Performance Evaluation of Multi-frame Super-resolution Algorithms

Kyle Nelson, Asim Bhatti, Saeid Nahavandi
Centre for Intelligent Systems Research
Deakin University
Geelong, Victoria, Australia

Abstract—Multi-frame super-resolution algorithms aim to increase spatial resolution by fusing information from several low-resolution perspectives of a scene. While a wide array of super-resolution algorithms now exist, the comparative capability of these techniques in practical scenarios has not been adequately explored. In addition, a standard quantitative method for assessing the relative merit of super-resolution algorithms is required. This paper presents a comprehensive practical comparison of existing super-resolution techniques using a shared platform and 4 common greyscale reference images. In total, 13 different super-resolution algorithms are evaluated, and as accurate alignment is critical to the super-resolution process, 6 registration algorithms are also included in the analysis. Pixel-based visual information fidelity (VIFP) is selected from the 12 image quality metrics reviewed as the measure most suited to the appraisal of super-resolved images. Experimental results show that Bayesian super-resolution methods utilizing the simultaneous autoregressive (SAR) prior produce the highest quality images when combined with generalized stochastic Lucas-Kanade optical flow registration.

Keywords—super-resolution, multi-frame, image enhancement, image quality, performance evaluation, comparison

I. INTRODUCTION

Super-resolution is a method of post-processing image enhancement that increases the spatial resolution of video or images. This paper focuses on multi-frame super-resolution techniques, which aim to improve image resolution by fusing information from multiple low-resolution images. Multi-frame super-resolution algorithms require several slightly different perspectives of the same scene, a non-integer pixel shift between images, such that each input view captures a marginally different representation of the scene. The subtle differences between images are then exploited to create an image that exceeds the spatial resolution of the input frames.

While a wide array of super-resolution algorithms now exist, far less attention has been paid to evaluating and comparing the capability of these techniques in practical scenarios. When publishing a new algorithm a majority of researchers provide results obtained using their method but seldom compare its standing relative to existing techniques. In addition, super-resolution algorithms are most commonly evaluated subjectively, further inhibiting comparison. Hence, the primary goal of this paper is to present a comprehensive and practical comparison of existing super-resolution techniques.

In total, 13 different super-resolution algorithms are evaluated using a common platform. Each method is examined using 4 common grayscale images; algorithm execution times are also compared along with visual results. When assessing the practical capabilities of different approaches to super-resolution it is imperative that image registration is also investigated, as techniques rely on the smallest of differences between frames. While registration parameters are often assumed known, 6 registration algorithms used in conjunction with super-resolution are also included in our analysis.

This paper also aims to identify the image quality metric most suited to the appraisal of super-resolved images. In the event that super-resolution results are objectively assessed this is often limited to evaluating the peak signal-to-noise ratio (PSNR), with respect to an original high-resolution image, compared to bicubic or cubic-spline interpolation. However, existing research has suggested that objective measures such as mean squared error (MSE) and PSNR are not suited to assessing the quality of super-resolution results [1,2] and that subjective evaluation can give a better indication of visual image quality [3]. While subjective inspections are indeed an ideal evaluation method, in practice the time and resources needed to conduct a full subjective assessment of each super-resolved image produced are simply not available. In practical situations, a method of quickly measuring the quality of a super-resolved image is required.

It is generally agreed that measures such as MSE and PSNR can assess the difference between two images, but do not necessarily reflect the level of image quality as perceived by the human visual system (HVS). However, this should not imply that other metrics cannot accurately assess super-resolved image quality. Despite there being a wide array of image quality metrics, many of which have been designed with the HVS in mind, very few have been utilized in conjunction with super-resolution. The universal quality index (UQI) and pixel-based visual information fidelity (VIFP) are noticeably absent from existing super-resolution performance evaluations. In order to ensure that super-resolution results are assessed using a broad range of image quality measures, 12 different metrics have been used in this review.
This performance evaluation was purposely implemented using 4 standard images and readily available implementations of super-resolution algorithms in the MATLAB platform commonly used by image processing researchers so that our results may easily be reproduced. We hope that by providing the datasets used and ground truth registration data online this will encourage other researchers to repeat our experiments with different datasets or to evaluate their own algorithms and report their comparative results, leading to benchmark image quality results for new methods in the field. We envisage that this may eventually result in an accepted group of datasets and a standard assessment methodology for the super-resolution community, allowing researchers to evaluate new algorithms and demonstrate their improvement over the current state of the art.

