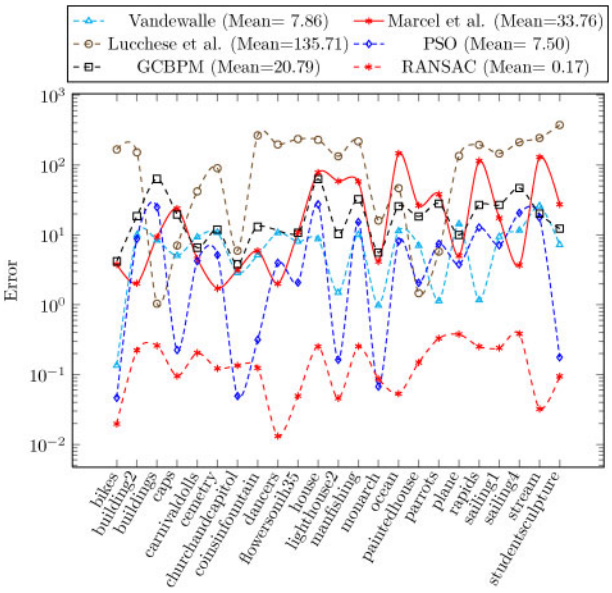
The above factorisation is legal, because the distribution of multiplication over addition is hold for '.' operator (which denotes '.*' operator in MATLAB):

$$X\cdot Z+Y\cdot Z=(X+Y)\cdot Z \quad (29)$$

For simplicity, we temporary drop the summation operator $\sum_{\mathrm{x}}$, from equation (28); $B=0$ or $(S_{\mathrm{E}}^{\mathrm{T}}B+2AS_{\mathrm{I}}^{\mathrm{T}})=0$ are the sufficient conditions such that $\partial SSD/\partial\Delta\mathbf{p}=0$.

If $B=0$ then from our definitions in equation (27) and regarding the summation, we will have:

$$\Delta\mathbf{p}=-\frac{e}{S_{\mathrm{I}}}=\frac{\sum_{\mathrm{x}}[T(\mathbf{x})-I(\mathbf{W}(\mathbf{x};\mathbf{p}))]}{\sum_{\mathrm{x}}[\nabla I(\partial\mathbf{W}/\partial\mathbf{p})]^{\mathrm{T}}} \quad (30)$$



8 Registration error

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:

$$S_{\mathrm{E}}^{\mathrm{T}} B + 2S_{\mathrm{I}}^{\mathrm{T}} A = 0 \Rightarrow$$

$$[\nabla E_{SDIS} \frac{\partial W}{\partial p}]^{\mathrm{T}} \left[I(\mathbf{W}(\mathbf{x};\mathbf{p})) + \nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \Delta\mathbf{p} - T(\mathbf{x}) \right]$$

$$+ 2\left(\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right)^{\mathrm{T}} \left[E_{\mathrm{SDIS}}(\mathbf{W}(\mathbf{x};\mathbf{p})) + \nabla E_{\mathrm{SDIS}} \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \Delta\mathbf{p} \right] = 0 \Rightarrow$$

$$S_{\mathrm{E}}^{\mathrm{T}}(I + S_{\mathrm{I}}\Delta\mathbf{p} - T) + 2S_{\mathrm{I}}^{\mathrm{T}}(E + S_{\mathrm{E}}\Delta\mathbf{p}) = 0 \Rightarrow$$

$$S_{\mathrm{E}}^{\mathrm{T}} I + S_{\mathrm{E}}^{\mathrm{T}} S_{\mathrm{I}} \Delta\mathbf{p} - S_{\mathrm{E}}^{\mathrm{T}} T + 2S_{\mathrm{I}}^{\mathrm{T}} E + 2S_{\mathrm{I}}^{\mathrm{T}} S_{\mathrm{E}} \Delta\mathbf{p} = 0 \Rightarrow$$

$$S_{\mathrm{E}}^{\mathrm{T}} I - S_{\mathrm{E}}^{\mathrm{T}} T + 2S_{\mathrm{I}}^{\mathrm{T}} E + (S_{\mathrm{E}}^{\mathrm{T}} S_{\mathrm{I}} + 2S_{\mathrm{I}}^{\mathrm{T}} S_{\mathrm{E}}) \Delta\mathbf{p} = 0 \Rightarrow$$

