Robust Multiframe Super-Resolution Employing Iteratively Re-Weighted Minimization

Thomas Köhler, Xiaolin Huang, *Member, IEEE,* Frank Schebesch, André Aichert, Andreas Maier, *Member, IEEE,* and Joachim Hornegger

***Abstract*—Multiframe super-resolution algorithms reconstruct high-resolution images by exploiting complementary information in multiple low-resolution frames. However, despite their success under ideal conditions, most existing methods rely on simplistic approximations to the physics of image acquisition and show limited robustness in real-world applications. This paper proposes spatially adaptive Bayesian modeling and an iterative algorithm for robust super-resolution imaging. In particular, we introduce a weighted Gaussian observation model to consider space variant noise and weighted bilateral total variation to exploit sparsity of natural images. Based on this model, we develop a majorization-minimization algorithm implemented as iteratively re-weighted minimization. The proposed method simultaneously estimates model parameters and the super-resolved image in an iterative coarse-to-fine scheme. Compared to prior work, our approach combines the benefits of achieving robust and edge preserving image reconstruction with small amount of parameter tuning, while being flexible in terms of motion models, computationally efficient and easy to implement. Our experimental evaluation confirms that our approach outperforms state-of-the-art algorithms under various practical conditions, e. g. inaccurate geometric and photometric registration or invalid measurements.**

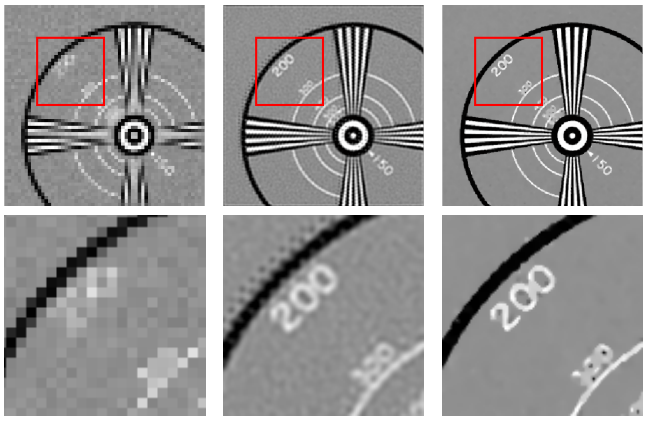
***Index Terms*—Super-resolution, sparse regularization, automatic parameter selection, majorization-minimization**

I. Introduction

THE spatial resolution of an imaging system is a crucial parameter to characterize its ability to capture images with a high level of detail. In many practical applications, this is one of the major quality indicators. Super-resolution is a method to increase the resolution of a camera by an enhancement of its pixel sampling by means of image processing, i. e. without the need to modify the sensor. One important paradigm is the reconstruction of high-resolution images from a set of low-resolution frames taken from the same scene [1]. Most reconstruction-based algorithms exploit subpixel displacements due to camera or object motion to combine complementary information present in multiple frames.

Even if a variety of super-resolution algorithms have been proposed over the past years for monochromatic [2]–[4] or color images [5], [6], their success in practical applications is often limited. In addition to theoretical limitations explained by the perturbation theory of linear systems [7], there are several practical constraints, which make super-resolution difficult. 1) If the subpixel motion is unknown, super-resolution requires an accurate estimate. This is hard to obtain using standard motion estimation techniques, especially if non-rigid motion, independently moving objects or occlusion must be taken into account. Some approaches are also limited to simplistic parametric models, e. g. globally rigid or affine transformations, which are inappropriate for many applications. 2) Reconstruction-based approaches usually approximate the physics of image acquisition with a simplified imaging model and are sensitive to deviations between assumed and true model parameters. Some examples for aspects, which are rarely included in such models are space variant noise, image compression, camera white-balancing or invalid pixels. 3) As super-resolution is known to be an ill-posed problem, many approaches make use of regularization techniques derived from prior knowledge regarding the appearance of images. This has a crucial impact on the performance of super-resolution but many commonly used general-purpose priors are inadequate to model natural images. 4) Algorithms often rely on user-defined hyperparameters, e. g. regularization parameters, that must be selected a priori. However, parameter selection is often cumbersome and limits the robustness of super-resolution if not done appropriately.

In real-world applications, super-resolution needs to build on robust estimation in order to achieve reliable results. The influence of robustness is demonstrated in Fig. 1, where the resolution chart in Fig. 1a is super-resolved in the presence of



(a) Original (b) Algorithm in [2] (c) Ours

Fig. 1: Super-resolution reconstruction (5× magnification) provided by our robust algorithm compared to a non-robust model using the algorithm in [2].

The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German National Science Foundation (DFG) in the framework of the excellence initiative.

T. Köhler, X. Huang, F. Schebesch, A. Aichert, A. Maier and J. Hornegger are with the Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany. (e-mail: {thomas.koehler, xiaolin.huang, frank.schebesch, andre.aichert, andreas.maier, joachim.hornegger}@fau.de)

T. Köhler, A. Maier and J. Hornegger are with the Graduate School in Advanced Optical Technologies (SAOT), 91058 Erlangen, Germany.

inaccurate motion estimation. Fig. 1b depicts the super-resolved image obtained by a model that can be considered non-robust, e. g. using the maximum a-posteriori (MAP) formulation in [2], and is severely affected by the motion estimation uncertainty. Our work aims at robust reconstruction as depicted in Fig. 1c.

A. Related Work

A large class of super-resolution algorithms is derived from a Bayesian perspective due to the simplicity and flexibility to integrate prior knowledge and different imaging models. Here, the design of robust algorithms is focused on two aspects.