II. EXISTING WORK

A number of performance evaluations of super-resolution techniques have been conducted in recent years, the most comprehensive being [1-4]. Five common super-resolution techniques are assessed in [4] using a range of subjective and objective measures, while the analysis in [3] reviews six different super-resolution algorithms, the most previously evaluated in a single publication. Techniques are assessed using a range of datasets and the results analyzed both subjectively and objectively using the structural similarity index (SSIM), PSNR, CIEDE2000 and S-CIELAB quality metrics. The authors find that a POCS and non-uniform interpolation method perform well in the ideal case, while restoration based methods prevail when prior information is unknown. The color difference equation CIEDE2000 was found to agree more closely with subjective appraisals than SSIM and PSNR. In [2] the results from four super-resolution methods are compared to bilinear interpolation objectively using the MSE, PSNR and multi-scale SSIM index (MS-SSIM). Experiments conducted with the well-known Lena and Mandrill images favor a method by Vandewalle et al., while the authors conclude that MS-SSIM is better suited to measuring the quality of super-resolved images than MSE and PSNR. The review in [1] compares five different techniques using four reference images, but image quality is only evaluated using MSE and PSNR, which again are deemed unsuitable for the task.

III. SUPER-RESOLUTION ALGORITHMS

The super-resolution algorithms compared in this evaluation are all readily available MATLAB implementations obtained from three super-resolution research packages [5-7]. While comprehensive, this is by no means an exhaustive review of all available super-resolution implementations, several others were not included for a range of reasons. Some were omitted as they implemented single frame super-resolution techniques or learning-based methods which require training images. Others were excluded simply because they produced exceptionally poor results, took an extremely long time to execute or the code would not execute at all. Code written in languages other than MATLAB, such as C, was not included nor were protected MATLAB functions, as they could not be interfaced with our evaluation platform. Commercial software was also omitted from this survey, though there are a number of available packages which utilize super-resolution technology.

In order to perform an experimental comparison of the identified registration and super-resolution techniques an evaluation platform was created to interface with the different routines. This was designed in a modular fashion so that each of the 4 datasets could be tested with any of the 6 registration and 13 super-resolution algorithms. Each of the routines reviewed were modified only to use common data structures so that it was possible to mix-and-match algorithms from different sources, their function was not altered in any way. Existing routines created by other labs were used in this evaluation to ensure impartial and objective assessment of results. Hence, all evaluated algorithms were coded as intended by the original authors and are therefore not influenced or in any way hindered by our programming. Also, those wishing to reproduce our results can easily do so by obtaining the datasets and ground truth registration parameters from http://www.deakin.edu.au/cisr, and the registration and super-resolution algorithms directly from [5-7].

A. Image Registration

The following section examines the image registration techniques implemented in this review.

1) Lucas-Kanade Optical Flow

Optical flow registration is primarily concerned with analyzing the flow or displacement of pixels across a sequence of images. The Lucas-Kanade method [8] assumes that displacement between sequential frames is small and that neighboring points display similar motion. The displacement of a given pixel is therefore estimated by finding the least squares solution to a series of optical flow equations over a local region of interest. In this review, two different implementations of Lucas-Kanade optical flow are analyzed, one using a Gaussian pyramid (GPLK) and the other a generalized stochastic version (GSLK).

