Accurate Image Registration for MAP Image Super-Resolution

Michalis Vrigkas, Christophoros Nikou and Lisimachos P. Kondi

Department of Computer Science, University of Ioannina, Greece
{mvrigkas, cnikou, lkon}@cs.uoi.gr

Abstract

The accuracy of image registration plays a dominant role in image super-resolution methods and in the related literature, landmark-based registration methods have gained increasing acceptance in this framework. In this work, we take advantage of a maximum *a posteriori* (MAP) scheme for image super-resolution in conjunction with the maximization of mutual information to improve image registration for super-resolution imaging. Local as well as global motion in the low-resolution images is considered. The overall scheme consists of two steps. At first, the low-resolution images are registered by establishing correspondences between image features. The second step is to fine-tune the registration parameters along with the high-resolution image estimation, using the maximization of mutual information criterion. Quantitative and qualitative results are reported indicating the effectiveness of the proposed scheme, which is evaluated with different image features and MAP image super-resolution computation methods.

Keywords:
Maximum *a posteriori* (MAP) super-resolution, image registration, mutual information, feature extraction, Harris corners, Scale Invariant Feature Transform (SIFT), Speed Up Robust Features (SURF).

1. Introduction

Image super-resolution (SR) reconstruction has gained lots of prominence in the last two decades. Many applications, ranging from medical imaging to image recognition and video applications, are driving the need for better

reconstruction techniques to enhance image resolution. The objective of image super-resolution is to reconstruct a high-resolution (HR) image from a sequence of low-resolution (LR) images. The SR methods aim to improve the spatial resolution by fusing the set of LR images to produce an image with more visible detail in the high spatial frequency features. The LR images experience different degradations such as motion, point spread function blurring, subsampling and additive noise. The HR image is estimated from a sequence of LR aliased images, which is possible if there exists sub-pixel motion between the LR images. Thus, each frame of the LR sequence brings complementary information on the original HR image.

Super-resolution reconstruction is achieved in three main steps: (i) registration, (ii) interpolation and (iii) restoration. Registration is the process of estimating an image transformation model derived directly from the LR data set. In the interpolation step, the LR images are superimposed onto the HR image grid, while restoration removes noise and blur that is present in the LR images. The direct inverse solution from interpolation, motion compensation and inverse filtering is ill-posed due to the existence of additive noise, even in cases of perfect motion registration and accurate knowledge of the point spread function of the acquisition system. Since the seminal work of Tsai and Huang [33], many methods have been proposed to seek a stable solution with high visual quality to overcome the ill-posed nature of the problem. Among them, methods based on the Fourier transform [31, 35] and projections onto convex sets [13] have gained popularity.

Close attention has been paid to stochastic methods, which impose a prior distribution on the image to be reconstructed. In this context, a basic maximum *a posteriori* (MAP) multi-frame SR framework exists, where the posterior distribution of the HR image is maximized [16, 18, 19, 26, 9, 11, 14, 38]. Bayesian approaches are also very popular. They seek to find a solution where all the unknown parameters, such as registration parameters, have the maximum probability [26, 25, 32, 20, 28, 14]. A number of studies have been applied in reconstruction from only a single LR frame [39]. These methods called *quasi-super-resolution* methods.

A key issue in the quality of the super-resolved image is the accuracy of the employed image registration technique. Also, knowledge of the involved motion model facilitates the task. This may include simple translational, rigid body or affine motion as well as projective or even photometric transformations. The standard approach is to estimate the registration parameters separately from the HR image [9, 11], either by aligning the LR images once,

at the beginning of the algorithm or iteratively before or after each update of the HR image [16, 18, 19]. The method of Farsiu *et al.* [14] focuses in this direction. The use of L_1 norm and a robust regularization term achieves high accuracy and results in images with sharp edges. Also, there exist techniques where the registration parameters are assumed to be random variables and they are marginalized in a Bayesian formulation [26, 25]. Apart from using block matching or phase correlation techniques, the majority of the registration methods used in the SR literature are related to standard optical flow methods and their variants [23, 5, 15]. In the same context, the study of Zhou *et al.* [42] try to estimate the registration parameters between a reference and a sensed image using a limited number of control points. Their application to image super resolution shows the potential of the method to correctly estimate the registration parameters under several affine deformations. In the sense of feature extraction techniques Baboulaz and Dragotti [4] developed a method for accurate registration of LR images.

Super-resolution lies at the heart of many aspects of image analysis theory and it therefore requires the understanding of several fields. In this framework, image fusion has efficiently been used to extract relevant information between LR images in a MAP-based scheme [38], where the unknown misregistrations can easily be handled.

Following the trends in computer vision, feature matching has also been used [9]. The parameters of the geometric transformation between the LR images are estimated by automatic detection and analysis of corresponding features among the input images. Typically, some hundreds of points of interest, such as the Harris corner features [17], are detected with subpixel accuracy and correspondences are established by examining the image neighborhoods around them.

Several image registration methods have been applied to image super-resolution reconstruction algorithms. Methods relying on image features such as Harris and SIFT [9] do not provide subpixel accuracy. Methods based on block matching [16] and optical flow [8] are generally time consuming. Algorithms using automatically computed segmentation maps [10] and tracking algorithms [13, 6], which have also been applied are both slow and prone to localization errors. All these landmark-based registrations are limited to least-squares based solutions. Mutual information (MI) [24, 36] is a method that was originally proposed for medical image registration. It is widely used in many domains, achieves sub-pixel registration accuracy and has never been employed in super-resolution reconstruction. Maximization

of MI is a very general and powerful criterion, because no assumptions are made regarding the nature of the statistical dependence between the two images and no limiting constraints are imposed on the image content of the involved degradations. In the last 15 years, the maximization of the mutual information has revolutionized image registration theory and applications as it considers the whole gray level image information and consistently provides sub-pixel precision. If mutual information is not initialized close to the global maximum, local extrema impede the registration process [27] and consequently, they rule out sub-pixel accuracy. To our knowledge, mutual information has not been applied as a registration method to the problem of multiple image super-resolution. A work that involves mutual information and SR was proposed by Zhang *et al.* [41]. However, the goal of the algorithm was to enhance the quality of a single image by generating multiple LR images from the same single image. Also, in [12], mutual information was employed as a regularization term in Bayesian image restoration.

In this paper, we propose to register the LR images by building correspondences between the LR frames followed by a gentle step of fine-tuning in synergy with the HR image estimation, by maximization of the mutual information criterion [24, 36] between the estimation of the HR image and each upsampled LR image. A four page summary of this work was presented in [37]. Herein, we present more experimental results comparing different feature-based registration methods and combining them with the maximization of mutual information criterion. Corresponding robust features are obtained in three different manners, using Harris corners [17], SIFT [22] and SURF descriptors [7]. Also, four MAP image super-resolution algorithms are put to test [16, 18, 14, 38], in order to demonstrate the effectiveness of the approach.

The main contribution of this paper is the development of a registration approach based on mutual information for SR reconstruction. Numerical results demonstrate that the reconstructed HR images are of higher quality with respect to standard MAP-based SR approaches not employing the mutual information criterion in the registration step.

2. Method

Given a collection of blurred LR images, which differ by a rigid transformation (rotation and translation) and are corrupted by white Gaussian noise, the goal is to automatically estimate a high resolution image. A flowchart

of the proposed method is shown in Figure 1. A feature extraction algorithm is firstly applied, which is followed by a least squares estimation of the rigid transformation parameters [34] based on the correspondences between features. This is, in general, the standard approach for any MAP-based super-resolution algorithm which then estimates the high resolution image.

However, feature based registration is prone to detection errors and errors in establishing correspondences between features which makes the result optimal only in the least squares sense. The least squares solution is relatively stable and close to the optimal minimizer but in most cases it needs further improvement. In order to overcome this limitation, we propose to employ the feature correspondence as the initialization of a registration algorithm relying on the maximization of the mutual information criterion [24, 36] which is a very powerful tool for image registration. Let us notice, that, the mutual information alone is very sensitive to local minima and needs a good initialization close to the optimal solution. This is true even if the registration problem is set in noise free conditions. Therefore, it cannot be employed in SR problems as the Gaussian noise may lead to large misregistrations. The synergy of least squares feature-based registration and mutual information is a proposed for increasing the quality of the super-resolved image.

2.1. Iterative MAP Image Super-Resolution

The image degradation process [18] is modeled by motion (rotation and translation), a linear blur, and subsampling by pixel averaging along with additive Gaussian noise. We assume that p LR images, each of size $M = N_1 \times N_2$, are obtained from the acquisition process. The following observation model is assumed, where all images are ordered lexicographically:

$$\mathbf{y} = \mathbf{W}\mathbf{z} + \mathbf{n}. \tag{1}$$

The set of LR frames is described as $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_p^T]^T$, where $\mathbf{y}_k$, for $k = 1, ...p$, are the p LR images. The desired HR image $\mathbf{z}$ is of size $N = l_1 N_1 \times l_2 N_2$, where l_1 and l_2 represent the up-sampling factors in the horizontal and vertical directions, respectively. The term $\mathbf{n}$ represents zero-mean additive Gaussian noise. In (1), the degradation matrix $\mathbf{W} = [\mathbf{W}_1^T, \mathbf{W}_2^T, \ldots, \mathbf{W}_p^T]^T$ performs the operations of motion, blur and subsampling. Thus, matrix $\mathbf{W}_k$, for the k-th frame, may be written as

$$\mathbf{W}_k = \mathbf{D}\mathbf{B}_k\mathbf{M}(\mathbf{s}_k), \tag{2}$$

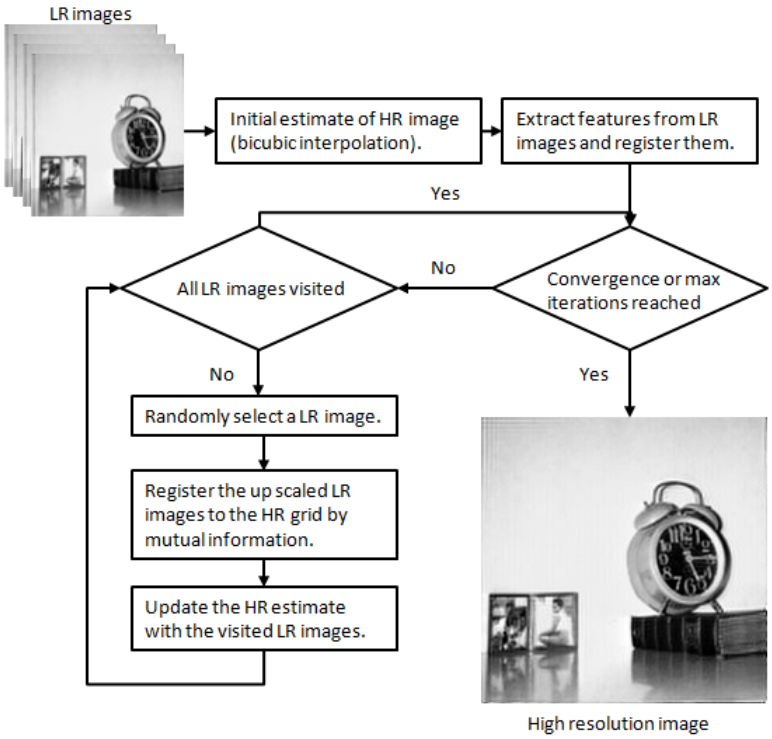


Figure 1: Flowchart of the proposed method. First, we perform one step of registration based on feature extraction and then iteratively, we register with mutual information and update the HR estimation.