1) Robust Observation Models and Optimization: One primary goal is the development of robust reconstruction algorithms based on outlier-insensitive observation models. For this purpose, a common approach is to use outlier detection techniques prior to super-resolution reconstruction. These are typically designed for specific types of outliers, e. g. in motion estimation [8] or saturated pixels [9]. As a more general approach, Zomet et al. [3] have proposed a robust gradient descent scheme using outlier-insensitive update equations. Similarly, Farsiu et al. [4] have introduced MAP estimation based on the L_1 norm as robust measure, which is statistically optimal in case of Laplacian noise and tolerate model outliers. However, this approach is limited by the fact that inliers and outliers are uniformly weighted, which is not necessarily optimal to describe inliers. For this purpose, re-descending M-estimators [10], [11] or hybrid error norms [12], [13] have been investigated. With a similar motivation, pixel-wise weighted least-square optimization [14]–[16] has been proposed. One common issue of these methods is that they require user supervision to tune parameters or employ ad-hoc methods to select them prior to super-resolution. In the work of He and Kondi [17], an algorithm accounting for different noise levels and adaptive regularization weight selection has been developed. A similar method has been introduced by Vrigkas et al. [18], which uses an outlier-insensitive M-estimator. However, these derivations are based on Tikhonov regularization limiting edge reconstruction as discussed below.

A complementary trend to the aforementioned techniques is the use of joint methods to estimate model parameters simultaneously to super-resolution, e. g. to improve the initial motion estimation [19] or to estimate an unknown blur kernel [20], [21]. A different approach is based on a variational Bayesian formulation [22]. Even if these approaches are able to improve accuracy of super-resolution and the utilized model parameters substantially, they are either limited to simplified models, computationally demanding or both. Typically such methods are based on non-robust observation models to make joint optimization or variational formulations tractable.

2) Regularization techniques: As super-resolution is an ill-posed problem, another key aspect to achieve robustness is the image prior that is used to regularize the image reconstruction. Tikhonov regularization based on the L_2 norm of the image gradient or the curvature [2] is one of the most commonly used techniques to alleviate the ill-posedness. While this often leads to efficient algorithms as the prior can be treated analytically, one inherent limitation of Tikhonov regularization is the ability of edge reconstruction since discontinuities are over-penalized. Alternatively, priors that exploit piecewise smoothness of natural images [23] or sparse priors such as total variation (TV) [24] or bilateral total variation (BTV) [4] have been introduced to handle this issue. Yuan et al. [25] and Li et al. [26] have presented spatially adaptive versions of these priors to further enhance edge reconstruction. However, the successes of these methods is highly dependent on additional feature extraction algorithms. Regularization of ill-posed problems has also been widely investigated for compressed sensing. Here, sparse regularization is based on iterated L_1 norm optimization that has been studied for signal recovery [27], [28]. It has been demonstrated that this approach leads to sparser solutions compared to the conventional TV priors making it attractive for regularization in super-resolution. Other fields, where this concept has already been successfully applied include denoising, blind deconvolution or range imaging [29], [30].

B. Contribution and Outline

In view of these findings, we propose a novel super-resolution method to simultaneously address robustness in the observation and the prior model as well as automatic parameter selection. We introduce an optimization scheme as extension of [15] that is flexible in terms of motion models and does not rely on additional feature extraction as used, e. g. in [13], [25], [26]. In summary, our contributions are as follows:

- We formulate robust super-resolution from a Bayesian perspective using a weighted Gaussian distribution to consider mixed noise described by inliers and outliers.
- We propose weighted BTV as a novel image prior for sparse and edge preserving regularization.
- We propose iterative noise and regularization parameter estimation to avoid manual parameter tuning.
- We develop an iteratively re-weighted minimization algorithm implemented in a coarse-to-fine scheme to solve the underlying optimization problem efficiently.

To facilitate reproducible research, Matlab code of our method is public available on our web page[1].

The remainder of this paper is organized as follows. Section II formulates the underlying mathematical framework for super-resolution used in this work. Section III introduces the proposed numerical optimization algorithm formulated as iteratively re-weighted minimization scheme. Section IV gives an in-depth theoretical analysis of our algorithm in terms of convergence. Section V presents a comprehensive quantitative analysis, a comparison to related methods and experimental evaluations on real data. Finally, Section VI summarizes our work.

II. Problem Formulation

A. Image Formation Model

Our method is based on a mathematical model that describes the physical process of obtaining low-resolution frames from a high-resolution image. We employ a generative image model [2] that considers subpixel displacements between low-resolution frames as well as linear, space invariant (LSI) blur due to

[1] www5.cs.fau.de/research/software/multi-frame-super-resolution-toolbox/

the camera point spread function (PSF) and additive noise. A single frame $\boldsymbol{y}^{(k)}$ of size $M = M_1 \cdot M_2$ pixels acquired at time step k out of a sequence $\boldsymbol{y}^{(1)}, \dots, \boldsymbol{y}^{(K)}$ is related to the unknown high-resolution image $\boldsymbol{x}$ of size $N = N_1 \cdot N_2$ pixels in matrix-vector notation according to:

$$\boldsymbol{y}^{(k)} = \boldsymbol{D}\boldsymbol{H}\boldsymbol{M}^{(k)}\boldsymbol{x} + \boldsymbol{\epsilon}^{(k)} = \boldsymbol{W}^{(k)}\boldsymbol{x} + \boldsymbol{\epsilon}^{(k)}, \qquad (1)$$

where $\boldsymbol{D}$ models subsampling, $\boldsymbol{H}$ models LSI blur and $\boldsymbol{M}^{(k)}$ encodes subpixel motion of $\boldsymbol{y}^{(k)}$ with respect to the reference coordinate frame defined by $\boldsymbol{x}$. $\boldsymbol{\epsilon}^{(k)}$ models additive observation noise. For convenience, the different transformations are combined to the system matrix $\boldsymbol{W}^{(k)} \in \mathbb{R}^{M \times N}$ to describe the formation of low-resolution frames from a high-resolution image [23]. The joint model for all frames is given by:

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{\epsilon}, \qquad (2)$$

where $\boldsymbol{y} = (\boldsymbol{y}^{(1)\top}, \dots, \boldsymbol{y}^{(K)\top})^\top$, $\boldsymbol{W} = (\boldsymbol{W}^{(1)\top}, \dots, \boldsymbol{W}^{(K)\top})^\top$ and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}^{(1)\top}, \dots, \boldsymbol{\epsilon}^{(K)\top})^\top$.

