

CONFIDENCE-AWARE LEVENBERG-MARQUARDT OPTIMIZATION FOR JOINT MOTION ESTIMATION AND SUPER-RESOLUTION

Cosmin Bercea, Andreas Maier, and Thomas Köhler

Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

ABSTRACT

Motion estimation across low-resolution frames and the reconstruction of high-resolution images are two coupled subproblems of multi-frame super-resolution. This paper introduces a new joint optimization approach for motion estimation and image reconstruction to address this interdependence. Our method is formulated via non-linear least squares optimization and combines two principles of robust super-resolution. First, to enhance the robustness of the joint estimation, we propose a confidence-aware energy minimization framework augmented with sparse regularization. Second, we develop a tailor-made Levenberg-Marquardt iteration scheme to jointly estimate motion parameters and the high-resolution image along with the corresponding model confidence parameters. Our experiments on simulated and real images confirm that the proposed approach outperforms decoupled motion estimation and image reconstruction as well as related state-of-the-art joint estimation algorithms.

Index Terms— Super-Resolution, Motion estimation, Levenberg-Marquardt optimization

1. INTRODUCTION

In digital imaging, super-resolution (SR) refers to a class of computational techniques to improve the spatial resolution of an imaging system. Over the past years, a variety of algorithms have been introduced that approach SR either from a single-frame [1] or a multi-frame perspective. Multi-frame methods exploit subpixel displacements across a sequence of low-resolution (LR) frames to reconstruct a high-resolution (HR) image [2]. This process can be considered as the combination of an initial motion estimation followed by image reconstruction. If motion is estimated properly, this approach holds the potential to overcome aliasing artifacts due to under-sampling. Unfortunately, in the presence of inaccurate motion estimation, its actual performance deteriorates.

Principally, SR can be approached by two basic methods: 1) The *two-stage* approach [3] considers motion estimation and image reconstruction as independent subproblems and employs image registration on LR frames to determine subpixel motion that is directly utilized for image reconstruction. The main drawback of this approach is the limited accuracy

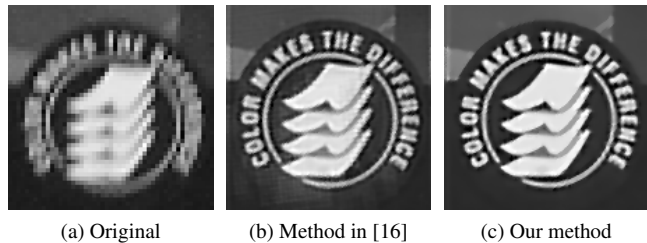


Fig. 1: Super-resolution under inaccurate motion estimation including model outliers related to a camera zoom.

of motion estimation on LR frames [4], which leads to a poor robustness of the overall process. State-of-the-art methods aim at compensating inaccurate motion estimation explicitly by tailor-made outlier detection techniques [5] or implicitly by robust optimization [6]. In this context, SR has been studied with outlier-insensitive observation models including robust error norms [7, 8] or, more recently, by space variant Bayesian models [9]. Another trend are hybrid methods that combine single-frame and multi-frame SR to locally enhance the image reconstruction [10].

2) *Joint* estimation as a complementary class of algorithms considers motion estimation and image reconstruction as two coupled subproblems. This aims at simultaneously estimating motion parameters and the associated HR image to enhance the accuracy of both tasks. For this purpose, alternating minimization regarding both subproblems [11, 12] is a common optimization scheme. More recently, different Bayesian formulations including marginalization over the HR image [13] or the motion parameters [14], as well as variational inference [15] have been developed to overcome the poor convergence of alternating minimization. Other state-of-the-art approaches related to our work include joint Gauss-Newton iterations [16, 17] or linear programming [18]. Compared to the two-stage approach, these schemes are able to correct inaccurate motion estimation up to a certain degree. However, one of their main limitations is that they are often based on simplified models and are sensitive to outliers in the image formation. Fig. 1 depicts this issue on an example dataset affected by outliers in the motion model with a comparison of the Gauss-Newton algorithm in [16] to our proposed method.