where $\mathbf{D}$ is the $N_1N_2 \times N$ subsampling matrix, $\mathbf{B}_k$ is the $N \times N$ blurring matrix. The transformation model represented by matrix $\mathbf{M}(\mathbf{s}_k)$ is described by:

$$\mathbf{z}_i^{'} = \mathbf{R}\mathbf{z}_i + \mathbf{T} + d_i,$$

where $\mathbf{z}_i$ is the i-th pixel of the high resolution image and $\mathbf{z}_i^{'}$ is the corresponding transformed pixel. $\mathbf{R}$ and $\mathbf{T}$ denote the global rotation and translation parameters and d_i is the local translational motion vector of the i-th pixel. Thus, matrix $\mathbf{M}(\mathbf{s}_k)$ now implies global rotation, global translation and local shift and $\mathbf{s}_k = \mathbf{R}, \mathbf{T}, d_i$ for the k-th LR image.

Formulating the super-resolution problem in a probabilistic framework

[16], a smooth Gaussian prior is generally considered for the HR image:

$$p(\mathbf{z}) = \frac{(\alpha|\mathbf{Q}^T\mathbf{Q}|)^{N/2}}{(2\pi)^{N/2}} \prod_{i=1}^{N} \exp\left(-\frac{1}{2}\alpha(\mathbf{Q}\mathbf{z})^T(\mathbf{Q}\mathbf{z})\right), \tag{3}$$

where $\mathbf{Q}\mathbf{z}$ is the Laplacian of the HR image $\mathbf{z}$ and parameter α controls the precision (inverse covariance) and consequently the shape of the distribution. The above zero-mean normal distribution assigns a high probability to images not exhibiting rich edge information. The simplest approach is to consider parameter α spatially constant, yielding a stationary model for the whole image. This implies that the statistics for $\mathbf{Q}\mathbf{z}$ are Gaussian, independent and identically distributed. Small values of α indicate the presence of a large variation. By this means, this prior may maintain edges and suppress noise in smooth areas of the image. Given the HR image $\mathbf{z}$ and the registration parameters between the LR images $\mathbf{s} = \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_k\}$, the likelihood of the LR images is also a Gaussian [18]:

$$p(\mathbf{y}|\mathbf{z}) = \frac{1}{(2\pi)^{\frac{pM}{2}}\sigma_\eta^{pM}} \exp\left(-\frac{(\mathbf{y}-\mathbf{W}\mathbf{z})^T(\mathbf{y}-\mathbf{W}\mathbf{z})}{2\sigma_\eta^2}\right), \tag{4}$$

where σ_η^2 is the variance of the observation noise $\mathbf{n}$.

Employing a MAP approach and maximizing $p(\mathbf{z}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{z})p(\mathbf{z})$ leads to the following MAP functional to be minimized with respect to the HR image $\mathbf{z}$ and the rigid transformation parameters $\mathbf{s}$:

$$L(\mathbf{z}, \mathbf{s}) = \sum_{k=1}^{p} \|\mathbf{y}_k - \mathbf{W}_k(\mathbf{s}_k)\mathbf{z}\|^2 + \frac{\sigma_\eta^2}{\lambda}\|\mathbf{Q}\mathbf{z}\|^2. \tag{5}$$

Notice the change in notation to explicitly underpin the dependence of matrix $\mathbf{W}_k$ on the registration parameters $\mathbf{s}_k$.

Using a gradient descent method with a properly calculated step size it can be shown that the update equation minimizing (5) can be written as

$$\hat{\mathbf{z}}^{n+1} = \hat{\mathbf{z}}^n - \varepsilon^n \nabla_{\mathbf{z}} L(\mathbf{z}, \mathbf{s})|_{\mathbf{z}=\hat{\mathbf{z}}^n, \mathbf{s}=\hat{\mathbf{s}}^n}. \tag{6}$$

Parameter ε^n is the step size at the n-th iteration which may be obtained in closed form from the data [16]. In general, the estimation of the regularization parameter λ, which depends on the noise standard deviation σ_η, and the

parameter α controlling the variance in the prior (3), is a difficult task. Generally, the parameters σ_η and λ are empirically selected [16].

In the same spirit, He and Kondi [18] perform a refinement step in order to avoid a blurred version of the high-resolution image. The regularization parameter λ and the noise standard deviation parameter σ_η are automatically computed from the data in an iterative scheme. The step size parameter ε^n is also computed in a closed form at each iteration.

Moreover, Farsiu *et al.* [14] proposed an efficient MAP estimation method to fuse a sequence of LR images. This technique can also cope with color images and reduce any color artifacts. The proposed cost function is also based on Eq. (5) but introduces three more penalty terms. The first penalty term refers to the spatial luminance, as it is important that the edges in the estimated HR image remain sharp. The second one, refers to the regularization of the color effects and finally, the third term penalizes the edge location and orientation across the different color bands. It is worth noticing that this method determines the value of step size ε^n heuristically.

Following a more general model for the prior distribution $p(\mathbf{z})$, Šroubek and Flusser [38] employ a Markov random field with a Gibbs distribution $p(\mathbf{z}) \propto \exp(-F(\mathbf{z})/C)$, where C is a constant and F is an energy function. To overcome the problem of no prior knowledge on the blurring functions, a Markov random field is also employed for the shape of the prior distribution of the degradation matrix $p(\mathbf{W})$. Employing a MAP approach maximizing $p(\mathbf{z}, \mathbf{W}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{z}, \mathbf{W})p(\mathbf{z})p(\mathbf{W})$ leads to a minimization problem with respect to the HR image $\mathbf{z}$ and the unknown blurs $\mathbf{W}$. Consequently, the updating of the HR image comprises two steps, the estimate of the HR image itself and the estimate of the blur.

2.2. Image Registration for Super-Resolution

2.2.1. Feature Extraction

A standard approach in MAP super-resolution algorithms is to register the LR images prior to the computation of the HR image. This is performed once and the registration parameters are fixed during the iterative estimation of the super-resolved image. In computer vision registration problems, it is common to estimate geometric transformations by computing corresponding features between LR images [9].

The Harris corner detector is a very useful technique for finding point-to-point correspondences among the LR images [17]. The basic idea behind the Harris corner detector is that it finds a point where two edges meet,

which also means that this point is in an area with high gradient in two directions. Although the extracted features are robust due to their invariance to rotation, affine intensity change and image noise, they often require expert supervision to prevent from registration errors as the registration parameters are computed in the least squares sense.

Another technique for finding interest points in the input LR images is the Scale Invariant Feature Transform (SIFT) [22]. SIFT is a descriptor of length 128, computed in four steps: (i) scale-space extrema detection, (ii) keypoint localization (iii) orientation assignment and (iv) keypoint descriptor generation. SIFT features are generally more robust than corner features. Features extracted from LR image are detectable even under changes in rotation, scale, noise, illumination and viewpoint [22].

Finally, the Speed Up Robust Features (SURF) is a SIFT-like scale, rotation and noise invariant keypoint detector and descriptor [7]. SURF features can be computed much faster than SIFT. They are based on integral images for fast image convolution and they use the sum of Haar wavelet responses around the point of interest.

2.2.2. Mutual Information Registration

The maximization of mutual information, originally proposed for medical image registration, is considered to be one of the most accurate methods for image registration [24, 36] as it provides sub-pixel accuracy. It relies on gray level information by considering each image pixel as a random variable.

Let $\mathbf{y}_k$ with $k = 1, 2, \ldots, p$ and $\hat{\mathbf{z}}^n$ be the two images with marginal probability density functions (computed from their histograms) $p_{\mathbf{y}_k}(i)$ and $p_{\hat{\mathbf{z}}^n}(j)$ respectively. Let also their joint density be $p_{\mathbf{y}_k\hat{\mathbf{z}}^n}(i,j)$. The mutual information between $\mathbf{y}_k$ and $\hat{\mathbf{z}}^n$ measures the degree of dependence between them and it is defined by

$$\begin{aligned} I(\mathbf{y}_k, \hat{\mathbf{z}}^n) &= H(\mathbf{y}_k) + H(\hat{\mathbf{z}}^n) - H(\mathbf{y}_k, \hat{\mathbf{z}}^n) I(\mathbf{y}_k, \hat{\mathbf{z}}^n) \\ &= \sum_i \sum_j p_{\mathbf{y}_k\hat{\mathbf{z}}^n}(i,j) \log \frac{p_{\mathbf{y}_k\hat{\mathbf{z}}^n}(i,j)}{p_{\mathbf{y}_k}(i) \cdot p_{\hat{\mathbf{z}}^n}(j)}, \end{aligned} \tag{7}$$

where $H(\mathbf{y}_k)$ and $H(\hat{\mathbf{z}}^n)$ are the marginal entropies of the random variables $\mathbf{y}_k$ and $\hat{\mathbf{z}}^n$ and $H(\mathbf{y}_k, \hat{\mathbf{z}}^n)$ is their joint entropy. If the images are correctly registered, their mutual information is maximized.

In order to provide invariance to the overlapping areas between the two images, a more robust measure is the normalized mutual information (*NMI*)

[30]:

$$NMI(\mathbf{y}_k, \hat{\mathbf{z}}^n) = \frac{H(\mathbf{y}_k) + H(\hat{\mathbf{z}}^n)}{H(\mathbf{y}_k, \hat{\mathbf{z}}^n)}. \tag{8}$$

A drawback of the mutual information (and *NMI*) is that, if it is not initialized close to the optimal solution it is trapped by local maxima [27]. To overcome this issue, a good initialization is important.