B. Observation Model

The proposed super-resolution method is derived from a Bayesian perspective. For this purpose, the sequence of low-resolution frames and the unknown high-resolution image are modeled as random variables. Assuming statistically independent observations, the posterior of the high-resolution image $\boldsymbol{x}$ given the low-resolution frames $\boldsymbol{y}^{(1)}, \dots, \boldsymbol{y}^{(K)}$ is:

$$p(\boldsymbol{x} \,|\, \boldsymbol{y}^{(1)}, \dots, \boldsymbol{y}^{(K)}) \propto p(\boldsymbol{x}) \cdot \prod_{k=1}^{K} p(\boldsymbol{y}^{(k)} \,|\, \boldsymbol{x}), \qquad (3)$$

where $p(\boldsymbol{x})$ describes the prior probability for an image $\boldsymbol{x}$ and $p(\boldsymbol{y}^{(k)} \,|\, \boldsymbol{x})$ describes the conditional probability of observing $\boldsymbol{y}^{(k)}$ from $\boldsymbol{x}$ under the image formation model in (2).

The distribution $p(\boldsymbol{y}^{(k)} \,|\, \boldsymbol{x})$ depends on the underlying observation noise $\boldsymbol{\epsilon}^{(k)}$ for the low-resolution frames. A normal distribution can be used assuming that $\boldsymbol{\epsilon}^{(k)}$ represents additive, white Gaussian noise (AWGN) as done in most super-resolution algorithms [2], [17], [19]. Alternatively, in order to be robust to outliers, e. g. due to invalid pixels or the uncertainty of motion estimation, robust methods based on the Laplacian distribution [4] or M-estimators [18] can be employed. However, with these models the distribution of $\boldsymbol{\epsilon}^{(k)}$ is implicitly assumed to be space invariant and may not be optimal for all observations. In contrast to these space invariant formulations, we follow the finding that image noise in real-world applications typically follows mixed distributions [13], [31]. For this purpose, we propose a spatially weighted Gaussian distribution $\mathcal{N}(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x} \,|\, 0, \sigma_{\text{noise}}, \boldsymbol{\beta})$ defined via the confidence map $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{KM})^\top$ as:

$$p(\boldsymbol{y} \,|\, \boldsymbol{x}, \boldsymbol{\beta}) \propto \exp\left\{ -\frac{(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x})^\top \boldsymbol{B}\,(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x})}{2\sigma_{\text{noise}}^2} \right\}, \qquad (4)$$

with scale parameter σ_{noise}, where $\boldsymbol{B} = \operatorname{diag}(\beta_1, \dots, \beta_{KM})$ is a diagonal matrix assembled from the confidence weights. The influences of the individual observations are controlled by the weights $\beta_m \in [0, 1]$ to model their confidence. Here, $\beta_m = 0$ indicates an outlier under the weighted Gaussian distribution while $\beta_m = 1$ indicates an inlier. Assuming independent observations, (4) can also be written as the Gaussian distribution with spatially varying standard deviation:

$$p(\boldsymbol{y} \,|\, \boldsymbol{x}, \boldsymbol{\beta}) \propto \prod_{m=1}^{KM} \exp\left\{ -\frac{1}{2\sigma_m^2} \Big[\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x} \Big]_m^2 \right\}, \qquad (5)$$

where σ_m is the standard deviation associated with the m-th observation y_m and $[\boldsymbol{z}]_m$ is the m-th element of the vector $\boldsymbol{z}$. For $\sigma_m = \sigma_{\text{noise}}/\sqrt{\beta_m}$ with $\beta_m \neq 0$, (4) and (5) are equivalent and the confidence weights locally define the noise standard deviation. Our distribution in (4) shares some similarities with related outlier detection techniques defined by means of confidence weighting [8]. However, unlike outlier detection that determines confidence weights on low-resolution data using ad-hoc methods, our weights are treated as latent variables and are inferred jointly with the super-resolved image.

C. Image Prior Model

The prior $p(\boldsymbol{x})$ describes the appearance of $\boldsymbol{x}$ in a statistical way. Despite priors introduced for specific applications, e. g. recognition-based models [32], a wide class of algorithms exploits sparsity of natural images in a transform domain. Here, the TV prior has been successfully employed for denoising and image restoration due to its ability of preserving sharp edges. Our approach is based on BTV [4] as generalization of the TV prior. This prior is given in the exponential form:

$$p(\boldsymbol{x}) \propto \exp\big\{ -\lambda R(\boldsymbol{x}) \big\}, \qquad (6)$$

where $\lambda \geq 0$ is a hyperparameter referred to as regularization weight and the BTV regularization term is given by:

$$R(\boldsymbol{x}) = \sum_{m=-P}^{P} \sum_{n=-P}^{P} \alpha_0^{|m|+|n|} \left|\left| \boldsymbol{x} - \boldsymbol{S}_{\text{v}}^m \boldsymbol{S}_{\text{h}}^n \boldsymbol{x} \right|\right|_1, \qquad (7)$$