This paper proposes a new algorithm for joint motion estimation and image reconstruction. As the main contribution, we propose non-linear least squares optimization regarding the motion parameters and the HR image based on a confidence-aware formulation of both subproblems. In addition, we develop a tailor-made Levenberg-Marquardt iteration scheme to jointly estimate motion parameters, the HR image and model confidence parameters. MATLAB code of this method is available on our webpage as part of our super-resolution toolbox¹.

2. CONFIDENCE-AWARE JOINT MOTION ESTIMATION AND SUPER-RESOLUTION

This section introduces our joint motion estimation and SR algorithm. First, Section 2.1 presents the confidence-aware formulation for our method. Then, in Section 2.2 and Section 2.3, we develop non-linear least-squares estimation for this model and the proposed Levenberg-Marquardt optimization.

2.1. Confidence-Aware Modeling

We describe the formation of a set of LR frames assembled to $\mathbf{y} = (\mathbf{y}^{(1)\top}, \dots, \mathbf{y}^{(K)\top})^\top$, $\mathbf{y}^{(k)} \in \mathbb{R}^M$ from the HR image $\mathbf{x} \in \mathbb{R}^N$ by:

$$\mathbf{y} = \mathbf{W}(\boldsymbol{\theta})\mathbf{x} + \boldsymbol{\epsilon}, \quad (1)$$

where $\mathbf{W}(\boldsymbol{\theta}) = \mathbf{D}\mathbf{H}\mathbf{M}(\boldsymbol{\theta})$ denotes the system matrix and $\boldsymbol{\epsilon}$ is additive noise [3]. We decompose $\mathbf{W}(\boldsymbol{\theta})$ into \mathbf{D} to model subsampling, \mathbf{H} to model space invariant blur related to the camera point spread function (PSF) and $\mathbf{M}(\boldsymbol{\theta})$ to describe subpixel motion of the different frames according to the motion parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(K)})^\top$ relative to \mathbf{x} . In this paper, we limit ourselves to a rigid motion model described by the parameters $\boldsymbol{\theta}^{(k)} = (\cos(\varphi), \sin(\varphi), t_u, t_v)^\top$ with rotation angle φ and translation $\mathbf{t} = (t_u, t_v)^\top$ for the frame $\mathbf{y}^{(k)}$.

We propose to couple SR and motion estimation in a joint optimization framework. Accordingly, we infer the unknown HR image \mathbf{x} and the motion parameters $\boldsymbol{\theta}$ as the minimum of:

$$E(\mathbf{x}, \boldsymbol{\theta}) = (\mathbf{y} - \mathbf{W}(\boldsymbol{\theta})\mathbf{x})^\top \mathbf{B} (\mathbf{y} - \mathbf{W}(\boldsymbol{\theta})\mathbf{x}) + \lambda \cdot R(\mathbf{x}), \quad (2)$$

where $R(\mathbf{x})$ is a regularizer with weight $\lambda \geq 0$. To make the joint estimation in (2) robust regarding outliers, we adopt the space variant observation model developed in our prior work [9] that is defined by the confidence map $\mathbf{B} = \text{diag}(\beta_1, \dots, \beta_{KM})$ to weight the influence of the individual observations by adaptive weights $\beta_i \in [0; 1]$. Moreover, we use weighted bilateral total variation (WBTv) [9] as a sparse prior for edge preserving regularization. We define $R(\mathbf{x})$ by the sparsifying transform \mathbf{S} according to:

$$R(\mathbf{x}) = \|\mathbf{A}\mathbf{S}\mathbf{x}\|_1 = \sum_{l=-P}^P \sum_{m=-P}^P \left\| \mathbf{A}^{l,m} \mathbf{S}^{l,m} \mathbf{x} \right\|_1, \quad (3)$$

where $\mathbf{S}^{l,m} = \alpha_0^{|l|+|m|} (\mathbf{I} - \mathbf{S}_v^l \mathbf{S}_h^m)$, and \mathbf{S}_v^l and \mathbf{S}_h^m denote vertical and horizontal shifts of \mathbf{x} weighted by $\alpha_0 \in [0; 1]$ over a $(2P+1) \times (2P+1)$ window. $\mathbf{A} = \text{diag}(\alpha_1, \dots, \alpha_N)$ with $\alpha_i \in [0; 1]$ are spatially adaptive weights to control the influence of the prior for edge preserving regularization.