2.3. The Overall Algorithm

In this framework, we propose to estimate the registration parameters in two steps. In the first step, the registration procedure is initialized by a landmark-based registration scheme. To this end, to register the LR images, we employ features such as the ones described in the previous section. Considering a LR image as the reference, the rigid transformation parameters (translation and rotation) are estimated through minimization of the mean square error between the locations of the features between the reference image and each LR image [34]. Thus, we obtain a good initialization for the unknown registration parameters.

In the next step, during the iterative update of the HR image, a fine tuning of the registration parameters is accomplished by the maximization of the mutual information between the current estimate of the HR image and each upscaled LR image. Upscaling is performed by deblurring (inverse filtering) and upsampling. As the estimate of the HR image changes at each iteration, the registration parameters are updated based on this estimate. By these means, the registration accuracy is improved at each iteration step. The overall algorithm is summarized in Algorithm 1.

The proposed algorithm requires as input the LR image sequence and the up-sampling factors l_1 and l_2. First, we extract features from the LR images and establish the correspondences between them. The correspondences are obtained by searching for points that are maximally correlated with each other within a window of 3×3 size surrounding each keypoint. Then, the corresponding features are used to estimate the transformation parameters using least squares [34]. Notice that the first estimate of the HR image is taken by random selection of a LR image which is then upscaled to the HR grid by bicubic interpolation. The random selection also holds for the subsequent LR frames, which are upscaled and registered to the current LR estimate. At each internal iteration, once a LR frame is registered to the current HR image, the LR image is updated using (6). The procedure is

Algorithm 1 Feature extraction based super-resolution image reconstruction algorithm.

Input: Low-Resolution images $\mathbf{y}_k$, $k = 1, 2, \cdots, p$ and l_1, l_2 up-sampling factors.

Output: High-Resolution image estimate $\hat{\mathbf{z}}$.

- Extract features from the LR images and establish correspondences.
- Estimate rotations and translations using least squares [34].
- Estimate local motion by applying block matching algorithm.
- First estimate of the HR image $\hat{\mathbf{z}}^0$ using (6).
- $n := 1$;
- **do**
  - **do**
    - Random selection of a LR image $\mathbf{y}_k$.
    - **if** $\mathbf{y}_k$ is visited.
      - Register by mutual information, given in (8), the upscaled $\mathbf{y}_k$ to $\hat{\mathbf{z}}^n$.
      - Estimate local motion by applying block matching algorithm.
      - Update $\hat{\mathbf{z}}^n$ using (6) only for the the visited $\mathbf{y}_k$.
    - **end**
    - Declare $\mathbf{y}_k$ visited.
  - **until all** $\mathbf{y}_k$ are visited.
  - $n := n + 1$;
  - Declare all $\mathbf{y}_k$, $k = 1, ...p$ unvisited.
- **until** $\|\hat{\mathbf{z}}^{n+1} - \hat{\mathbf{z}}^n\|/\|\hat{\mathbf{z}}^n\| < \epsilon$ or a predefined number of iterations is reached.

repeated until a convergence criterion is satisfied or a maximum predefined number of iterations is reached.

The method is iterative and consists of two steps. A first step is the registration by mutual information which converges to the optimal solution provided that the algorithm is initialized close to the global optimum [24, 36]. This is generally the case in super-resolution problems where the misalignment is below one degree in rotation and one pixel in translation. Moreover, the problems examined here contain larger misalignments which are resolved by the registration of the images using keypoints. The second step concerns a family of MAP super-resolution reconstruction methods which converge according to their iterative schemes [16, 18, 14, 38]. The convergence of these iterative algorithms is guaranteed by the *contraction mapping theorem* [21]. According to this theorem, the iterative model (6) converges to a unique solution $\hat{\mathbf{z}}$. Therefore, our algorithm, relying on the combination of the above schemes always converges.

3. Experimental Results

In this work, we sought to establish a methodology for efficiently registering LR images in the context of SR reconstruction problem. Since image alignment is critical to SR reconstruction, the effect of registration error was investigated experimentally. In order to evaluate the proposed methodology, several set of experiments were carried out using four state-of-the-art MAP based SR algorithms [16], [18], [14] and [38]. The majority of the images used in these experiments are from the USC-SIPI image database [1]. Sequences of five LR images were created by rotating, translating, blurring, downsampling and degrading by noise an original image. Translation parameters were randomly drawn from a uniform distribution in $[-3, 3]$ (in units of HR pixels) and rotation angles were also uniformly selected in $[-5, 5]$ (in degrees). The images were then downsampled by a factor of two (four pixels to one). Then, a point spread function of a 5×5 Gaussian kernel with standard deviation of 1 pixel was applied. Finally, the resulting images were degraded by white Gaussian noise in order to obtain signal to noise ratios of (i) 25 dB, (ii) 30 dB and (iii) 35 dB. Additionally, in order to obtain a first estimate of the HR image, a LR image was chosen and it was upscaled by bicubic interpolation in all the experiments. In order to evaluate the algorithm over the local translational model, further experiments on an artificially generated dataset have also been conducted. The dataset consists of three shapes, one triangle

and two stars. The stars remain still whereas the triangle moves along the horizontal and vertical axes with respect to the stars.

A quantitative evaluation of the obtained HR images is given by the peak signal to noise ratio (PSNR):

$$\mathrm{PSNR} = 10 \log_{10} \frac{255^2}{\|\hat{\mathbf{z}} - \mathbf{z}\|^2}, \tag{9}$$

where $\mathbf{z}$ and $\hat{\mathbf{z}}$ denote the ground truth and the estimated HR image, respectively.

The improvement signal to noise ratio (ISNR) was also used, which is defined as:

$$\mathrm{ISNR} = 10 \log_{10} \frac{\|\hat{\mathbf{z}}_{\mathrm{ref}} - \mathbf{z}\|^2}{\|\hat{\mathbf{z}} - \mathbf{z}\|^2}, \tag{10}$$

where $\hat{\mathbf{z}}_{\mathrm{ref}}$ denotes a reference HR image.

The structural similarity measure index (SSIM) [40] is a metric that represents a visual distortion between a reference image and the observe LR image. The SSIM is regarded as a function between two images $\mathbf{z}$ and $\hat{\mathbf{z}}$ and it is expressed as:

$$\mathrm{SSIM}(\mathbf{z}, \hat{\mathbf{z}}) = \frac{(2\mu_{\mathbf{z}}\mu_{\hat{\mathbf{z}}} + C_1)(2\sigma_{\mathbf{z}\hat{\mathbf{z}}} + C_2)}{(\mu_{\mathbf{z}}^2 + \mu_{\hat{\mathbf{z}}}^2 + C_1)(\sigma_{\mathbf{z}}^2 + \sigma_{\hat{\mathbf{z}}}^2 + C_2)}, \tag{11}$$

where $\mu_{\mathbf{z}}$ and $\mu_{\hat{\mathbf{z}}}$ denotes the mean intensity of the ground truth and the estimated HR image, respectively. $\sigma_{\mathbf{z}}$ and $\sigma_{\hat{\mathbf{z}}}$ are the standard deviations of the two images and C_1 and C_2 are constants added to avoid instability.

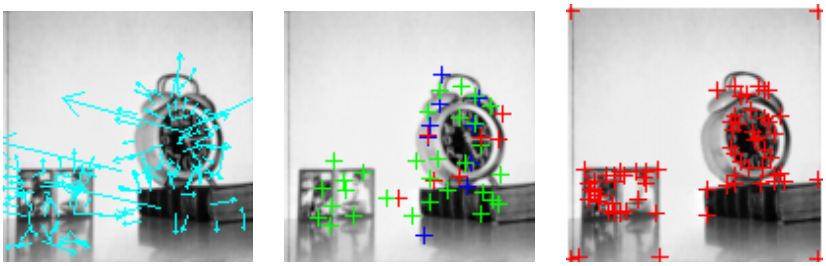
Finally, we have further used the visual information fidelity measure (VIF) [29] in order to assess the quality of the estimated HR image. The construction of VIF relies on successfully modeling the image distortion. It is a measure of statistical modeling described as:

$$\mathrm{VIF} = \frac{\sum_{j \in \mathrm{subbands}} I(\overrightarrow{C}^j; \overrightarrow{\hat{\mathbf{z}}}^j | s^j)}{\sum_{j \in \mathrm{subbands}} I(\overrightarrow{C}^j; \overrightarrow{\mathbf{z}}^j | s^j)}, \tag{12}$$

where we sum over the subbands of interest. The $I(\overrightarrow{C}^j; \overrightarrow{\hat{\mathbf{z}}}^j | s^j)$ and $I(\overrightarrow{C}^j; \overrightarrow{\mathbf{z}}^j | s^j)$ are the corresponding mutual information measures for the j-th subband. $\overrightarrow{C}$ is a collection of N realizations of a random coefficient vector field from a subband in the reference image $\hat{\mathbf{z}}$ and $s = \{s_1, s_2 \ldots, s_N\}$ denotes a realization of the particular reference image. Notice that both SSIM and VIF

range between zero and one. The highest the measure value, the better the reconstructed HR image is.

At first, in order to evaluate the proposed registration method, we performed a number of experiments in a number of registration problems. Registration errors were computed in terms of pixels and not in terms of transformation parameters. Registration accuracies in terms of rotation angles and translation vectors are not easily evaluated due to parameter coupling. Therefore, the registration errors are defined as deviations of the corners of the registered image with respect to the ground truth position. Let us notice that these registration errors are less forgiving at the corners of the image (where their values are larger) with regard to the center of the image frame.



(a) (b) (c)

Figure 2: Representative extracted features of the LR *clock* sequence. (a) SIFT, (b) SURF and (c) Harris corners.

Figure 2 shows representative extracted features in an observed LR image using SIFT (129 keypoints), SURF (50 keypoints) and Harris corners (66 keypoints), respectively.

Hence, we examined the dependence of the registration quality on the registration methods. The experiments for all images in the data set were realized 15 times using different transformation parameters and noise realizations. In the first experiment, we compared the performances of different feature-based registration methods, namely SIFT, SURF and Harris corner with and without the mutual information criterion. Notice that the mutual information is initialized by the transformation estimated by the corresponding feature-based registration method. Table 1 summarizes the statistics on the registration errors. It is worth noticing that SIFT and SURF descriptors produce the smallest registration errors when combined with the mutual information. It may also be noticed that Harris corners perform worse than

Table 1: Statistics for the registration errors in pixels for the *clock* sequence.