2) Method of Keren, Peleg and Brada

In [9] an iterative registration process is proposed to measure translation and rotation to sub-pixel accuracy. The error between two images, $f$ and $g$, after rotation and translation is expressed using Taylor expansions:

$$E(a, b, \theta) = \sum \left[ f(x, y) + (a - y\theta - \frac{x\theta^2}{2}) \frac{\partial f}{\partial x} + (b + x\theta - \frac{y\theta^2}{2}) \frac{\partial f}{\partial y} - g(x, y) \right]^2,$$

where $a$ is a horizontal translation, $b$ a vertical translation and rotation $\theta$. The error is minimized iteratively by first moving the second image $g$ by the accumulated translation and rotation values using interpolation, then re-estimating the difference between the two images. The implementation evaluated in this review (KPB), again applies a Gaussian pyramid to increase speed and robustness.
3) Frequency-domain Methods

Vandewalle, Süssfrank and Vetterli (VSV) present a frequency domain registration technique in [10] which aligns images based on their low-frequency components, to avoid aliasing effects and high-frequency noise. The rotation between two frames is first estimated from the amplitude of their Fourier transforms, which are unaffected by translations in the spatial domain. The maximal correlation is found by transforming the image samples to polar coordinates and expressing frequency content as a function of an angle determined by integrating over radial lines. The second image is then rotated and translation is calculated from the linear phase difference between Fourier transforms.

In [11] Lucchese and Cortelazzo (LC) propose a frequency domain method for planar motion that operates in Cartesian coordinates. It is shown that the difference between the magnitude of the Fourier transform of one image, and a mirrored version of the Fourier transform magnitude for a second image, produces a pair of orthogonal zero-crossing lines. The angle between these lines and the frequency axis is equal to half the rotation angle between the two images. A three stage coarse-to-fine process is used to refine the rotation estimate; phase correlation is used to distinguish between pairs of possible solutions and again to estimate the translation difference between the images. An implementation of the method by Marcel, Briot, and Murrieta (MBM) included in this review employs the same phase correlation method to estimate both the rotation and translation between images [12].

B. Super-resolution

In this section we review the super-resolution algorithms evaluated in this paper.

1) Bayesian Methods

This performance evaluation compares a number of Bayesian super-resolution algorithms, which vary primarily in their use of prior probability distribution. The method presented in [13] utilizes a quadratic approximation of the total variation (TV) prior, given by

\[
p(x|\alpha) = c\alpha^{LN/2}\exp\left\{-\frac{\alpha}{2}\sum_{i=1}^{LN}\left[(\Delta_h^i(x))^2 + (\Delta_v^i(x))^2\right]\right\},
\]

where \(x\) is the unknown high-resolution image, \(\Delta_h^i(x)\) and \(\Delta_v^i(x)\) are the horizontal and vertical first order differences at pixel \(i\), \(L\) the resolution enhancement factor, \(N\) the number of pixels in the observed low-resolution images, \(\alpha\) a model parameter and \(c\) is a constant. The TV prior is known to preserve image edge information, while imposing overall smoothness. The posterior distribution in this case is approximated by minimizing the Kullback-Leibler distance between the set of model parameters and the posterior.

Villena et al. [14] propose a prior based on the L1-norm (L1) of vertical and horizontal first order differences of image pixel values. The L1 prior probability is described by

\[
p(x|\alpha^h, \alpha^v) \propto (\alpha^h, \alpha^v)^{LN}\exp\left\{-\sum_{i=1}^{LN}\left[\alpha^h(\Delta_h^i(x))^2 + \alpha^v(\Delta_v^i(x))^2\right]\right\},
\]

where \(\alpha^h\) and \(\alpha^v\) are model parameters. The use of two parameters in this case, for the horizontal and vertical directions, makes the L1 model more adaptable to image characteristics than the single-parameter TV model above.

The third Bayesian method investigated in this review implements algorithm 3.1 in [15], which employs the following simultaneous autoregressive (SAR) prior:

\[
p(x|\alpha) \propto \alpha^{LN/2}\exp\left\{-\frac{\alpha}{2}\|Cx\|^2\right\},
\]

where \(C\) is the Laplacian operator. In [16] the sparse, edge-preserving TV prior is combined with the texture-preserving non-sparse SAR model. The implementation evaluated in this paper is therefore labeled (TVSAR). A high-resolution estimate is recovered in this method by minimizing the linear convex combination of Kullback-Leibler divergences.