2.2. Non-Linear Least-Squares Estimation

Since our energy function in (2) is non-linear w.r.t. $\boldsymbol{\theta}$, we propose a non-linear least-squares estimation of \mathbf{x} and $\boldsymbol{\theta}$. We first derive the least-squares approximation of (3) as:

$$\|\mathbf{S}\mathbf{x}\|_1 \approx \mathbf{x}^\top \mathbf{L}\mathbf{x} = \mathbf{x}^\top \left(\sum_{l=-P}^P \sum_{m=-P}^P \mathbf{L}^{l,m} \right) \mathbf{x}, \quad (4)$$

where $\mathbf{L}^{l,m} = (\mathbf{S}^{l,m})^\top \mathbf{U}^{l,m} \mathbf{S}^{l,m}$. Here, $\mathbf{U}^{l,m}$ is a diagonal matrix associated with the shift (l, m) and is assembled by $U_{ii}^{l,m} = (\max(|z_i|, \tau))^{-1}$, where $\mathbf{z} = \mathbf{S}^{l,m} \mathbf{x}$ and τ is a small threshold ($\tau = 10^{-2}$) to avoid numerical instabilities close to zero. Next, we reformulate (2) as the least-squares term:

$$E(\mathbf{x}, \boldsymbol{\theta}) = \left\| \left(\mathbf{B}^{\frac{1}{2}} \mathbf{r}(\mathbf{x}, \boldsymbol{\theta}), \sqrt{\lambda} \mathbf{A}^{\frac{1}{2}} \mathbf{L}^{\frac{1}{2}} \mathbf{x} \right)^\top \right\|_2^2, \quad (5)$$

where $\mathbf{r}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{y} - \mathbf{W}(\boldsymbol{\theta})\mathbf{x}$ is the residual error. Accordingly, starting from an initial guess \mathbf{x}^0 and $\boldsymbol{\theta}^0$, both parameter sets are iteratively updated according to $\mathbf{x}^t = \mathbf{x}^{t-1} + \Delta \mathbf{x}$ and $\boldsymbol{\theta}^t = \boldsymbol{\theta}^{t-1} + \Delta \boldsymbol{\theta}$ for $t \geq 1$. The updates $\Delta \mathbf{x}$ and $\Delta \boldsymbol{\theta}$ at iteration t are derived by a linear approximation of (5) and are determined as the minimum of the least-squares term:

$$\left\| \mathbf{P}^t \begin{pmatrix} \Delta \boldsymbol{\theta} \\ \Delta \mathbf{x} \end{pmatrix} - \mathbf{f}^t \right\|_2^2, \quad (6)$$

$$\text{where } \mathbf{P}^t = \begin{pmatrix} (\mathbf{B}^t)^{\frac{1}{2}} \mathbf{J}(\mathbf{x}, \boldsymbol{\theta}) & (\mathbf{B}^t)^{\frac{1}{2}} \mathbf{W}(\boldsymbol{\theta}) \\ \mathbf{0} & \sqrt{\lambda} (\mathbf{A}^t)^{\frac{1}{2}} \mathbf{L}^{\frac{1}{2}} \end{pmatrix}, \quad (7)$$

$$\text{and } \mathbf{f}^t = \begin{pmatrix} (\mathbf{B}^t)^{\frac{1}{2}} \mathbf{r}(\mathbf{x}, \boldsymbol{\theta}) \\ -\sqrt{\lambda} (\mathbf{A}^t)^{\frac{1}{2}} \mathbf{L}^{\frac{1}{2}} \mathbf{x} \end{pmatrix}. \quad (8)$$

$\mathbf{J}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{D}\mathbf{H} \frac{\partial(\mathbf{M}(\boldsymbol{\theta})\mathbf{x})}{\partial \boldsymbol{\theta}}$ is the Jacobian of the system matrix $\mathbf{W}(\boldsymbol{\theta})\mathbf{x}$ w.r.t. $\boldsymbol{\theta}$. The derivative of $\mathbf{M}(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$ is computed numerically using bilinear interpolation [16].