$$\Delta\mathbf{p} = -\frac{S_{\mathrm{E}}^{\mathrm{T}}(I - T) + 2S_{\mathrm{I}}^{\mathrm{T}} E}{S_{\mathrm{E}}^{\mathrm{T}} S_{\mathrm{I}} + 2S_{\mathrm{I}}^{\mathrm{T}} S_{\mathrm{E}}} = -\frac{S_{\mathrm{E}}^{\mathrm{T}} e + 2S_{\mathrm{I}}^{\mathrm{T}} E}{S_{\mathrm{E}}^{\mathrm{T}} S_{\mathrm{I}} + 2S_{\mathrm{I}}^{\mathrm{T}} S_{\mathrm{E}}} \qquad (31)$$

Based on the definitions in equation (27) and regarding the summation, we will have:

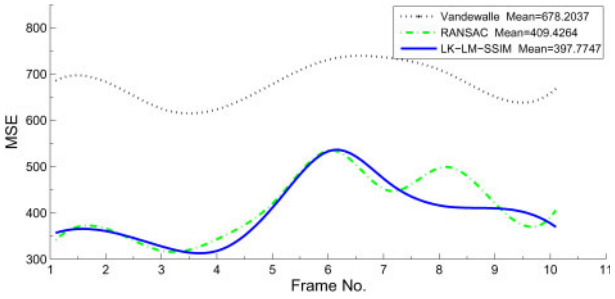
$$\Delta\mathbf{p} = -H^{-1} \sum_{\mathrm{x}} \qquad (32)$$

$$\left\{ \left[\nabla E_{\mathrm{SDIS}} \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right]^{\mathrm{T}} [I(\mathbf{W}(\mathbf{x};\mathbf{p})) - T(\mathbf{x})] + 2\left[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right]^{\mathrm{T}} E_{\mathrm{SDIS}} \right\}$$

where H is:

$$H = \sum_{x} ([\nabla E_{SDIS} \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^{T}$$

$$[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}] + 2[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]^{T} [\nabla E_{SDIS} \frac{\partial \mathbf{W}}{\partial \mathbf{p}}]) \qquad (33)$$



9 Reconstruction error, when some other registration methods are used in Algorithm 2, instead of 'LK–LM–SSIM'

But our implementation based on equation (33) 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 equation (32). This makes the computations to be wrong and even affects the singularity of H in some examples.

APPENDIX 2

A2.1 Comparing some other registration methods

In section 3.1.3, the result of the proposed video enhancement algorithm (Algorithm 2) was compared when different variations of the LK algorithm are used as the registration method. The results achieved showed the benefits of the proposed method; the goal of the paper is not comparing the registration methods but it may be interesting to have a comparison with other registration algorithms. There are many registration methods with their own pros and cons. Figure 8 demonstrates our comparison between some of the registration methods on some images of LIVE dataset LIVE dataset.[22]

The methods compared here are:

- the frequency method of Vandewalle[25]
- the method proposed by Marcel *et al.*[20]
- the method of Lucchese and Cortelazzo[19]
- The registration method based on particle swarm optimisation (PSO)[24]
- grey-coded bit-plane matching[15]
- outlier detection and parameter estimation using RANSAC,[10,12,29] in which used in many papers.

In Fig. 8, for each image, a distorted version of that image, under Eucledian transformation with known parameters and with SNR = 70 dB, is produced for evaluation. Since the motion parameters is known, the registration error can be computed. Here the

pixel location error is reported as the registration error.

As can be seen in Fig. 8, the RANSAC, PSO and Vandewalle methods produced lower errors, in average. The PSO method is a very time consuming approah; hence, only the RANSAC and the Vandewalle methods selected here for further comparison with the proposed approach 'LK–LM–SSIM' when these methods are used in Algorithm 2. The result is shown in Fig. 9 for some frames of 'Tehran Park' sequence. As can be seen, the proposed approach has the lowest error.