Registration errors					
Methods	mean	std	median	min	max
SIFT	0.102	0.036	0.115	0.013	0.169
SIFT & MI	0.049	0.024	0.045	0.014	0.089
SURF	0.121	0.077	0.111	0.020	0.317
SURF & MI	0.064	0.039	0.057	0.014	0.089
Harris	0.142	0.063	0.134	0.035	0.252
Harris & MI	0.101	0.051	0.112	0.019	0.191

SIFT and SURF. In the general case, Harris corners are less accurate in registering the LR images, which leads to bad initialization for the mutual information criterion and affects the registration accuracy.

A next experiment consists in applying a feature-based registration scheme followed by the maximization of the mutual information and estimate SR images using the methods of Hardie *et al.* [16], He and Kondi [18], Farsiu *et al.* [14] and Šroubek and Flusser [38]. For the implementation of the last two methods we used the code provided by the authors [2, 3]. The numerical results are summarized in Table 2 showing the PSNR, and Table 3 showing the ISNR for 35 dB, 30 dB and 25 dB degradation noise, where the mean values, the standard deviations and the median values of the PSNR and ISNR for the *clock* sequence are presented. Tables 4 and 5 preset the SSIM and VIF values for the *clock* sequence for the same experiments. These values are obtained through 15 random realizations of the experiment using different transformation parameters and noise realizations.

In Table 2 it may be seen that the combination of feature-based initialization of the registration parameters followed by fine tuning by the maximization of the mutual information criterion provides consistently higher accuracy. The PSNR values in bold indicate the best quality reconstructed images with respect to the registration method (along columns). In terms of PSNR, the method of He and Kondi [18] achieves better reconstruction results in most cases. The use of mutual information consistently improves the results in all of the methods.

The results in Table 3 show the ISNR statistics for the compared SR methods. Registration using SURF was taken to be the reference method ($\hat{\mathbf{z}}_{\text{ref}}$ in Eq. (10)) for computing the ISNR value. The ISNR values in bold

Table 2: PSNR statistics (in dB) for the compared super-resolution methods for the *clock* sequence.

PSNR	Registration Method	Hardie *et al.* [16]			He and Kondi [18]			Farsiu *et al.* [14]			Šroubek and Flusser [38]		
		mean	std	median	mean	std	median	mean	std	median	mean	std	median
35 dB Gaussian noise	SIFT	24.26	0.64	24.29	24.93	0.47	24.91	24.50	0.73	24.57	23.26	0.32	23.24
	SIFT & MI	25.34	0.57	24.72	25.80	0.10	25.61	**25.06**	0.34	24.37	24.11	0.19	24.59
	SURF	24.31	0.62	24.13	24.88	0.76	25.19	23.19	0.25	22.05	23.02	0.20	22.94
	SURF & MI	**25.40**	0.64	24.19	**25.84**	0.11	25.54	23.58	0.51	23.85	24.07	0.38	24.27
	Harris	23.33	0.28	22.83	23.57	0.18	23.43	22.47	0.66	22.57	22.91	0.84	23.19
	Harris & MI	25.17	0.69	25.31	25.13	0.22	25.07	24.47	0.17	24.50	**24.63**	0.19	24.41
30 dB Gaussian noise	SIFT	22.86	0.46	21.86	22.96	0.61	23.21	22.25	0.51	22.58	22.34	0.18	22.30
	SIFT & MI	**23.60**	0.21	23.32	**24.42**	0.21	24.80	**23.73**	0.29	22.81	22.42	0.27	22.57
	SURF	22.29	0.73	22.17	23.32	0.28	23.89	22.33	0.20	22.13	22.56	0.26	22.94
	SURF & MI	23.47	0.71	23.98	23.07	0.84	23.75	22.56	0.71	22.15	23.13	0.41	23.28
	Harris	21.66	0.32	21.29	21.11	0.26	21.09	22.42	0.70	21.66	22.23	0.31	22.26
	Harris & MI	22.71	0.47	22.29	23.99	0.34	23.38	22.77	0.25	22.65	**23.31**	0.16	23.07
25 dB Gaussian noise	SIFT	22.13	0.39	22.02	22.48	0.26	22.37	22.56	0.21	22.46	22.19	0.47	22.27
	SIFT & MI	**22.65**	0.17	22.86	22.54	0.14	22.03	22.69	0.14	22.87	22.25	0.18	22.25
	SURF	21.76	0.25	21.77	22.34	0.21	22.35	21.30	0.63	21.67	22.26	0.33	22.43
	SURF & MI	22.49	0.43	22.59	**22.76**	0.04	22.79	**22.71**	0.79	21.89	**22.66**	0.44	22.80
	Harris	21.45	0.74	21.51	21.34	0.64	21.21	21.42	0.82	21.56	21.09	0.80	21.27
	Harris & MI	21.59	0.66	21.77	22.73	0.59	22.65	22.66	0.45	22.08	22.01	0.27	22.66

Table 3: ISNR statistics (in dB) for the compared super-resolution methods for the *clock* sequence. Baseline is the SURF-based registration method

ISNR	Registration Method	Hardie *et al.* [16]			He and Kondi [18]			Farsiu *et al.* [14]			Šroubek and Flusser [38]		
		mean	std	median	mean	std	median	mean	std	median	mean	std	median
35 dB Gaussian noise	SIFT	1.61	0.15	1.58	1.47	0.17	1.38	1.31	0.32	1.53	1.88	0.40	1.17
	SIFT & MI	**1.87**	0.22	1.61	1.40	0.14	1.45	1.57	0.29	1.59	**2.29**	0.21	2.13
	SURF	–	–	–	–	–	–	–	–	–	–	–	–
	SURF & MI	1.42	0.23	1.68	1.39	0.37	1.47	1.80	0.17	1.87	2.02	0.15	2.06
	Harris	1.47	0.16	1.15	**1.97**	0.28	1.77	1.80	0.55	1.81	1.39	0.17	1.21
	Harris & MI	1.52	0.53	1.71	1.71	0.17	1.77	**1.88**	0.41	1.63	1.51	0.19	1.84
30 dB Gaussian noise	SIFT	1.45	0.21	1.42	1.61	0.64	1.48	2.90	0.42	2.40	1.34	0.17	1.46
	SIFT & MI	**1.60**	0.21	1.65	**1.97**	0.21	1.61	**2.99**	0.16	2.71	**1.97**	0.18	1.79
	SURF	–	–	–	–	–	–	–	–	–	–	–	–
	SURF & MI	1.43	0.13	1.39	1.92	0.33	1.55	1.73	0.45	1.69	1.77	0.27	1.92
	Harris	1.44	0.18	1.45	1.75	0.11	1.73	2.29	0.35	1.79	1.45	0.50	1.19
	Harris & MI	1.46	0.18	1.42	1.73	0.31	1.79	2.36	0.55	2.02	1.48	0.22	1.35
25 dB Gaussian noise	SIFT	1.41	0.19	1.65	1.77	0.43	1.92	2.26	0.34	2.13	1.31	0.44	1.61
	SIFT & MI	**1.69**	0.26	1.71	**1.78**	0.21	1.86	**2.55**	0.16	2.51	**1.73**	0.27	1.93
	SURF	–	–	–	–	–	–	–	–	–	–	–	–
	SURF & MI	1.37	0.18	1.31	1.75	0.18	1.78	1.78	0.25	1.47	1.68	0.13	1.43
	Harris	1.42	0.53	1.13	1.69	0.24	1.70	1.88	0.44	1.53	1.49	0.21	1.34
	Harris & MI	1.39	0.16	1.34	1.61	0.16	1.52	1.89	0.52	1.53	1.70	0.32	1.76

indicate the best performance with respect to the registration method (along columns). The method of Farsiu *et al.* [14] is very competitive and provides superior performance in most cases. Notice that the use of mutual information improves the results almost in every case. It only underperforms when

Table 4: SSIM statistics for the compared super-resolution methods for the *clock* sequence. registration method

SSIM	Registration Method	Hardie *et al.* [16]			He and Kondi [18]			Farsiu *et al.* [14]			Šroubek and Flusser [38]		
		mean	std	median	mean	std	median	mean	std	median	mean	std	median
35 dB Gaussian noise	SIFT	0.87	0.02	0.88	0.85	0.01	0.85	0.89	0.01	0.88	0.81	0.01	0.82
	SIFT & MI	0.88	0.01	0.88	0.87	0.02	0.88	0.89	0.04	0.88	0.84	0.02	0.84
	SURF	0.86	0.01	0.86	0.86	0.01	0.86	0.80	0.04	0.80	0.84	0.01	0.83
	SURF & MI	0.87	0.02	0.87	**0.88**	0.02	0.87	0.88	0.03	0.85	**0.86**	0.01	0.85
	Harris	0.85	0.03	0.84	0.81	0.02	0.81	0.83	0.02	0.83	0.85	0.02	0.86
	Harris & MI	**0.89**	0.01	0.88	0.87	0.03	0.88	**0.89**	0.03	0.90	0.85	0.02	0.84
30 dB Gaussian noise	SIFT	0.82	0.01	0.82	0.82	0.01	0.82	0.83	0.01	0.83	0.82	0.01	0.82
	SIFT & MI	0.83	0.03	0.83	0.83	0.01	0.82	0.80	0.04	0.81	**0.83**	0.02	0.84
	SURF	0.84	0.01	0.84	0.83	0.01	0.82	0.78	0.04	0.80	0.80	0.01	0.81
	SURF & MI	**0.85**	0.02	0.84	0.82	0.02	0.82	0.80	0.03	0.80	0.82	0.01	0.83
	Harris	0.83	0.02	0.83	0.82	0.01	0.83	0.83	0.02	0.83	0.81	0.01	0.81
	Harris & MI	**0.85**	0.03	0.87	**0.85**	0.01	0.85	**0.87**	0.04	0.86	0.82	0.03	0.82
25 dB Gaussian noise	SIFT	0.67	0.01	0.67	0.60	0.02	0.60	0.71	0.01	0.71	0.64	0.01	0.65
	SIFT & MI	0.68	0.02	0.67	**0.61**	0.02	0.60	**0.78**	0.04	0.78	0.72	0.02	0.72
	SURF	0.67	0.01	0.67	0.60	0.01	0.60	0.68	0.04	0.70	0.64	0.01	0.64
	SURF & MI	0.68	0.01	0.68	**0.61**	0.02	0.60	0.70	0.01	0.70	0.72	0.01	0.72
	Harris	0.66	0.01	0.66	0.57	0.02	0.56	0.73	0.02	0.73	0.70	0.02	0.72
	Harris & MI	**0.70**	0.01	0.70	0.60	0.02	0.60	0.75	0.05	0.73	**0.74**	0.03	0.75

combined with Harris corners detectors, which is due to the sensitivity to noise of the corner detectors. This drawback may lead to bad initialization of the mutual information criterion and thus the registration method may fail to register the LR images correctly.