For confidence-aware optimization, we estimate the observation weights \mathbf{B}^t from $\mathbf{r} = \mathbf{y} - \mathbf{W}(\boldsymbol{\theta}^{t-1})\mathbf{x}^{t-1}$ according to:

$$\beta_i^t = \begin{cases} 1 & \text{if } |r_i| \leq \sigma_{\text{noise}}^t \\ \sigma_{\text{noise}}^t / |r_i| & \text{otherwise} \end{cases}, \quad (9)$$

and WBTv weights \mathbf{A}^t from $\mathbf{z} = \mathbf{S}\mathbf{x}^{t-1}$ according to:

$$\alpha_i^t = \begin{cases} 1 & \text{if } |z_i| \leq \sigma_{\text{prior}}^t \\ p \left(\sigma_{\text{prior}}^t / |z_i| \right)^{1-p} & \text{otherwise} \end{cases}, \quad (10)$$

where $p \in [0, 1]$ denotes a sparsity parameter. The scale parameters $\sigma_{\text{noise}}^t = 1.4826 \cdot \text{mad}(\mathbf{y} - \mathbf{W}(\boldsymbol{\theta}^{t-1})\mathbf{x}^{t-1}, \mathbf{B}^{t-1})$

¹www.cs.fau.de/research/software/multi-frame-super-resolution-toolbox

Algorithm 1 Confidence-aware Levenberg-Marquardt optimization

- 1: Initialize $\mathbf{x}^0, \boldsymbol{\theta}^0$ and $t = 1$
 - 2: **while** Convergence criterion not fulfilled **do**
 - 3: Get \mathbf{B}^t and \mathbf{A}^t from $(\mathbf{x}^{t-1}, \boldsymbol{\theta}^{t-1})$ according to (9) and (10)
 - 4: Estimate μ^t from (12) with search range $[\log \mu_l; \log \mu_u]$
 - 5: Estimate $\Delta \mathbf{x}$ and $\Delta \boldsymbol{\theta}$ with $\mu = \mu^t$ according to (11)
 - 6: Update $\mathbf{x}^t = \mathbf{x}^{t-1} + \Delta \mathbf{x}$ and $\boldsymbol{\theta}^t = \boldsymbol{\theta}^{t-1} + \Delta \boldsymbol{\theta}$
 - 7: Set $t \leftarrow t + 1$ and proceed with next iteration
 - 8: **end while**
-

and $\sigma_{\text{prior}}^t = \text{mad}(\mathbf{S}\mathbf{x}^{t-1}, \mathbf{A}^{t-1})$ are determined via the weighted median absolute deviation (MAD) rule under the weights \mathbf{A}^{t-1} and \mathbf{B}^{t-1} as derived in [9].

2.3. Levenberg-Marquardt Iterations

Related SR algorithms formulated via least-squares estimation [16, 17] employ Gauss-Newton iterations to iteratively minimize (5). However, the convergence of a Gauss-Newton scheme relies on an initial solution close to the desired global minimum, which is hard to guarantee under practical conditions, e.g. if the initial estimate is affected by outliers. To alleviate this issue, we propose Levenberg-Marquardt iterations, which combines the Gauss-Newton method with gradient descent in our method summarized in Algorithm 1. Instead of solving (6), we compute the updates $\Delta \boldsymbol{\theta}$ and $\Delta \mathbf{x}$ in our Levenberg-Marquardt iteration scheme according to:

$$\begin{pmatrix} \Delta \boldsymbol{\theta} \\ \Delta \mathbf{x} \end{pmatrix} = \left[(\mathbf{P}^t)^\top \mathbf{P}^t + \mu \cdot \text{diag}((\mathbf{P}^t)^\top \mathbf{P}^t) \right]^{-1} (\mathbf{P}^t)^\top \mathbf{f}^t, \quad (11)$$

where μ is the damping factor to control the contributions of gradient descent ($\mu \gg 0$) and Gauss-Newton ($\mu = 0$). The system in (11) is solved by the conjugate gradient (CG) method with T_{cg} iterations in our inner optimization loop to avoid a direct inversion of \mathbf{P}^t . In our outer optimization loop, we use T_{lm} Levenberg-Marquardt iterations to minimize (5).