with $P \geq 1$ and $\alpha_0 \in]0, 1]$. This term exploits image derivatives in a $(2P+1) \times (2P+1)$ window using $\boldsymbol{S}_{\text{v}}^m \in \mathbb{R}^{N \times N}$ and $\boldsymbol{S}_{\text{h}}^n \in \mathbb{R}^{N \times N}$, which denote a vertical and horizontal shift of $\boldsymbol{x}$ by m and n pixels, respectively. For convenience, we reformulate (7) to:

$$R(\boldsymbol{x}) = \sum_{m=-P}^{P} \sum_{n=-P}^{P} \left|\left| \boldsymbol{S}^{m,n} \boldsymbol{x} \right|\right|_1 = \left|\left| \boldsymbol{S}\boldsymbol{x} \right|\right|_1, \qquad (8)$$

where the matrix $\boldsymbol{S} \in \mathbb{R}^{N' \times N}$ with $N' = (2P+1)^2 N$ defines a linear sparsifying transform and is assembled as:

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}^{-P,-P} & \boldsymbol{S}^{-P+1,-P} & \dots & \boldsymbol{S}^{P,P} \end{pmatrix}^\top, \qquad (9)$$

where $\boldsymbol{S}^{m,n} = \alpha_0^{|m|+|n|} (\boldsymbol{I}_{N \times N} - \boldsymbol{S}_{\text{v}}^m \boldsymbol{S}_{\text{h}}^n)$.

Even if (7) is defined via a robust L_1 norm assuming a Laplacian distribution in the transform domain, the properties of BTV in terms of edge preservation are limited. In particular, this prior is not spatially adaptive and discontinuities are penalized proportional to the magnitude of $\boldsymbol{S}\boldsymbol{x}$. For this reason, the use of heavier-tailed distributions than the Laplacian distribution has been proposed for image restoration [29], [30] and compressed sensing [27], [28]. The motivation of these techniques is to better exploit sparsity of natural images. In the spirit of these

findings and similarly to the observation model, we generalize (7) to the spatially adaptive prior:

$$p(\boldsymbol{x} \,|\, \boldsymbol{\alpha}) \propto \exp\left\{-\lambda \frac{||\boldsymbol{\alpha} \odot \boldsymbol{S}\boldsymbol{x}||_1}{\sigma_{\text{prior}}}\right\}, \tag{10}$$

which follows a zero-mean weighted Laplacian distribution $L(\boldsymbol{S}\boldsymbol{x} \,|\, 0, \sigma_{\text{prior}}, \boldsymbol{\alpha})$ with scale parameter σ_{prior} and spatially varying weights $\boldsymbol{\alpha}$. The regularization term associated with this prior referred to as weighted BTV (WBTV) is given by:

$$R(\boldsymbol{x} \,|\, \boldsymbol{\alpha}) = ||\boldsymbol{\alpha} \odot (\boldsymbol{S}\boldsymbol{x})||_1 = ||\boldsymbol{A}\boldsymbol{S}\boldsymbol{x}||_1, \tag{11}$$

where $\boldsymbol{A} = \operatorname{diag}(\alpha_1, \ldots, \boldsymbol{\alpha}_{N'})$ is assembled as a diagonal matrix from $\boldsymbol{\alpha}$. The continuous weights $\alpha_i \in [0, 1]$ allows us to locally adapt $p(\boldsymbol{x} \,|\, \boldsymbol{\alpha})$ in order to enhance the reconstruction of discontinuities. These weights need to be selected such that $\alpha_i = 0$ for discontinuities to lower the influence of regularization and $\alpha_i = 1$ to increase its influence in order to perform smoothing in homogeneous regions. This model is similar to the locally adaptive BTV proposed by Li et al. [26]. However, our weights are handled as latent variables and are inferred jointly with the super-resolved image, while [26] determines the weights in a preprocessing step.

D. Bayesian Estimation

Our goal is to reconstruct a high-resolution image that best explains the set of low-resolution frames. More formally, this is done using MAP estimation derived from (3) as:

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \left\{ p(\boldsymbol{x} \,|\, \boldsymbol{\alpha}) \cdot \prod_{k=1}^{K} p(\boldsymbol{y}^{(k)} \,|\, \boldsymbol{x}, \boldsymbol{\beta}^{(k)}) \right\}. \tag{12}$$

Since we do not know the weights $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ a priori, we need to treat them as latent variables. Let us first review two approaches to infer the weights that are related to our method.

One approach is to solve for $\boldsymbol{x}$ by marginalization over the unknown weights. Under the assumption that $\boldsymbol{\beta}$ is statistically independent of both $\boldsymbol{x}$ and $\boldsymbol{\alpha}$, this can be done according to:

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \int_{\mathbb{R}^{KM}} \int_{\mathbb{R}^{N'}} p(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta} \,|\, \boldsymbol{x}) p(\boldsymbol{x} \,|\, \boldsymbol{\alpha})\, d\boldsymbol{\alpha}\, d\boldsymbol{\beta}, \tag{13}$$

where the weights are marginalized out of the problem. In general, this requires integration over $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathbb{R}^{N'+KM}$, which is computationally prohibitive. If one employs simplified priors $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\beta})$ to model the weights, i. e. conjugate priors to the data likelihood distribution, this marginalization can be treated in closed-form. However, the main limitation of this formulation is that conjugate priors are too simplistic to model the statistical appearance of the weights and rely on additional hyperparameters that need to be chosen appropriately.

Another approach is to solve for $\boldsymbol{x}$ using joint MAP estimation for the confidence weights and the super-resolved image [33]. This leads to the optimization problem:

$$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \arg\max_{\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}} \left\{ p(\boldsymbol{x} \,|\, \boldsymbol{\alpha}) p(\boldsymbol{y} \,|\, \boldsymbol{x}, \boldsymbol{\beta}) p(\boldsymbol{\alpha}) p(\boldsymbol{\beta}) \right\}, \tag{14}$$

that can be solved by alternating minimization for all coupled variables. Similar to marginalization, this requires modeling of prior distributions $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\beta})$. However, the joint estimation is only computationally tractable for simplistic priors.