2) Set Theoretic Methods

The set theoretic methods of super-resolution primarily revolve around the popular projection onto convex sets (POCS) algorithm. The POCS technique limits the solution space for super-resolution reconstruction by intersecting a space containing possible high-resolution images with a series of constraint sets representing desirable image characteristics. Any high-resolution estimate that lies within the intersection of the sets will be, by definition, consistent with the input data and meet all additional constraints and is therefore a feasible solution.

The Papoulis-Gerchberg (PG) algorithm, based on the iterative signal extrapolation method proposed independently by Papoulis [17] and Gerchberg [18], may be thought of as a special case of the POCS technique. By continually projecting onto the space of known pixels and the space of bandlimited images, the algorithm iteratively converges to the desired image at the intersection of the two spaces.
3) **Iterative Back-Projection**

The iterative back-projection (IBP) method proposed by Irani and Peleg [19] iteratively uses the current best guess for the super-resolved image in conjunction with a model of the imaging process to create simulated low-resolution images, which are then compared to the original low-resolution input images. Difference images are formed, by subtracting the real low-resolution images from the simulated low-resolution images, and then used to improve the initial guess by ‘back-projecting’ each value into the super-resolution image space.

4) **Interpolation of Non-uniformly spaced Samples**

Interpolation techniques employed in multi-frame super-resolution fuse low-resolution frames to create a composite image of non-uniformly spaced samples. To improve image resolution, the irregular points are then interpolated and re-sampled on a regularly spaced high-resolution lattice. Included in this analysis is an implementation of the structure-adaptive normalized convolution (NC) method [20].

5) **Robust Super-resolution**

Robust super-resolution algorithms try to accommodate for inaccurate imaging models and outlying data points caused by registration errors, noise or motion blur, which may adversely affect the quality of the high-resolution estimate. Zomet et al. present a robust method in [21] based on iterative back-projection which achieves robustness by minimizing the scaled pixel-wise median instead of minimizing the sum of the difference images. This pixel-wise approach eliminates points that are inconsistent with the chosen imaging model without removing key information arising from image differences such as aliasing. Two separate implementations of this algorithm are evaluated in this review, labeled (ZMT) and (ZRP). Another method by Farsui et al. [22] employs L1-norm minimization in pursuit of robustness against motion errors, blurring and outliers. A robust regularizer called bilateral TV is proposed to preserve edge information, improve the rate of convergence and assist in the search for a stable solution. Three algorithms based on [22] are assessed, robust methods denoted (RSR) and (FREM), and a fast and robust method (FFREM).

C. **Image Quality Metrics**

In pursuit of the metric most suited to the appraisal of super-resolved images, we consider a number of widely used measures such as the MSE, PSNR and SSIM, as well as several less common metrics. Mean-squared error (MSE) is simply the cumulative squared error between a super-resolved image and the reference high-resolution image. The most commonly used measure of the difference between two images is the peak signal-to-noise-ratio (PSNR), given by

$$PSNR = 10 \log_{10} \frac{MAX^2}{MSE},$$

where $MAX$ is the maximum possible pixel value (255 for 8-bit images). The signal-to-noise ratio (SNR) is also calculated for each super-resolution result which utilizes the mean of the reference image, rather than the maximum value.

The structural similarity (SSIM) index [23] is designed to reflect the way the human visual system (HVS) processes structural information. SSIM estimates the luminance, contrast and structure of an image using the mean pixel intensity, variance between the two images and covariance between the images respectively. The local similarity is calculated for small windows across the images, which are then averaged to obtain the mean structural similarity for the entire image. The multi-scale structural similarity index (MS-SSIM) is an extension to SSIM which calculates structural similarity over multiple image resolutions [24]. The predecessor to SSIM, the universal quality index (UQI) is calculated as the product of three components: a correlation coefficient measuring the degree of linear correlation between the two images, the mean luminance similarity and the contrast similarity between the reference and super-resolution result [25].