Table 4 shows the statistics using the SSIM index. As it can be seen, all methods are consistent in the whole set of experiments. The mutual information criterion improves the performance of the method in all super-resolution algorithms. The SSIM values in bold show the best performance with respect to the registration method.

In Table 5, the numerical results of the HR estimated image using the VIF measure are shown. As it can be observed, the method of Hardie *et al.* [16] provides better results on this dataset. SIFT & MI and Harris & MI are considered to be the best registration methods for these experiments. The VIF values in bold denote the best performance with respect to the registration method. In the general case, mutual information improves the results for the majority of the methods for all experiments.

Experiments over an artificially generated dataset applying a local motion transformation model to the LR images have also been conducted. Image registration is performed by first determining the corresponding features between the estimated HR image and the LR images, which are then pro-

Table 5: VIF statistics for the compared super-resolution methods for the *clock* sequence. registration method

VIF	Registration Method	Hardie *et al.* [16]			He and Kondi [18]			Farsiu *et al.* [14]			Šroubek and Flusser [38]		
		mean	std	median	mean	std	median	mean	std	median	mean	std	median
35 dB Gaussian noise	SIFT	0.80	0.07	0.81	0.78	0.06	0.81	0.47	0.06	0.51	0.59	0.04	0.58
	SIFT & MI	0.77	0.04	0.80	**0.79**	0.02	0.87	**0.57**	0.09	0.56	**0.67**	0.06	0.68
	SURF	0.79	0.07	0.81	0.78	0.06	0.78	0.52	0.09	0.49	0.66	0.02	0.65
	SURF & MI	0.77	0.07	0.78	**0.79**	0.08	0.81	0.53	0.06	0.55	0.63	0.02	0.63
	Harris	0.65	0.18	0.61	0.58	0.12	0.56	0.50	0.06	0.52	0.64	0.05	0.66
	Harris & MI	**0.92**	0.10	0.94	0.76	0.18	0.84	0.56	0.07	0.58	0.64	0.09	0.60
30 dB Gaussian noise	SIFT	0.76	0.07	0.78	0.54	0.10	0.55	0.51	0.04	0.53	0.61	0.05	0.63
	SIFT & MI	0.73	0.11	0.77	0.52	0.10	0.53	**0.58**	0.09	0.54	**0.67**	0.05	0.66
	SURF	0.78	0.04	0.80	0.54	0.10	0.51	0.47	0.08	0.49	0.63	0.02	0.64
	SURF & MI	0.75	0.06	0.74	**0.55**	0.08	0.55	0.51	0.07	0.50	0.64	0.03	0.63
	Harris	0.75	0.13	0.70	0.49	0.11	0.51	056	0.06	0.57	0.59	0.03	0.57
	Harris & MI	**0.84**	0.20	0.93	0.56	0.07	0.56	0.44	0.08	0.42	0.66	0.09	0.64
25 dB Gaussian noise	SIFT	0.74	0.06	0.73	0.50	0.09	0.49	0.44	0.02	0.43	0.56	0.04	0.54
	SIFT & MI	0.70	0.05	0.71	0.50	0.10	0.52	0.55	0.07	0.54	0.65	0.05	0.63
	SURF	0.71	0.06	0.70	0.51	0.10	0.46	0.48	0.07	0.50	0.60	0.02	0.59
	SURF & MI	0.74	0.05	0.71	**0.52**	0.09	0.49	0.51	0.05	0.49	0.63	0.02	0.61
	Harris	0.68	0.13	0.65	0.41	0.04	0.42	0.40	0.06	0.36	0.55	0.05	0.56
	Harris & MI	**0.84**	0.08	0.84	0.50	0.07	0.52	**0.57**	0.09	0.59	**0.64**	0.07	0.66

jected onto the high resolution grid. Then, the mutual information criterion is optimized to refine the registration parameters. Next, the local motion vectors are estimated by applying a block matching algorithm and motion compensation provides us the estimated transformation of the reference image with respect to the current HR estimate. Tables 6, 7, 8 and 9, present the PSNR, ISNR, SSIM and VIF numerical results for this experiment respectively, comparing the local motion against the global motion estimation for Hardie *et al.* [16] and He and Kondi [18] methods. The values in bold indicate the best performance for the corresponding registration method. As it can be observed, mutual information improves the super-resolution results in all experiments. Some representative HR images for the *Star* sequence using the local motion compensation technique are depicted in Figure 5.

An advantage of the proposed scheme is that not only is the reconstructed HR image of better quality but also the algorithm converges faster. This is depicted in Figure 3, where the cost function (5) is drawn with respect to the iteration number for the methods of Hardie *et al.* [16] (3(a)) and the method of He and Kondi [18] (3(b)). We may observe that in all cases the use of mutual information improves the convergence rate compared with the corresponding feature-based registration method.

Convergence of the super-resolution algorithm was achieved when $\|\hat{\mathbf{z}}^{n+1} -$

Table 6: PSNR statistics (in dB) for the compared super-resolution methods for the *star* sequence.

PSNR	Registration Method	Hardie *et al.* [16] Local motion			He and Kondi [18] Local motion			Hardie *et al.* [16] Global motion			He and Kondi [18] Global motion		
		mean	std	median	mean	std	median	mean	std	median	mean	std	median
35 dB Gaussian noise	SIFT	27.72	0.18	27.29	28.65	0.52	28.65	20.59	0.76	20.27	19.92	0.25	19.99
	SIFT & MI	27.85	0.59	27.60	**29.14**	0.90	28.43	20.09	0.33	20.08	19.85	0.81	19.80
	SURF	22.11	0.48	21.81	24.12	0.69	24.33	20.18	0.36	20.14	19.90	0.82	19.70
	SURF & MI	22.37	0.86	21.15	24.67	0.30	25.42	20.18	0.36	20.14	**20.88**	0.42	20.83
	Harris	27.48	0.93	27.34	28.41	0.20	28.04	20.27	0.97	20.55	19.99	0.59	19.78
	Harris & MI	**28.12**	0.18	28.56	28.51	0.70	28.81	**21.21**	0.77	20.95	20.48	0.51	20.41
30 dB Gaussian noise	SIFT	27.66	0.89	27.38	26.24	0.89	26.61	20.12	0.77	20.38	19.10	0.34	19.70
	SIFT & MI	27.84	0.44	27.58	27.94	0.55	27.36	20.27	0.36	20.36	19.94	0.11	19.36
	SURF	22.28	0.91	22.29	21.88	0.37	21.66	20.19	0.35	20.15	19.10	0.83	19.44
	SURF & MI	23.12	0.42	22.22	22.36	0.42	21.51	20.19	0.35	20.15	**20.87**	0.41	20.82
	Harris	27.11	0.24	27.79	27.94	0.40	27.56	20.55	0.49	20.95	19.98	0.59	19.88
	Harris & MI	**28.27**	0.97	28.88	**28.65**	0.82	28.00	**20.59**	0.61	20.39	20.53	0.35	20.53
25 dB Gaussian noise	SIFT	25.47	0.13	25.64	26.09	0.68	25.48	20.27	0.41	20.13	18.83	0.59	18.35
	SIFT & MI	**26.38**	0.87	26.42	26.36	0.88	26.14	**20.63**	0.89	20.23	18.34	0.24	18.10
	SURF	24.17	0.78	24.65	18.98	0.12	18.27	20.15	0.40	20.03	15.92	0.33	15.29
	SURF & MI	24.62	0.27	25.06	21.82	0.39	20.79	20.09	0.33	20.08	**20.53**	0.39	20.50
	Harris	25.81	0.87	26.10	25.79	0.69	25.64	19.61	0.78	20.11	19.80	0.66	19.39
	Harris & MI	26.01	0.13	26.85	**26.97**	0.24	27.25	19.44	0.13	20.09	19.66	0.35	19.69

Table 7: ISNR statistics (in dB) for the compared super-resolution methods for the *star* sequence. Baseline is the SURF-based registration method

ISNR	Registration Method	Hardie *et al.* [16] Local motion			He and Kondi [18] Local motion			Hardie *et al.* [16] Global motion			He and Kondi [18] Global motion		
		mean	std	median	mean	std	median	mean	std	median	mean	std	median
35 dB Gaussian noise	SIFT	0.99	0.39	1.04	1.67	0.68	1.46	0.25	0.76	0.57	0.56	0.25	0.99
	SIFT & MI	**1.68**	0.13	1.25	1.17	0.42	0.54	0.43	0.43	0.30	**0.63**	0.81	0.69
	SURF	–	–	–	–	–	–	–	–	–	–	–	–
	SURF & MI	1.34	0.86	2.07	**1.93**	0.40	1.42	**0.65**	0.36	0.70	0.60	0.42	0.65
	Harris	1.10	0.79	1.79	1.28	0.92	1.84	0.57	0.97	0.30	0.49	0.59	0.78
	Harris & MI	0.96	0.02	0.96	1.39	0.70	1.19	0.62	0.77	0.89	0.35	0.52	0.43
30 dB Gaussian noise	SIFT	1.08	0.75	1.24	1.42	0.88	2.07	0.55	0.75	0.31	0.11	0.34	0.44
	SIFT & MI	**2.11**	0.17	2.39	**2.02**	0.41	2.25	**0.56**	0.36	0.48	**1.17**	0.13	1.74
	SURF	–	–	–	–	–	–	–	–	–	–	–	–
	SURF & MI	1.55	0.43	1.47	1.30	0.42	2.19	0.46	0.07	0.47	0.24	0.40	0.28
	Harris	1.28	0.28	1.78	1.50	0.19	1.88	0.19	0.49	0.73	0.13	0.60	0.26
	Harris & MI	0.95	0.02	0.97	2.01	0.73	1.31	0.11	0.60	0.27	0.15	0.34	0.17
25 dB Gaussian noise	SIFT	1.70	0.55	1.29	1.56	0.18	1.85	0.17	0.27	0.34	0.12	0.66	0.39
	SIFT & MI	**1.73**	0.31	1.23	1.77	0.21	1.60	0.15	0.89	0.42	**1.56**	0.15	1.02
	SURF	–	–	–	–	–	–	–	–	–	–	–	–
	SURF & MI	1.21	0.78	1.53	1.77	0.38	1.78	0.43	0.43	0.30	0.29	0.51	0.19
	Harris	1.02	0.73	1.59	1.76	0.27	1.55	**0.77**	0.75	0.40	0.72	0.11	0.76
	Harris & MI	0.97	0.01	0.98	**1.80**	0.65	20.2	0.15	0.15	0.50	0.82	0.35	0.87