III. Robust Super-Resolution Algorithm

Our approach is based on MAP estimation according to (12) while simultaneously estimating the latent variables $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. In statistical parameter estimation, this class of optimization problems is commonly treated by means of expectation maximization (EM) [34]. This section introduces a robust iterative method, which falls into the class of majorization-minimization (MM) methods as generalization of EM [35]. The main merit of the MM formulation over the closely related Bayesian marginalization approach as well as the joint MAP approach is that our method does not require a pure statistical modeling of the latent weights. In particular, the prior distributions $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\beta})$ are replaced by a weighting scheme that acts as a surrogate for the priors and enables an iterative and closed-form estimation of the weights. Our formulation yields a computationally efficient iteratively re-weighted minimization scheme [27], [36] consisting of the following steps at each iteration $t \geq 1$:

1) Let $\boldsymbol{x}^{t-1}$ be the super-resolved image reconstructed at iteration $t-1$. Then, refined estimates for the weights $\boldsymbol{\alpha}^t$ and $\boldsymbol{\beta}^t$ are determined in closed form by:
$$\begin{aligned} \boldsymbol{\alpha}^t &:= \alpha\big(\boldsymbol{x}^{t-1} \,|\, \sigma^t_{\text{prior}}\big) \\ \boldsymbol{\beta}^t &:= \beta\big(\boldsymbol{x}^{t-1} \,|\, \boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(K)}, \sigma^t_{\text{noise}}\big), \end{aligned} \tag{15}$$
where $\alpha : \mathbb{R}^{N'} \to \mathbb{R}_0^{+N'}$ and $\beta : \mathbb{R}^{KM} \to \mathbb{R}_0^{+KM}$ are weighting functions, see Section III-A.
2) In order to avoid manual parameter tuning, the unknown scale parameters σ^t_{noise} and σ^t_{prior} as well as the regularization weight λ are automatically selected per iteration, see Section III-B and Section III-C.
3) Once $\boldsymbol{\alpha}^t$, $\boldsymbol{\beta}^t$ and λ^t are estimated, $\boldsymbol{x}^t$ is obtained by MAP estimation, see Section III-D.

These steps yield a sequence of estimates $\{\boldsymbol{x}^t\}_{t=0}$ and are embedded into a coarse-to-fine scheme to perform iteratively re-weighted minimization efficiently. The following subsections introduce the details of our method.

A. Weight Estimation

Let us now proceed with the derivation of the weighting functions to estimate the weights of our Bayesian model.

1) Observation Weights: To determine the observation confidence weights, we analyze the residual error $r : \mathbb{R}^N \to \mathbb{R}^{KM}$ where $r(\boldsymbol{x} \,|\, \boldsymbol{y}) = \boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}$ measures the deviation between a backprojected estimate $\boldsymbol{x}$ and all observations $\boldsymbol{y}$. Our estimation scheme follows the idea that unreliable observations can be identified by their residuals. Accordingly, these observations should be gradually downweighted over the iterations. For this purpose, our weighting function is defined element-wise as:

$$\beta(\boldsymbol{x} \,|\, \boldsymbol{y}, \sigma_{\text{noise}}) := \Big(\beta_1(\boldsymbol{r} \,|\, \sigma_{\text{noise}}), \ldots, \beta_{KM}(\boldsymbol{r} \,|\, \sigma_{\text{noise}})\Big)^\top, \tag{16}$$

where $\boldsymbol{r} = r(\boldsymbol{x} \,|\, \boldsymbol{y})$ and $\beta_i : \mathbb{R}^{KM} \to \mathbb{R}_0^+$ denotes the function to determine the i-th weight from the decomposition:

$$\beta_i(\boldsymbol{r} \,|\, \sigma_{\text{noise}}) := \beta_{i,\text{bias}}(\boldsymbol{r}) \cdot \beta_{i,\text{local}}(\boldsymbol{r} \,|\, \sigma_{\text{noise}}). \tag{17}$$

In the proposed algorithm, two weighting functions are used to remove global outliers detected by $\beta_{i,\text{bias}}(\boldsymbol{r})$ as well as local

outliers identified by $\beta_{i,\text{local}}(\boldsymbol{r} \,|\, \sigma_{\text{noise}})$. For their definition, let $\boldsymbol{r} = r(\boldsymbol{x}^{t-1} \,|\, \boldsymbol{y})$ be the residual error at iteration $t-1$.

First, we detect global outliers on a per-frame basis by means of bias detection. Practically, this should suppress outlier frames, e. g. individual images with global erroneous motion estimation or images affected by motion blur. Our detection is justified by the assumption that the residual error associated with each low-resolution frame $\boldsymbol{y}^{(k)}$ needs to have zero-mean according to the observation model in (4) and its distribution needs to be symmetric [3]. For this reason, bias detection needs to identify frames violating this assumption by analyzing their residual errors. This is done using the binary weighting function $\beta_{i,\text{bias}} : \mathbb{R}^{KM} \to \{0, 1\}$ given by:

$$\beta_{i,\text{bias}}(\boldsymbol{r}) = \begin{cases} 1 & \text{if } \left|\text{MED}(\boldsymbol{r}^{(k)})\right| \leq r_{\text{max}} \\ 0 & \text{otherwise} \end{cases}, \tag{18}$$

where $\boldsymbol{r}^{(k)}$ denotes the residual error of the k-th frame associated with the i-th observation. For robust detection, we replace the mean by the sample median $\text{MED}(\cdot)$ and classify $\boldsymbol{y}^{(k)}$ as biased, if its absolute median residual error exceeds the maximum deviation r_{max}. We set r_{max} to 2 % of the maximum intensity range, i. e. $r_{\text{max}} = 0.02$ if $\boldsymbol{y}^{(k)}$ is given in $[0, 1]^M$, which generalizes fairly well in real-world scenarios.