The performance of the Levenberg-Marquardt iterations in (11) is highly depended on the damping parameter μ . The common damping approach adaptively selects μ from only two candidates per iteration [19], which may lead to inadequate adaptations between gradient descent and Gauss-Newton. In contrast to this approach, we propose to select the damping parameter adaptively per iteration as μ^t to minimize the confidence weighted residual error $\mathbf{r}(\mathbf{x}, \boldsymbol{\theta})$. This leads to the one-dimensional optimization problem:

$$\mu^t = \arg \min_{\mu} \left\| (\mathbf{B}^t)^{\frac{1}{2}} [\mathbf{y} - \mathbf{W}(\boldsymbol{\theta}(\mu)) \mathbf{x}(\mu)] \right\|_2^2, \quad (12)$$

where $\mathbf{x}(\mu)$ and $\boldsymbol{\theta}(\mu)$ are the HR image and the motion parameters obtained from (11) with damping factor μ . To make this parameter selection tractable, we approximate (12) by a one-dimensional search with T_{μ} iterations over the logarithmic scaled range $[\log \mu_l; \log \mu_u]$ in our algorithm. Subsequently, we solve (11) with the optimal parameter $\mu = \mu^t$.

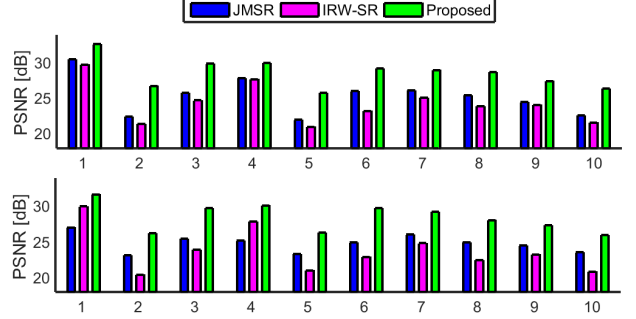


Fig. 2: Mean PSNR over ten simulated datasets with ten randomly generated image sequences per dataset. We compared SR under inaccurate motion estimation without outliers (top row) and with outliers due to invalid pixels (bottom row).

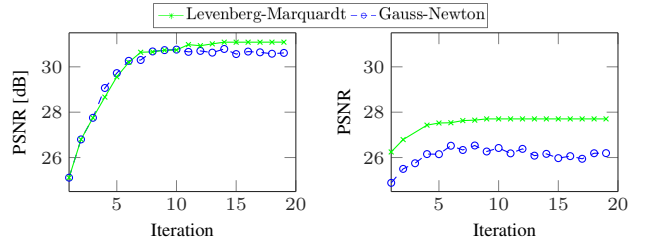


Fig. 4: PSNR over the iterations of our algorithm using Gauss-Newton and the proposed Levenberg-Marquardt iterations. Left: Iterations without outliers. Right: Iterations in the presence of outliers due to invalid pixels.

3. EXPERIMENTS AND RESULTS

This section studies the performance of the proposed method on real and simulated data. We compared our method to the confidence-aware two-stage algorithm based on iteratively re-weighted minimization (IRWSR) [9] and the joint motion estimation and SR using Gauss-Newton iterations (JMSR) [16]. Throughout all experiments, we set the WBTV parameters to $P = 2$, $\alpha_0 = 0.5$ and $p = 0.5$. For iterative minimization, we chose $T_{\text{lm}} = 25$, $T_{\text{cg}} = 25$ and $T_{\mu} = 5$ with the damping parameter search range $\log \mu_l = -4$ and $\log \mu_u = 4$. The regularization weights were selected on one training sequence per dataset using a grid search to enable fair comparisons.