$\hat{\mathbf{z}}^n\|/\|\hat{\mathbf{z}}^n\| < 10^{-5}$ or until 30 iterations were reached. This criterion was used for Hardie *et al.* [16] and He and Kondi [18] algorithms. For the

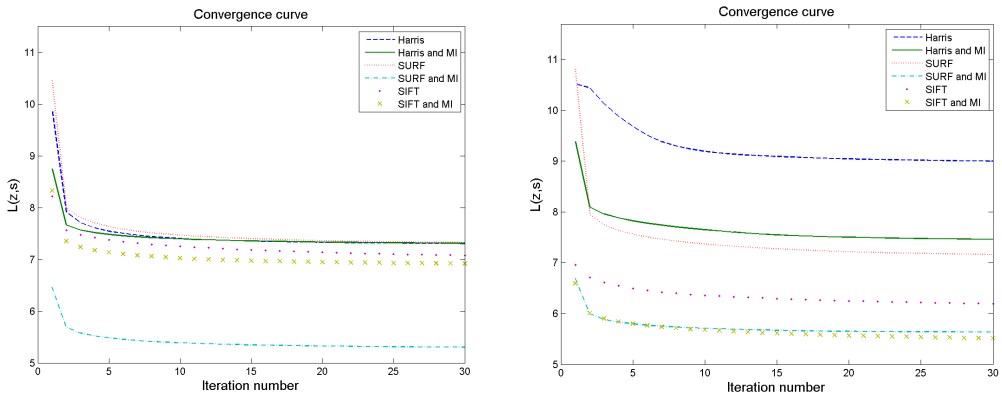
Table 8: SSIM statistics for the compared super-resolution methods for the *star* sequence.

SSIM	Registration Method	Hardie *et al.* [16] Local motion			He and Kondi [18] Local motion			Hardie *et al.* [16] Global motion			He and Kondi [18] Global motion		
		mean	std	median	mean	std	median	mean	std	median	mean	std	median
35 dB Gaussian noise	SIFT	0.93	0.01	0.92	0.94	0.03	0.93	0.49	0.11	0.46	0.82	0.02	0.83
	SIFT & MI	0.93	0.02	0.93	0.94	0.02	0.94	**0.57**	0.15	0.53	0.92	0.02	0.92
	SURF	0.95	0.02	0.98	0.94	0.03	0.93	0.46	0.06	0.48	0.30	0.13	0.22
	SURF & MI	**0.96**	0.02	0.95	0.95	0.02	0.95	0.46	0.06	0.48	**0.94**	0.01	0.94
	Harris	0.79	0.17	0.86	0.94	0.01	0.96	0.52	0.06	0.50	0.49	0.08	0.48
	Harris & MI	0.84	0.14	0.86	**0.96**	0.02	0.95	0.54	0.08	0.53	0.84	0.01	0.84
30 dB Gaussian noise	SIFT	0.92	0.03	0.92	0.94	0.04	0.97	0.44	0.10	0.47	0.83	0.02	0.85
	SIFT & MI	0.95	0.03	0.97	**0.95**	0.02	0.96	0.46	0.07	0.50	**0.94**	0.01	0.95
	SURF	0.93	0.02	0.93	**0.95**	0.02	0.95	0.46	0.07	0.47	0.64	0.17	0.68
	SURF & MI	**0.96**	0.01	0.96	**0.95**	0.03	0.95	0.46	0.07	0.47	**0.94**	0.01	0.94
	Harris	0.78	0.02	0.80	0.93	0.02	0.91	0.50	0.08	0.50	0.49	0.08	0.50
	Harris & MI	0.85	0.13	0.78	0.94	0.03	0.94	**0.52**	0.06	0.50	0.85	0.03	0.86
25 dB Gaussian noise	SIFT	0.95	0.03	0.97	0.90	0.02	0.93	0.44	0.01	0.45	0.83	0.02	0.83
	SIFT & MI	**0.96**	0.03	0.98	**0.96**	0.10	0.93	0.48	0.14	0.41	**0.94**	0.03	0.95
	SURF	0.94	0.01	0.96	0.95	0.02	0.97	0.43	0.07	0.43	0.49	0.19	0.37
	SURF & MI	0.95	0.03	0.95	**0.96**	0.03	0.97	**0.57**	0.15	0.53	0.93	0.01	0.93
	Harris	0.85	0.03	0.86	0.95	0.03	0.98	0.56	0.10	0.61	0.72	0.11	0.76
	Harris & MI	0.86	0.09	0.82	**0.96**	0.03	0.98	0.56	0.14	0.63	0.85	0.09	0.87

Table 9: VIF statistics for the compared super-resolution methods for the *star* sequence.

VIF	Registration Method	Hardie *et al.* [16] Local motion			He and Kondi [18] Local motion			Hardie *et al.* [16] Global motion			He and Kondi [18] Global motion		
		mean	std	median	mean	std	median	mean	std	median	mean	std	median
35 dB Gaussian noise	SIFT	0.77	0.05	0.79	0.80	0.07	0.80	0.47	0.05	0.46	0.46	0.06	0.47
	SIFT & MI	0.80	0.07	0.74	**0.83**	0.03	0.83	0.45	0.01	0.46	0.47	0.05	0.51
	SURF	0.54	0.17	0.52	0.60	0.11	0.64	0.46	0.06	0.48	0.34	0.23	0.38
	SURF & MI	0.55	0.13	0.50	0.63	0.21	0.72	0.47	0.01	0.48	**0.57**	0.01	0.58
	Harris	0.79	0.08	0.86	0.82	0.09	0.86	0.44	0.07	0.49	0.54	0.02	0.55
	Harris & MI	**0.81**	0.14	0.86	0.80	0.13	0.86	**0.49**	0.01	0.49	0.48	0.01	0.49
30 dB Gaussian noise	SIFT	0.76	0.02	0.78	0.75	0.03	0.74	0.44	0.05	0.46	0.45	0.06	0.43
	SIFT & MI	0.77	0.06	0.78	0.78	0.07	0.77	0.47	0.02	0.48	0.50	0.06	0.50
	SURF	0.57	0.16	0.54	0.59	0.16	0.54	0.46	0.07	0.47	0.39	0.18	0.42
	SURF & MI	0.59	0.17	0.54	0.55	0.20	0.51	0.47	0.01	0.48	0.55	0.01	0.56
	Harris	0.78	0.12	0.85	0.78	0.12	0.85	0.49	0.08	0.50	0.49	0.06	0.51
	Harris & MI	**0.78**	0.13	0.85	**0.81**	0.09	0.84	**0.52**	0.01	0.50	**0.51**	0.07	0.79
25 dB Gaussian noise	SIFT	0.74	0.06	0.75	0.75	0.07	0.76	0.44	0.01	0.45	0.47	0.08	0.44
	SIFT & MI	0.74	0.07	0.75	0.73	0.07	0.75	0.46	0.05	0.45	0.47	0.09	0.49
	SURF	0.59	0.10	0.65	0.37	0.08	0.34	0.45	0.01	0.46	0.20	0.19	0.19
	SURF & MI	0.63	0.16	0.71	0.49	0.18	0.41	0.45	0.01	0.46	**0.51**	0.01	0.52
	Harris	0.74	0.14	0.81	0.71	0.18	0.77	0.56	0.10	0.61	0.50	0.13	0.45
	Harris & MI	**0.76**	0.09	0.72	**0.76**	0.09	0.79	**0.56**	0.14	0.63	0.38	0.09	0.36

methods of Farsiu *et al.* [14] and Šroubek and Flusser [38] all the parameters required by the methods were set to their default as proposed by the authors (e.g. the algorithms converged in average at about 10 iterations). Figure



(a) (b)

Figure 3: The cost function $L(\mathbf{z}, \mathbf{s})$, with respect to the iteration number compared with (a) Hardie *et al.* [16] (b) He and Kondi [18] methods.

4 depicts representative results of the implementation of [16], [18], [14] and [38] for the *clock* sequence with respect to the optimal registration method. Equivalent results are also shown in Figures 6, 7, 8 and 9 for four additional samples of reconstructed HR images and with different noise degradations. The parameters used for Hardie *et al.* method [16] were manually set to $\lambda = 100$ and $\sigma_\eta^2 = 1$ while, the parameters used for He and Kondi [18] method were automatically computed as described in [18].

It is worth noticing that in Figure 8, for the reconstructed HR image *Artificial Lena*, the best registration method is the Harris corners combined with the mutual information. This is due to the high gradient information, which leads to better initialization of the mutual information criterion. Also, notice that similar numerical results are obtained for all the HR reconstructed images. The complete set of experiments may be seen online at `http://www.cs.uoi.gr/~mvrigkas/MAP_SR.html`.

4. Conclusion

In this paper, we presented a two-step registration approach for image super-resolution, which is supported by a feature-based image registration followed by a registration relying on the maximization of mutual information. First, an estimate of the transformation parameters in the least squares sense is provided. Second, the influence of misregistration is improved by the use