In addition, we detect local outliers on a per pixel basis using the weighting function $\beta_{i,\text{local}} : \mathbb{R}^{KM} \to \mathbb{R}_0^+$ defined by:

$$\beta_{i,\text{local}}(\boldsymbol{r} \,|\, \sigma_{\text{noise}}^t) = \begin{cases} 1 & \text{if } |r_i| \leq c\sigma_{\text{noise}}^t \\ \frac{c\sigma_{\text{noise}}^t}{|r_i|} & \text{otherwise} \end{cases}, \tag{19}$$

where σ_{noise}^t denotes an adaptive estimate of the scale parameter of the observation model in (4) to discriminate inliers and outliers with a fixed constant $c > 0$. Assuming a Gaussian distribution for inlier observations, $c = 2$ yields a correct classification of $\approx 95\,\%$ of the true inliers. For these observations, a constant confidence is assigned. Observations that are not explained as inliers are weighted by the inverse of their absolute residual error resulting in a lower confidence.

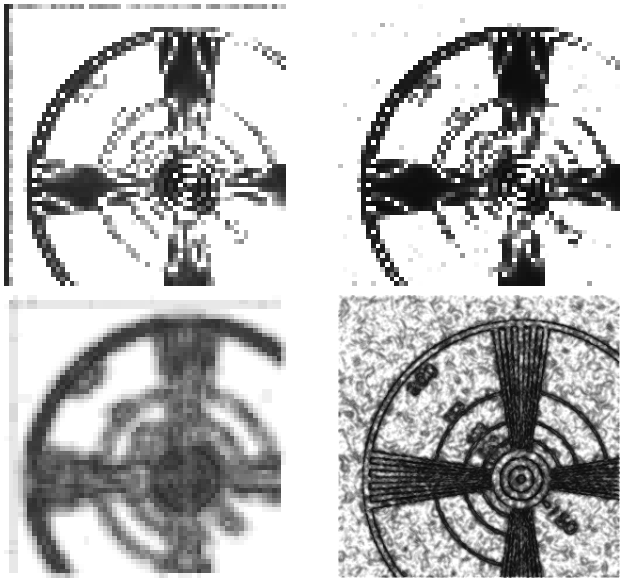
Fig. 2 (top row) depicts the combined observation confidence weights for a single frame mapped to a gray-scale, with insufficient motion estimation for the reconstruction in Fig. 1. In this example, the confidence weights are first slightly over-estimated. Over the course of iteratively re-weighted minimization, the weights are adaptively refined for outlier detection in low-resolution data. Accordingly, the confidence weights describe the observations by mixed noise that can be explained by a superposition of sensor noise and model parameter uncertainties, e. g. in terms of motion estimation.

2) Image Prior Weights: Similarly, the image prior weights are determined under the transform $\boldsymbol{z} = \boldsymbol{S}\boldsymbol{x}$ according to:

$$\alpha(\boldsymbol{x} \,|\, \sigma_{\text{prior}}) := \Big(\alpha_1(\boldsymbol{z} \,|\, \sigma_{\text{prior}}), \ldots, \alpha_{N'}(\boldsymbol{z} \,|\, \sigma_{\text{prior}})\Big)^\top. \tag{20}$$

The weighting function for the i-th element is defined as:

$$\alpha_i(\boldsymbol{z} \,|\, \sigma_{\text{prior}}^t) = \begin{cases} 1 & \text{if } \left|[\boldsymbol{Q}\boldsymbol{z}]_i\right| \leq c\sigma_{\text{prior}}^t \\ p\frac{(c\sigma_{\text{prior}}^t)^{1-p}}{\left|[\boldsymbol{Q}\boldsymbol{z}]_i\right|^{1-p}} & \text{otherwise} \end{cases}, \tag{21}$$



(a) Iteration 1 (b) Iteration 10

Fig. 2: Gray-scale visualization of the adaptive weights estimated by our algorithm in Fig. 1 after the 1st (a) and the 10th iteration (b), respectively (bright regions denote higher weights). Top row: Observation confidence weights $\boldsymbol{\beta}$ for a single frame in the presence of motion estimation uncertainty. Bottom row: Image prior weights $\boldsymbol{\alpha}$ for the sparsity parameter $p = 0.5$.

where $p \in [0, 1]$ is referred to as sparsity parameter, σ_{prior}^t is an adaptive estimate of the scale parameter in (10) and $c > 0$ is a constant to discriminate flat regions and discontinuities that is set to $c = 2$ analogous to the observation weights. In order to reduce the influence of isolated, noisy pixels, this weighting function is applied to a filtered version of the transformed image denoted by $\boldsymbol{Q}\boldsymbol{z}$. In our work, $\boldsymbol{Q}$ is implemented by edge preserving 3×3 median filtering in the transform domain.

This mixed norm weighting function explains images by a mixture of discontinuities and flat regions and assigns spatially adaptive weights to the underlying WBTV model. For small $|[\boldsymbol{Q}\boldsymbol{z}]_i|$ corresponding to flat regions affected by residual noise, a constant weight is used. In order to penalize discontinuities less strictly, a smaller weight that is chosen to be proportional to the inverse of the magnitude of the transformed image is used for large $|[\boldsymbol{Q}\boldsymbol{z}]_i|$. Fig. 2 (bottom row) visualizes this behavior by gray-scale visualizations of the image prior weights over the iterations corresponding to the reconstruction in Fig. 1. Our weights are gradually refined in iteratively re-weighted minimization in order to adapt to the local image structure in our regularization term.