In [26,27] Sheikh et al. present several quality metrics based on an information theoretic approach to image quality. The information fidelity criterion (IFC) proposed in [26] estimates mutual information by modeling natural scenes using Gaussian scale mixtures, while distortion is represented using a signal attenuation and additive Gaussian noise model. This concept is extended in [27] as the visual information fidelity criterion (VIF), with the quantification of reference image information content and HVS modeling. We also evaluate an efficient multi-scale pixel domain implementation of VIF known as VIFP which uses a scalar random field model for natural scenes, instead of the vector version used in VIF.

This paper also implements the weighted signal to noise ratio (WSNR), which perceptually weights the frequency domain using a contrast sensitivity function to approximate the HVS [28]. In [29] the authors model frequency distortion and noise separately, developing the noise quality measure (NQM) to measure the effect of additive noise. Chandler and Hemami propose the visual signal-to-noise ratio (VSNR) in [30] as a metric for quantifying the visual fidelity of natural images based on nearthreshold and suprathreshold properties of the HVS. In the interests of reproducibility, all quality metric calculations in this evaluation were performed using the implementation in [31].
IV. METHODOLOGY

In order to compare the registration and super-resolution techniques presented, each algorithm is practically evaluated using a range of input datasets created from the 4 common reference images shown in Fig 1. The bird and cameraman images to the top are 256×256 pixels in size, while the boat and elaine reference images on the bottom are 512×512 pixels.

Synthetic datasets have been used for experimentation instead of real images so that meaningful quality metric results can be calculated using the reference images. The input datasets created are designed to artificially mimic the image capture process and to evaluate the response of each algorithm to known super-resolution problem areas such as quantization, aliasing, motion blur and noise. Starting with a reference image from Fig. 1, a symmetric 3×3 Gaussian low-pass filter is first applied to simulate the effect of a camera point-spread function. The blurred image is then downsampled by a factor $M$ and noise is added to achieve a SNR of 20 dB. Finally, we create $K$ copies of the blurred, downsampled and noisy image, offset by a factor $O$. An offset factor of 2 for example, represents each of the low-resolution images being translated by a different but known amount within the range ±2 pixels horizontally, ±2 pixels vertically then rotated by ±2 degrees. The offset applied to each of the $K$ frames in a dataset is random; however the same $K$ offsets are used for datasets created from each of the reference images to allow registration comparisons.

To examine how the number of low-resolution frames can affect the super-resolution result we experimented with 5 different database sizes where $K \in \{4, 8, 16, 32, 64\}$. Offset factors were chosen to permit only minimal difference between frames, $O \in \{1, 2\}$, and to ensure that the super-resolved image has the same dimensions as the reference frame, the resolution enhancement factor $L = M$ where $M \in \{2, 4\}$. Fig. 2 shows example input images created with varying synthetic down sampling factors and varying levels of translation and rotation.
Our experimental methodology involved trialing every possible combination of the 4 reference images, 2 resolution enhancement factors, 2 offset factors, 5 database sizes, 6 registration algorithms and 13 super-resolution techniques. For each of the 6,240 separate experiments conducted the super-resolution execution time was recorded as well as measurements using the 12 quality metrics. All super-resolved output images were saved for future reference along with translations and rotations estimated during the registration phase to enable independent evaluation of the registration and super-resolution algorithms. All experiments were conducted on an Intel Core i7 Q720 CPU @ 1.60 GHz with 6GM RAM.

V. EXPERIMENTAL RESULTS

A. Image Quality Metrics

In order to assess the quality of super-resolved images, a quality metric suited to this field of image enhancement must first be selected. As we seek a metric that performs well across a wide quality spectrum, rather than subjectively classifying the best super-resolution results and comparing this order to rankings by various image quality metrics, we instead take the inverse approach and evaluate how 12 different quality metrics rank poor super-resolution results to identify the most appropriate measure. This approach greatly simplifies the evaluation of image quality metrics and largely removes subjective influences from the comparison.