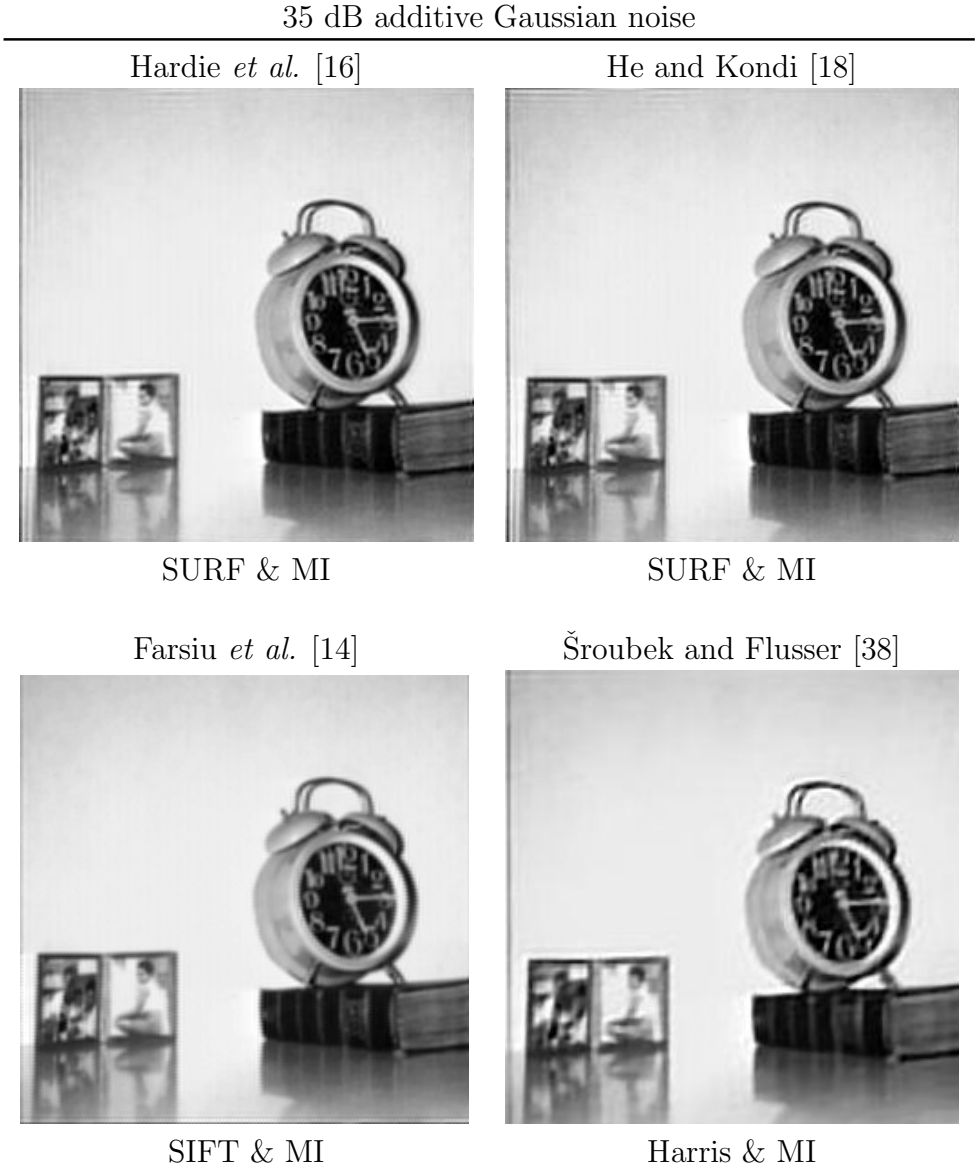


Figure 4: Reconstructed HR images for the 256×256 *clock* sequence (obtained from the USC-SIPI database [1]). For each SR method, representative HR images are shown with respect to the optimal registration method.

of mutual information. By these means, the main drawback of mutual information, that is, the large number of local maxima is overcome. A solution

of high accuracy is obtained for the super-resolved image when compared to images reconstructed without the registration step using the mutual information. The overall reconstruction algorithm converges faster than the standard solution based only on landmark correspondence and registration [9]. These issues were examined using several MAP SR algorithms.

Finally, let us notice that we have also tried to register the LR images by the mutual information method only, without initialization by the feature-based registration. In all cases the resulting estimation of the registration parameters was erroneous leading to a HR image of very low quality. The reason that mutual information itself may fail to register correct the LR images is its proneness of being trapped in local maxima.

References

[1] The USC-SIPI image database. `http://sipi.usc.edu/database/`, 1977.

[2] `http://users.soe.ucsc.edu/~milanfar/software/superresolution.html`, 2006.

[3] `http://staff.utia.cas.cz/sroubekf`, 2006.

[4] L. Baboulaz and P. L. Dragotti. Exact feature extraction using finite rate of innovation principles with an application to image super-resolution. *IEEE Transactions on Image Processing*, 18(2):281–298, February 2009.

[5] S. Baker and T. Kanade. Super resolution optical flow. Technical report, Robotics Institute, October 1999.

[6] B. Bascle, A. Blake, and A. Zisserman. Motion deblurring and super-resolution from an image sequence. In *Proc. European Conference on Computer Vision*, pages 573–582, 1996.

[7] H. Bay, T. Tuytelaars, and L. V. Gool. SURF: speeded up robust features. In *European Conference on Computer Vision*, pages 404–417, 2006.

[8] J. C. Brailean and A. K. Katsaggelos. Simultaneous recursive displacement estimation and restoration of noisy-blurred image sequences. *IEEE Transactions in Image Processing*, 4(9):1236–1251, September 1995.

[9] D. Capel and A. Zisserman. Computer vision applied to super-resolution. *IEEE Signal Processing Magazine*, 2003.

[10] M. M. Chang, A. M. Tekalp, and M. I. Sezan. Simultaneous motion estimation and segmentation. *IEEE Transactions in Imgage Processing*, 6(9):1326–1333, September 1997.

[11] G. K. Chantas, N. P. Galatsanos, and N. A. Woods. Super-resolution based on fast registration and maximum a posteriori reconstruction. *IEEE Transactions on Image Processing*, 16(7):1821–1830, July 2007.

[12] Y. Chen, H. Wang, T. Fang, and J. Tyan. Mutual information regularized Bayesian framework for multiple image restoration. In *Proc.* 10^{th} *IEEE International Conference on Computer Vision*, pages 190–197, 2005.

[13] P. E. Eren, M. I. Sezan, and A. M. Tekalp. Robust, object-based high-resolution image reconstruction from low-resolution video. *IEEE Transactions in Imgage Processing*, 6(10):1446–1451, October 1997.

[14] S. Farsiu, M. Elad, and P. Milanfar. Multi-frame demosaicing and super-resolution of color images. *IEEE Transactions on Image Processing*, 15:141–159, 2006.

[15] R. Fransens, C. Strecha, and L. V. Gool. Optical flow based super-resolution: A probabilistic approach. *Computuer Vision and Image Understanding*, 106:106–115, April 2007.

[16] R. C. Hardie, K. J. Barnard, and E. E. Armstrong. Joint MAP image registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing*, 6(12):1621–1633, December 1997.

[17] C. G. Harris and M. J. Stephens. A combined corner and edge detector. In *Proc.* 4^{th} *Alvey Vision Conference, Manchester*, pages 147–151, 1988.

[18] H. He and L. P. Kondi. Resolution enhancement of video sequences with simultaneous estimation of the regularization parameter. *SPIE Journal of Electronic Imaging*, 13(3):586–596, 2004.

[19] H. He and L. P. Kondi. An image super-resolution algorithm for different error levels per frame. *IEEE Transactions on Image Processing*, 15(3):592–603, March 2006.

[20] A. Kanemura, S. Maeda, and S. Ishii. Edge-preserving Bayesian image superresolution based on compound Markov random fields. In *Proc.* 17^{th} *International conference on Artificial Neural Networks*, pages 611–620, Porto, Portugal, 2007.

[21] A. K. Katsaggelos. Iterative image restoration algorithms. *Optical Engineering*, 28(7):735–748, 1989.

[22] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[23] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc.* 7^{th} *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

[24] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Mulimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, April 1997.

[25] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman. Bayesian image super-resolution, continued. In *Advances in Neural Information Processing Systems*, pages 1089–1096. MIT Press, 2006.

[26] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman. Overcoming registration uncertainty in image super-resolution: Maximize or marginalize? *EURASIP Journal on Advances in Signal Processing*, 2007:Article ID 23565, 14 pages, 2007.

[27] J. Pluim, A. Maintz, and M. Viergever. Interpolation artefacts in mutual information-based image registration. *Computer Vision and Image Understanding*, 77(9):211–232, 2000.

[28] C. A. Segall, A. K. Katsaggelos, R. Molina, and J. Mateos. Bayesian resolution enhancement of compressed video. *IEEE Transactions on Image Processing*, 13:898–911, 2004.

[29] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

[30] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image aligment. *Pattern Recognition*, 32(1):71–86, 1999.

[31] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan. High-resolution image reconstruction from lower-resolution image sequences and space varying image restoration. In *International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 169–172, 1992.

[32] M. E. Tipping and C. M. Bishop. Bayesian image super-resolution. In *Advances in Neural Information Processing Systems*, pages 1303–1310. MIT Press, 2003.

[33] R. Y. Tsai and T. S. Huang. *Advances in Computer Vision and Image Processing*, volume 1, chapter 7, pages 317–339. JAI Press, Greenwich, Conn., USA, 1984.

[34] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:376–380, 1991.

[35] P. Vandewalle, S. Süsstrunk, and M. Vetterli. A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP Journal on Applied Signal Processing*, pages Article ID 71459, 14 pages, 2006.

[36] P. Viola and W. M. Wells. Aligment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[37] M. Vrigkas, C. Nikou, and L. P. Kondi. On the improvement of image registration for high accuracy super-resolution. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 981–984, 2011.

[38] F. Šroubek and J. Flusser. Resolution enhancement via probabilistic deconvolution of multiple degraded images. *Pattern Recognition Letters*, 27(4):287–293, 2006.

[39] I. Vujovic and I. Kuzmanic. Wavelet quasi-superresolution in marine applications. In 48^{th} *International Symposium ELMAR-2006 focused on Multimedia Signal Processing and Communications*, pages 65–68, june 2006.

[40] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[41] B. Zhang, J. Liu, J. Chu, and J. Qiao. A mutual information based sub-pixel registration method for image super resolution. In 5^{th} *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 422–425, 2009.

[42] F. Zhou, W. Yang, and Q. Liao. A coarse-to-fine subpixel registration method to recover local perspective deformation in the application of image super-resolution. *IEEE Transactions on Image Processing*, 21(1):53–66, January 2012.