For our quantitative evaluation, we generated synthetic datasets from ten HR images taken from the LIVE database [20]. The LR data was simulated from the HR images according to (1) using a decimation factor of 2, a Gaussian PSF with width $6 \cdot \sigma_{\text{PSF}}$ ($\sigma_{\text{PSF}} = 0.5$), and uniformly distributed translations (-2 to 2 px) and rotations (-1° to 1°). Each frame was corrupted by additive, Gaussian noise with standard deviation $\sigma_{\text{noise}} = 0.025$. The fidelity of the SR reconstruction to the ground truth data was assessed by the peak-signal-to-noise ratio (PSNR) in decibels (dB) as well as structural similarity (SSIM). We simulated sequences consisting of $K = 12$



Fig. 3: SR reconstructions provided by the competing algorithms on simulated and real datasets. Top row: Simulated dataset with uncertainties in the initial motion parameters and outliers due to invalid pixels generated by salt-and-pepper noise. Bottom row: *Emily* dataset with outliers due to non-rigid motion across the LR frames related to movements of the head.

frames with two scenarios: 1) The initial motion parameters were corrupted by uniformly distributed errors for the translation (-0.3 to $+0.3$ px) and the rotation (-0.005° to $+0.005^\circ$) to simulate realistic accuracies of motion estimation in a baseline experiment. 2) We corrupted two frames per sequence by salt-and-pepper noise with noise level $\nu = 0.075$ that denotes the fraction of invalid pixels to simulate outliers in the input frames. Fig. 2 depicts the PSNR for ten datasets with ten randomly generated realizations of these experiments per dataset. We observed that on the one hand JMSR compensated uncertainties in the motion parameters but deteriorates in the presence of outliers. On the other hand, IRWSR was robust regarding outliers but could not correct uncertainties in the initial motion parameters. In contrast to these methods, our confidence-aware algorithm compensated both effects. On average, our method outperformed JMSR by 3.2 dB in the absence of outliers and by 3.0 dB in presence of outliers. See Fig. 3 (top row) for a qualitative comparison along with the PSNR and SSIM measures of the HR images. Here, our method corrected motion estimation uncertainties and was insensitive to invalid pixels.

In order to prove the benefit of the proposed Levenberg-Marquardt optimization, we studied the convergence of our algorithm using the proposed iteration scheme in comparison to Gauss-Newton iterations similar to [16]. Fig. 4 depicts this comparison on a simulated dataset generated with motion estimation uncertainties and outliers due to salt-and-pepper noise. In both scenarios, the proposed Levenberg-Marquardt scheme converged within the first ten iterations and outperformed simple Gauss-Newton iterations.

Finally, we evaluated our method on real image sequences

from the MDSP dataset [21]. Fig. 3 (bottom row) shows a comparison of the competing SR approaches on the *Emily* sequence, where the initial motion estimation was performed on the LR frames using the method in [22]. This sequence follows a translational motion in the first part and non-rigid motion related to head movements in the second part. This deviation of the actual motion to the rigid model caused outliers that cannot be compensated by JMSR and resulted in artifacts in the estimated HR image. The IRWSR method was able to compensate these outlier frames but could not refine the initial motion estimation for the remaining frames. Our method combines outlier removal with a refinement of the initial motion parameters resulting in less artifacts.

4. CONCLUSION

In this paper, we introduced a new joint motion estimation and SR algorithm. Unlike related methods, our algorithm is based on a confidence-aware formulation to consider outliers and space variant noise in the image formation. We developed Levenberg-Marquardt optimization that jointly estimates motion parameters, the unknown HR image and model confidence weights. In our experiments on real and simulated data, our algorithm outperformed state-of-the-art two-stage SR reconstruction as well as joint motion estimation and SR by combining the advantages of both approaches. In particular, the proposed algorithm increases robustness regarding inaccurate motion estimation and outlier observations.

In our future work, we aim at extending our model to more general types of motion, e. g. affine transformations.