Of the 6,240 experiments conducted, 4,320 resulted in blatantly poor super-resolved images. Some unfavorable results were caused by inaccurate registration, such as Fig. 3(a), however many more resulted directly from the super-resolution algorithm employed, as in Fig. 3(b). While the severity of the erroneous results varied, it is obvious that super-resolved images such as those in Fig. 3 are undesirable.

While such a large number of poor results are far from ideal, they provide a unique dataset with which to compare image quality metrics to the HVS and analyze their suitability in evaluating super-resolved images. It was found that both the WSNR and IFC metrics consistently ranked blatantly poor images as the highest quality results. Hence, one may question their suitability in evaluating the quality of super-resolved images. The NQM, VIF and VSNR metrics also ranked images such as those in Fig. 3 amongst the best performing results. As one may expect MSE, SNR and PSNR, all agreed on the ranking of super-resolution results; however they still classify some poor images as being high-quality. SSIM and MS-SSIM were far more successful at separating the blatantly poor images from the remaining results, while the best performing metrics were UQI and VIFP.
It was found that VIFP, UQI, MS-SSIM and SSIM generally agree when ranking the quality of super-resolved images, while the correlation between VIFP and UQI was much more pronounced. It is interesting to note that UQI appears to order the quality of super-resolution images better than its successors SSIM and MS-SSIM; the capability of the VIFP measure is even more surprising given that related measures IFC and VIF were among the worst performing metrics. While UQI and VIFP both appear to be suitable metrics for evaluating the quality of super-resolved images, over a large number of experiments VIFP outperformed the UQI metric. Our informal subjective observation also favors the VIFP measure; hence it has been used throughout this paper to evaluate super-resolution image quality.

Poor super-resolution results like Fig. 3 were consistently caused by a select group of registration and super-resolution algorithms. The LC and MBM registration algorithms continually produced super-resolution images such Fig. 3(a), regardless of the super-resolution algorithm employed. Similarly, the IBP and PG super-resolution implementations resulted in obviously inferior images, as well as all five robust methods: ZMT, RSR, ZRP, FFREM and FREM. Hence, these methods have been omitted from all further analysis.

B. Image Registration

In this section we compare the four best-performing registration algorithms using 4 different reference images, 2 synthetic downsampling factors and 2 offset factors with datasets of 64 images. Table I provides a comparison of the error between known values and the offsets estimated by each registration method. With respect to translation estimation the GPLK implementation performed the best on average, while the GSLK method failed to achieve sub-pixel accuracy in the vertical direction. However, GSLK produced the most accurate rotation estimation while VSV performed quite poorly considering that rotation angles were only varied by a maximum of ±2 degrees. Interestingly, GPLK outperformed VSV with respect to rotation estimation, despite the fact the GPLK implementation only estimates translation and ignores rotation altogether. The KPB method performed consistently well in both translation and rotation estimation.

As one may expect, increasing the synthetic downsampling factor \( M \) from 2 to 4 or increasing the offset factor \( O \) from 1 to 2 adversely affected the accuracy of registration estimation, however the relative performance of the techniques was largely unaltered.