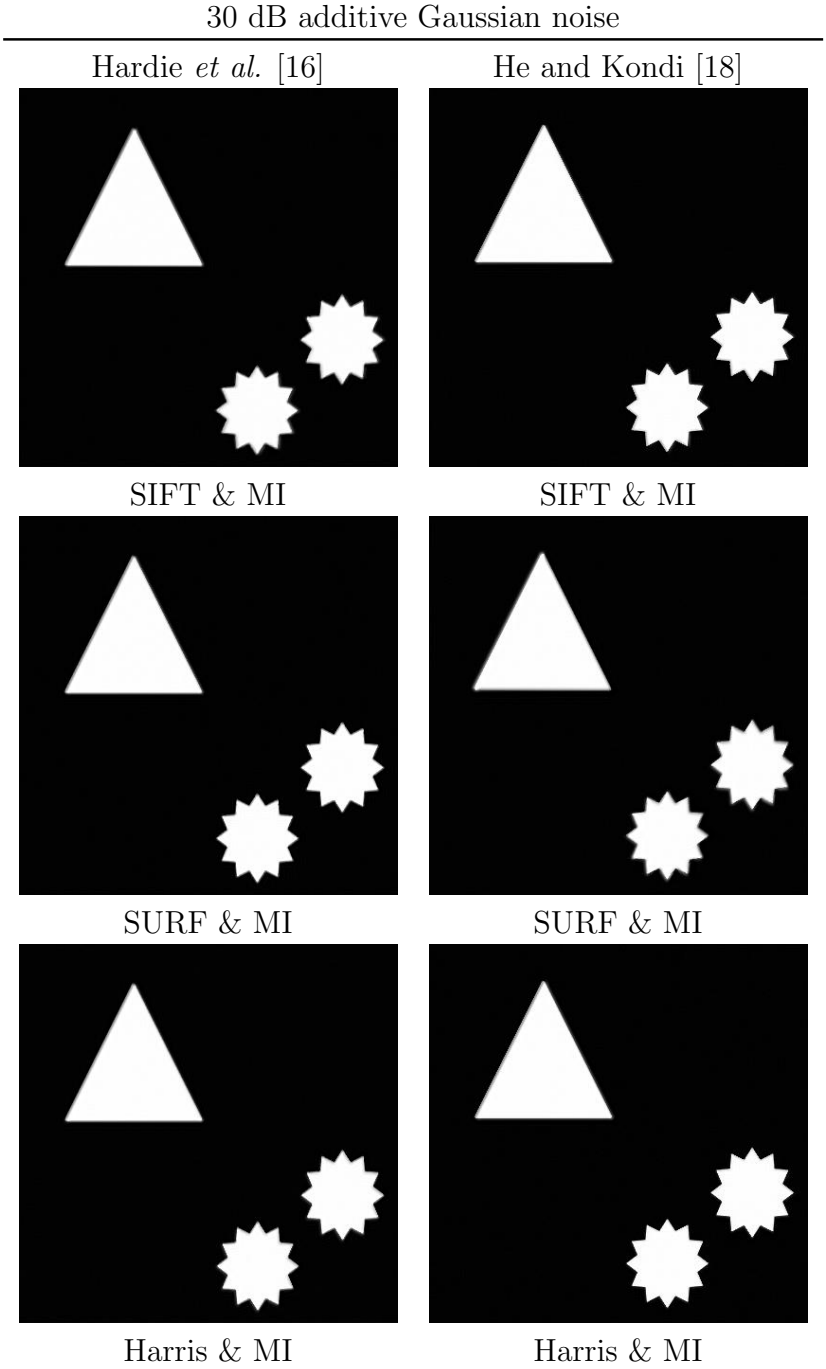


Figure 5: Reconstructed HR images with local motion compensation for the 400×400 *Star* sequence. For each SR method, representative HR images are shown with respect to the optimal registration method.

35 dB additive Gaussian noise

Hardie *et al.* [16]

SURF & MI

He and Kondi [18]

SURF & MI

Farsiu *et al.* [14]

SURF & MI

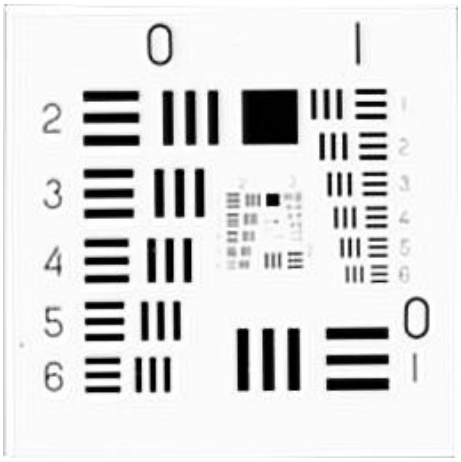
Šroubek and Flusser [38]

SURF & MI

Figure 6: Reconstructed HR images for the 256×256 *Boat* sequence (obtained from the USC-SIPI database [1]). For each SR method, representative HR images are shown with respect to the optimal registration method.

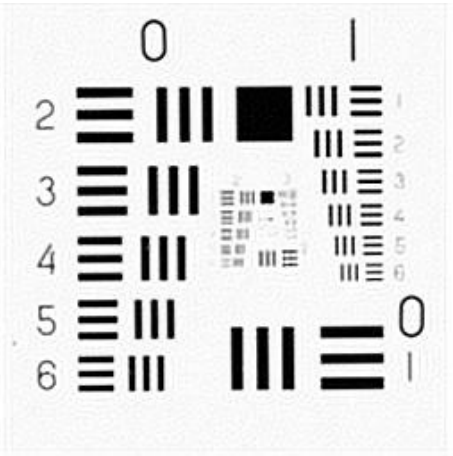
30 dB additive Gaussian noise

Hardie *et al.* [16]



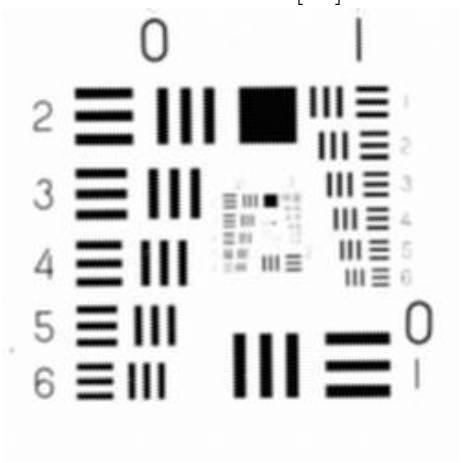
SIFT & MI

He and Kondi [18]



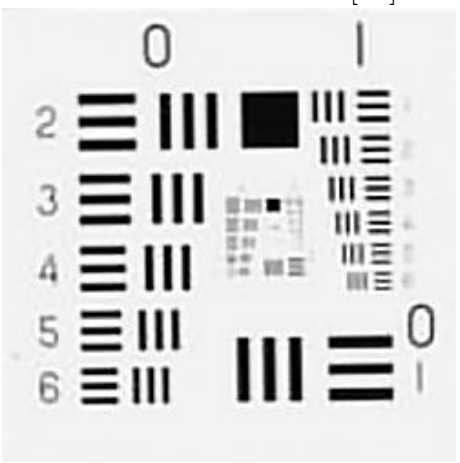
SIFT & MI

Farsiu *et al.* [14]



SIFT & MI

Šroubek and Flusser [38]



SIFT & MI

Figure 7: Reconstructed HR images for the 256×256 *Eye chart* sequence (obtained from the USC-SIPI database [1]). For each SR method, representative HR images are shown with respect to the optimal registration method.

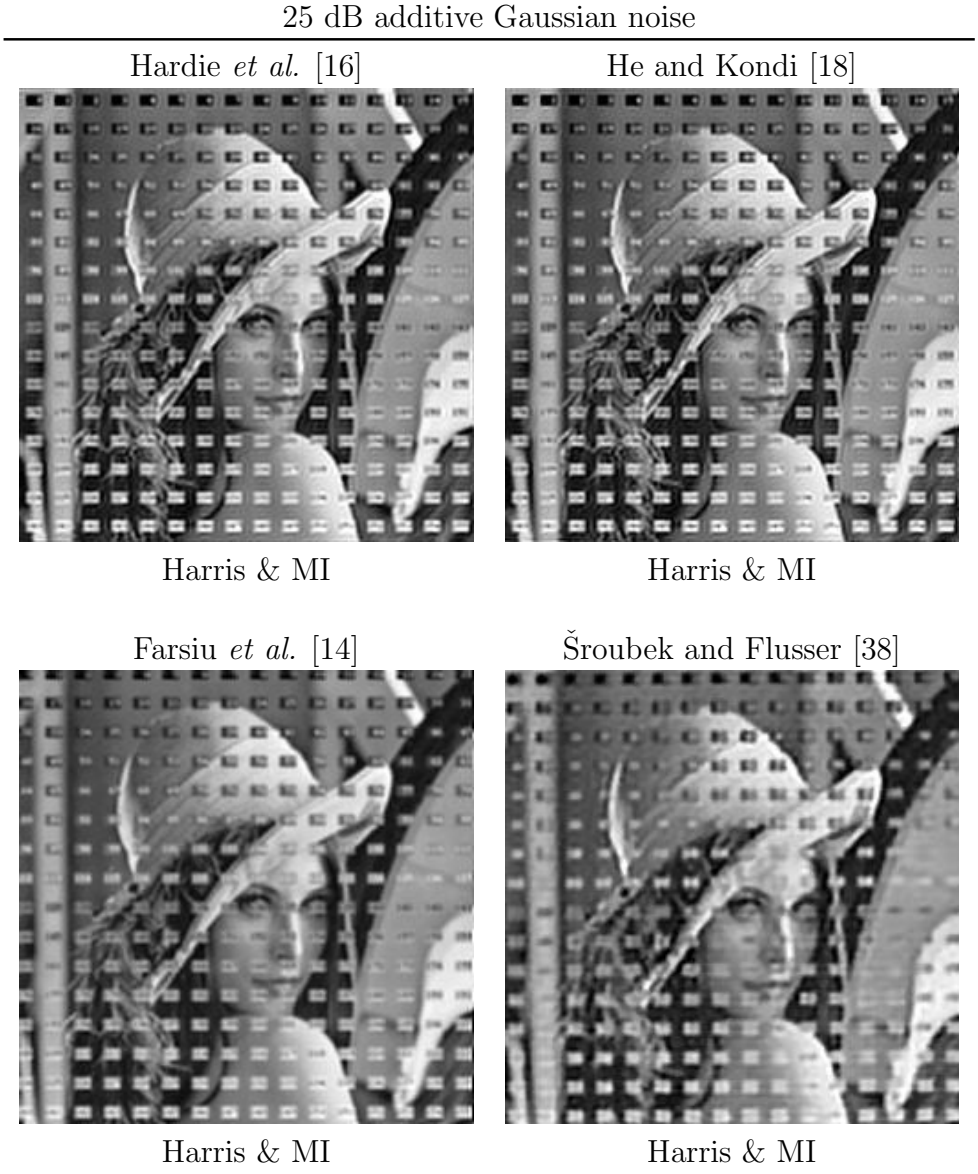


Figure 8: Reconstructed HR images for the 256×256 *Artificial Lena* sequence (obtained from the USC-SIPI database [1]). For each SR method, representative HR images are shown with respect to the optimal registration method.

30 dB additive Gaussian noise

Hardie *et al.* [16]

SIFT & MI

He and Kondi [18]

SIFT & MI

Farsiu *et al.* [14]

SIFT & MI

Šroubek and Flusser [38]

SIFT & MI

Figure 9: Reconstructed HR images for the 256×190 *Car* sequence. For each SR method, representative HR images are shown with respect to the optimal registration method.