5. REFERENCES

- [1] William T. Freeman, Egon C. Pasztor, and Owen T. Carmichael, "Learning Low-Level Vision," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 25–47, 2000.
- [2] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang, "Super-Resolution Image Reconstruction: a Technical Overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.
- [3] Michael Elad and Arie Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1646–1658, 1997.
- [4] Patrick Vandewalle, Sabine Süsstrunk, and Martin Vetterli, "A Frequency Domain Approach to Registration of Aliased Images with Application to Super-resolution," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, pp. 1–15, 2006.
- [5] Wenyi Zhao and Harpreet Sawhney, "Is super-resolution with optical flow feasible?," *European Conference on Computer Vision (ECCV)*, pp. 599–613, 2002.
- [6] Assaf Zomet, Alex Rav-Acha, and Shmuel Peleg, "Robust Super-Resolution," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001, pp. 645 – 650.
- [7] Sina Farsiu, M Dirk Robinson, Michael Elad, and Peyman Milanfar, "Fast and Robust Multiframe Super Resolution.," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.
- [8] Michalis Vrigkas, Christophoros Nikou, and Lisimachos P. Kondi, "Robust Maximum a Posteriori Image Super-Resolution," *Journal of Electronic Imaging*, vol. 23, no. 4, pp. 043016–043016, 2014.
- [9] Thomas Köhler, Xiaolin Huang, Frank Schebesch, Andre Aichert, Andreas Maier, and Joachim Hornegger, "Robust Multiframe Super-Resolution Employing Iteratively Re-Weighted Minimization," *IEEE Transactions on Computational Imaging*, vol. 2, no. 1, pp. 42–58, 2016.
- [10] Michel Batz, Andrea Eichenseer, Jurgen Seiler, Markus Jonscher, and Andre Kaup, "Hybrid Super-Resolution Combining Example-Based Single-Image and Interpolation-Based Multi-Image Reconstruction Approaches," in *IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 58–62.
- [11] Russell C. Hardie, Kenneth J. Barnard, and Ernest E. Armstrong, "Joint MAP Registration and High-Resolution Image Estimation using a Sequence of Undersampled Images.," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, 1997.
- [12] Ce Liu and Deqing Sun, "A Bayesian Approach to Adaptive Video Super Resolution," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 209–216.
- [13] Michael E. Tipping and Christopher M. Bishop, "Bayesian Image Super-resolution," in *Advances in Neural Information Processing Systems*. 2003, pp. 1303–1310, MIT Press.
- [14] Lyndsey C. Pickup, David P. Capel, Stephen J. Roberts, and Andrew Zisserman, "Overcoming Registration Uncertainty in Image Super-Resolution: Maximize or Marginalize?," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–15, 2007.
- [15] S Derin Babacan, Rafael Molina, and Aggelos K Katsagelos, "Variational Bayesian Super Resolution.," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 984–999, 2011.
- [16] Yu He, Kim-Hui Yap, Li Chen, and Lap-Pui Chau, "A Nonlinear Least Square Technique for Simultaneous Image Registration and Super-Resolution.," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2830–2841, 2007.
- [17] Xuesong Zhang, Jing Jiang, and Silong Peng, "Commutability of Blur and Affine Warping in Super-Resolution with Application to Joint Estimation of Triple-Coupled Variables.," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1796–808, 2012.
- [18] Kim-Hui Yap, Yu He, Yushuang Tian, and Lap-Pui Chau, "A Nonlinear L1-Norm Approach for Joint Image Registration and Super-Resolution," *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 981–984, 2009.
- [19] Donald W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM Journal on Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.
- [20] H.R. Sheikh and L. Z. Wang, "LIVE Image Quality Assessment Database Release 2," 2016.
- [21] Sina Farsiu, Dirk Robinson, and Peyman Milanfar, "Multi-Dimensional Signal Processing Dataset," 2016.
- [22] Georgios D Evangelidis and Emmanouil Z Psarakis, "Parametric Image Alignment using Enhanced Correlation Coefficient Maximization.," *IEEE TPAMI*, vol. 30, no. 10, pp. 1858–65, Oct. 2008.