C. Super-resolution

This section examines the six best-performing super-resolution implementations. Table II compares the mean quality of super-resolved images produced by each possible combination of the registration and super-resolution algorithms analyzed. Every combination was assessed over 80 experiments including 4 different reference images, 2 resolution enhancement factors, 2 offset factors and 5 database sizes. Table II shows that the SAR and TVSAR methods produce exactly the same results; hence SAR/TVSAR will be considered a single method throughout the remaining analysis. On average the SAR/TVSAR method combined with GSLK registration produced the highest quality results, while the TV and L1 methods also achieved their best results when combined with GSLK registration. The NC super-resolution algorithm achieved the second highest quality results on average, this time combined with the KPB registration method. The POCS algorithm generally resulted in the poorest quality images, its best results achieved with GPLK registration. The relatively high quality results achieved with GSLK and secondly KPB registration initially appears to imply that accurate rotation estimation is more valuable to super-resolution algorithms than translation estimation. The GSLK registration algorithm notably performed the worst with respect to translational motion estimation, yet resulted in the highest quality super-resolution results. However, the GSLK algorithm was also responsible for the worst average super-resolution results when combined with NC or POCS super-resolution. The relative success of GSLK

| TABLE I |
| Comparison of average absolute error (µ) and standard deviation of the error (σ) in translation and rotation for 1024 Images |

<table>
<thead>
<tr>
<th></th>
<th>GSLK</th>
<th>KPB</th>
<th>VSV</th>
<th>GPLK</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-translation (pixels)</td>
<td>0.01891</td>
<td>1.580818</td>
<td>0.001243</td>
<td>0.140776</td>
</tr>
<tr>
<td>Y-translation (pixels)</td>
<td>1.17525</td>
<td>1.613122</td>
<td>0.173094</td>
<td>0.222284</td>
</tr>
<tr>
<td>Rotation (degrees)</td>
<td>0.180635</td>
<td>0.353025</td>
<td>0.204412</td>
<td>0.314722</td>
</tr>
</tbody>
</table>

| TABLE II |
| Comparison of mean VIF (µ) and standard deviation of VIF (σ) over 80 experiments |

<table>
<thead>
<tr>
<th></th>
<th>GSLK</th>
<th>KPB</th>
<th>VSV</th>
<th>GPLK</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAR</td>
<td>0.335355</td>
<td>0.127545</td>
<td>0.26812</td>
<td>0.131538</td>
</tr>
<tr>
<td>TVSAR</td>
<td>0.335355</td>
<td>0.127545</td>
<td>0.26812</td>
<td>0.131538</td>
</tr>
<tr>
<td>NC</td>
<td>0.161379</td>
<td>0.126334</td>
<td>0.32386</td>
<td>0.100089</td>
</tr>
<tr>
<td>TV</td>
<td>0.319929</td>
<td>0.132822</td>
<td>0.244427</td>
<td>0.128213</td>
</tr>
<tr>
<td>L1</td>
<td>0.319224</td>
<td>0.133724</td>
<td>0.243204</td>
<td>0.128324</td>
</tr>
<tr>
<td>POCS</td>
<td>0.16694</td>
<td>0.129393</td>
<td>0.253761</td>
<td>0.088412</td>
</tr>
</tbody>
</table>
registration combined with Bayesian super-resolution techniques and the KPBN combination may instead imply a tight coupling between algorithms implemented by the same sources [5,6].

As it did for registration accuracy, increasing the offset factor $O$ from 1 to 2 adversely affected super-resolution results due to registration errors. However, the impact of increasing the resolution enhancement factor was much more pronounced, across all super-resolution algorithms image quality reduced by almost half when the resolution enhancement factor was doubled. A loss of quality is to be expected however, as the super-resolution algorithms essentially have to estimate a high resolution image from a fewer noisy samples.

During the experiments conducted super-resolution image quality was also greatly affected by the reference image employed, as shown in Fig. 4. On average, the bird reference image was able to be restored to a much higher quality from low-resolution frames than the other references, especially the cameraman image. The POCS method performed significantly worse than other algorithms for the Elaine and boat images, however in general the remaining methods responded similarly to each reference image. In Fig. 4 and all remaining analysis, the super-resolution results for each algorithm reflect only the average of the most favorable registration/super-resolution combination shown in bold in Table II, rather than the average of that super-resolution method across all registration algorithms.

![Figure 4](image4.png)

Figure 4. Mean VIFP for each reference image over 20 experiments.