B. Scale Parameter Estimation

In order to avoid an offline selection of the noise level σ_{noise}^t in (19), we propose to automatically adjust it in an optimal way at each iteration. Assuming a uniform distribution for the prior $p(\sigma_{\text{noise}}^t)$, σ_{noise}^t is inferred at iteration t by the maximum likelihood (ML) estimation:

$$\sigma_{\text{noise}}^t = \underset{\sigma_{\text{noise}}}{\arg\max}\, p(\boldsymbol{y} \,|\, \boldsymbol{x}^{t-1}, \boldsymbol{\beta}^{t-1}, \sigma_{\text{noise}}), \tag{22}$$

given $\boldsymbol{x}^{t-1}$ and $\boldsymbol{\beta}^{t-1}$ obtained at the previous iteration. For robust ML estimation, we propose the estimation of σ_{noise}^t from

the median absolute deviation (MAD) [36] of the residual $\boldsymbol{r}^{t-1}$ as a robust measure of statistical dispersion. The MAD is computed from the weighted median, where $\boldsymbol{\beta}^{t-1}$ is used as a weight vector. Hence, the scale parameter is determined by:

$$\begin{aligned}\sigma_{\text{noise}}^t &= \sigma_0 \cdot \text{MAD}\left(\boldsymbol{r}^{t-1} \,|\, \boldsymbol{\beta}^{t-1}\right) \\ &= \sigma_0 \cdot \underset{i=1,\ldots,KM}{\text{MED}} \left(\left| r_i^{t-1} - \text{MED}\big(\boldsymbol{r}^{t-1} | \boldsymbol{\beta}^{t-1}\big) \right| \,\middle|\, \boldsymbol{\beta}^{t-1} \right),\end{aligned} \tag{23}$$

where we set $\sigma_0 = 1.4826$ assuming a Gaussian distribution for the inlier observations [36]. $\text{MAD}(\boldsymbol{r} \,|\, \boldsymbol{\beta})$ denotes the MAD computed from the residual error $\boldsymbol{r}$ under the confidence weights $\boldsymbol{\beta}$ and $\text{MED}(\boldsymbol{r} \,|\, \boldsymbol{\beta}) := \text{MED}_{i=1,\ldots,KM}(r_i \,|\, \boldsymbol{\beta})$ is the weighted median of the elements r_i under the weights $\boldsymbol{\beta}$.

Similarly, the ML estimate for σ_{prior}^t employed for the image prior weighting function in (21) is obtained from the distribution of $\boldsymbol{S}\boldsymbol{x}^{t-1}$ and the weights $\boldsymbol{\alpha}^{t-1}$. In our robust approach, this is done according to:

$$\sigma_{\text{prior}}^t = \sigma_0 \cdot \text{MAD}\left(\boldsymbol{S}\boldsymbol{x}^{t-1} \,|\, \boldsymbol{\alpha}^{t-1}\right), \tag{24}$$

where we set $\sigma_0 = 1$ for the Laplacian distribution.

C. Hyperparameter Estimation

For image reconstruction, the hyperparameter λ weighting the observation model against the prior needs to be initialized. If this parameter is underestimated, super-resolution is ill-conditioned and affected by residual noise while in case of an overestimate, the super-resolved images get blurred. In general, an optimal λ is unknown and its selection also depends on the confidence weights of our Bayesian model. Common approaches to select λ use, e. g. the discrepancy principle [37], generalized cross validation (GCV) [38] or Bayesian methods [22], [39]. Typically these methods deal with specific forms of the prior $p(\boldsymbol{x})$ or use approximative schemes [39].

We employ a data-driven hyperparameter selection that generalizes fairly well to different forms of $p(\boldsymbol{x})$. Similar to [23], we use a cross validation like procedure that is embedded to our algorithm to estimate λ jointly with the super-resolved image. In our approach, we decompose the observations $\boldsymbol{y}$ into two disjoint subsets, where a fraction of δ_{cv}, $0 < \delta_{\text{cv}} < 1$ is used for parameter training and the remaining observations are used for model evaluation. Given a fixed weight λ, we denote the associated super-resolved image as:

$$\begin{aligned}\boldsymbol{x}(\lambda) &= \arg\max_{\boldsymbol{x}} \left\{ p(\boldsymbol{y} \,|\, \boldsymbol{x}, \boldsymbol{\beta}^t, \boldsymbol{I}_{\delta_{\text{cv}}}) p(\boldsymbol{x} \,|\, \boldsymbol{\alpha}^t, \lambda) \right\} \\ &= \arg\min_{\boldsymbol{x}} \left\{ \mathcal{L}_{\text{cv}}(\boldsymbol{x} \,|\, \boldsymbol{I}_{\delta_{\text{cv}}}) + \lambda \big|\big| \boldsymbol{A}^t \boldsymbol{S} \boldsymbol{x} \big|\big|_1 \right\}.\end{aligned} \tag{25}$$