Fig. 5 highlights the effect that a varying database size has on the quality of the super-resolution result. It would generally be expected that a greater number of input frames is desirable, as this provides a super-resolution algorithm with more samples from which to estimate the high-resolution image, however Fig. 5 demonstrates this is not always the case. The generally superior SAR/TVSAR and NC algorithms show a small decline in image quality as the database size is increased from 4 to 8 and a slight decrease again between 32 and 64 images. The NC algorithm maintains a relatively consistent level of quality across all database sizes, while all other methods display a significant improvement in image quality when the number of available low-resolution frames is increased from 8 to 16.

![Figure 5](image5.png)

Figure 5. Impact of $K$ on mean VIFP over 16 experiments.

In practical super-resolution applications not only is the quality of the resulting image important, but the algorithm execution time is also crucial. Super-resolution time is affected by the dimensions of the low-resolution frames, but even more so by the size of the input database. As expected, Fig. 6 shows a significant increase in processing time as $K$ increases across all algorithms.
except POCS, which generally maintains a super-resolution time between 4 and 8 seconds regardless of $K$. Ultimately, an algorithm producing the highest quality result in the shortest time is sought after. Fig. 7 demonstrates that of the algorithms evaluated, SAR/TVSAR and NC are able to produce comparatively high quality images in a relatively short period of time.

![Figure 6. Effect of $K$ on mean super-resolution time over 20 experiments.](image)

![Figure 7. Comparison of mean VIFP and super-resolution time over 16 experiments.](image)

In general we find that the POCS method produces results in a timely manner, however while they are of a higher quality than the implementations producing blatantly poor results it is frequently outperformed by other methods. It can also be seen that the TV and L1 algorithms are closely matched across all evaluations but are generally exceeded by the SAR and TVSAR methods. The NC algorithm demonstrates consistent image quality across a range of database sizes and exhibits relatively fast execution, however overall SAR and TVSAR were the best performing methods across all measures.

It must be noted that the relative merit of algorithms reviewed in this evaluation, such as the success of the SAR and TVSAR methods or poor results from the robust methods does not necessarily imply that Bayesian super-resolution techniques are superior, only that they produced higher quality results relative to the specific implementations evaluated. Also, as synthetic input datasets were used to enable quantitative analysis of each algorithm the results obtained do not necessarily give an indication of the comparative merits of each algorithm when applied to images from a real camera.

![Figure 8. Close-up super-resolution results where $L=2$, $O=1$, $K=4$: (a) $R=GSLK$, $SR=SAR$, VIFP=0.421015 (b) $R=KPB$, $SR=NC$, VIFP=0.399663 (c) $R=GSLK$, $SR=TV$, VIFP=0.394548 (d) $R=GSLK$, $SR=L1$, VIFP=0.395435 (e) $R=GPLK$, $SR=POCS$, VIFP=0.288345.](image)
VI. CONCLUSION

This paper presents a comprehensive practical comparison of readily available super-resolution implementations, enabling researchers to select the most applicable technique to a given practical scenario. In total 6 registration algorithms and 13 different super-resolution techniques have been evaluated using a common platform and 4 common greyscale reference images. 12 different image quality metrics were used to evaluate super-resolution results and the pixel-based visual information fidelity (VIFP) measure was found to be most suited to the task, followed by the UQI, MS-SSIM and SSIM metrics. Registration results were varied with a method by Keren, Peleg and Brada displaying good overall performance, a generalized stochastic Lucas-Kanade optical flow implementation producing the most accurate rotation estimation and a second optical flow technique demonstrating superior translation estimation. Two Bayesian super-resolution methods based on the simultaneous autoregressive (SAR) prior were found to produce the highest quality results when combined with generalized stochastic Lucas-Kanade optical flow registration, while a normalized convolution implementation also produced high quality super-resolved images in conjunction with registration by Keren, Peleg and Brada. Future analysis will include a subjective evaluation to verify that VIFP is indeed the most suited metric for the appraisal of super-resolved images and additional experiments employing larger offset factors and a greater range of reference frames, including color images.

REFERENCES