We use the negative log-likelihood to define the data fidelity term $\mathcal{L}_{\text{cv}}(\boldsymbol{x} \,|\, \boldsymbol{I}_{\delta_{\text{cv}}}) \propto -\log p(\boldsymbol{y} \,|\, \boldsymbol{x}, \boldsymbol{\beta}^t, \boldsymbol{I}_{\delta_{\text{cv}}})$, where:

$$\mathcal{L}_{\text{cv}}(\boldsymbol{x} \,|\, \boldsymbol{I}_{\delta_{\text{cv}}}) = \big(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}\big)^\top \boldsymbol{I}_{\delta_{\text{cv}}} \boldsymbol{B}^t \big(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}\big), \tag{26}$$

and $\boldsymbol{I}_{\delta_{\text{cv}}} \in \mathbb{R}^{KM \times KM}$ denotes a diagonal matrix, where $I_{\delta_{\text{cv}},i} = 1$ with probability δ_{cv} to specify the training observations and $I_{\delta_{\text{cv}},i} = 0$ with probability $1 - \delta_{\text{cv}}$. The optimal regularization weight λ^t at iteration t is determined by:

$$\lambda^t = \arg\min_{\lambda} \mathcal{L}_{\text{cv}}(\lambda \,|\, \overline{\boldsymbol{I}}_{\delta_{\text{cv}}}), \tag{27}$$

where the cross validation error associated with λ is given by:

$$\mathcal{L}_{\text{cv}}(\lambda \,|\, \overline{\boldsymbol{I}}_{\delta_{\text{cv}}}) = \big(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}(\lambda)\big)^\top \overline{\boldsymbol{I}}_{\delta_{\text{cv}}} \boldsymbol{B}^t \big(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}(\lambda)\big), \tag{28}$$

and $\overline{\boldsymbol{I}}_{\delta_{\text{cv}}}$ is obtained from $\boldsymbol{I}_{\delta_{\text{cv}}}$ by flipping the diagonal elements to minimize the error on the validation observations.

Notice that the cross validation error in (28) and the image reconstruction according to (25) are interdependent and hence the gradient of $\mathcal{L}_{\text{cv}}(\lambda \,|\, \overline{\boldsymbol{I}}_{\delta_{\text{cv}}})$ is not well-defined. Opposed to gradient-based optimization [23] that might get stuck in a local minimum, we propose an adaptive grid search to select λ^t according to (27) as follows: For $t = 1$, λ^1 is selected as the global minimum of $\mathcal{L}_{\text{cv}}(\lambda \,|\, \overline{\boldsymbol{I}}_{\delta_{\text{cv}}})$ using a grid search over the log-transformed range $[\log \lambda_l, \log \lambda_u]$ of regularization weights. For $t > 1$, we use λ^{t-1} as initial guess that is refined by a local search with the adaptive search range $\log \lambda_{u,l} = \log(\lambda^{t-1}) \pm \frac{1}{t}$. The number of steps is adaptively adjusted for each iteration. For $t = 1$, it is initialized by T_{cv}^1. In the subsequent iterations, it is gradually reduced according to $T_{\text{cv}}^t = \lceil 0.5 \cdot T_{\text{cv}}^{t-1} \rceil$.

D. Image Reconstruction

Given the estimates of the weights $\boldsymbol{\alpha}^t$ and $\boldsymbol{\beta}^t$ as well as the regularization parameter λ^t, we determine $\boldsymbol{x}^t$ according to:

$$\begin{aligned}\boldsymbol{x}^t &= \arg\max_{\boldsymbol{x}} \left\{ p(\boldsymbol{y} \,|\, \boldsymbol{x}, \boldsymbol{\beta}^t) p(\boldsymbol{x} \,|\, \boldsymbol{\alpha}^t) \right\} \\ &= \arg\min_{\boldsymbol{x}} F^t(\boldsymbol{x}),\end{aligned} \tag{29}$$

where the overall energy function at iteration t is:

$$F^t(\boldsymbol{x}) = \big(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}\big)^\top \boldsymbol{B}^t \big(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}\big) + \lambda^t \big|\big| \boldsymbol{A}^t \boldsymbol{S} \boldsymbol{x} \big|\big|_1. \tag{30}$$

We employ Scaled Conjugate Gradients (SCG) iterations [40] for numerical optimization of this convex minimization problem. For $t = 1$, the iterations are initialized by the temporal median of the motion-compensated low-resolution frames to provide an outlier-insensitive initial guess [41]. Subsequently, at each iteration $t > 1$, we update $\boldsymbol{x}^{t-1}$ determined at the previous iteration to obtain the refined estimate $\boldsymbol{x}^t$. For the implementation of SCG, we use a smooth and continuous differentiable approximation of (11) according to:

$$R(\boldsymbol{x} \,|\, \boldsymbol{\alpha}) = \big|\big| \boldsymbol{A}\boldsymbol{S}\boldsymbol{x} \big|\big|_1 \approx \sum_{i=1}^{N'} \alpha_i \sqrt{[\boldsymbol{S}\boldsymbol{x}]_i^2 + \tau}. \tag{31}$$

For small τ ($\tau = 10^{-4}$), this provides a reasonable approximation of the L_1 norm in the WBTV prior. This inner optimization loop to minimize (30) is performed until a maximum number of T_{SCG} iterations or a convergence tolerance is reached.

In order to avoid local minimums in our iterative algorithm, we propose an implementation in a coarse-to-fine scheme. The confidence weights are initialized by $\boldsymbol{\alpha}^0 = \mathbf{1}$ and $\boldsymbol{\beta}^0 = \mathbf{1}$, where $\mathbf{1}$ is an all-one vector. Moreover, we gradually increase the magnification factor by Δs per iteration such that $s^t = s^{t-1} + \Delta s$ until the desired magnification s is reached. In addition to the avoidance of local minimums, this approach also reduces the computational costs, as more iterations for the computational demanding hyperparameter selection are done more efficiently for smaller s. This coarse-to-fine optimization alternates between weight computation,

Algorithm 1 Super-resolution via iteratively re-weighted minimization

Set $\boldsymbol{x}^0$ to the motion-compensated median of $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(K)}$
Set $\boldsymbol{\alpha}^0 = \mathbf{1}$, $\boldsymbol{\beta}^0 = \mathbf{1}$, $s^0 = 0$ and $t = 1$
while (32) not fulfilled and $t \leq T_{\text{max}}$ **do**
  Set magnification factor s^t to $s^t = \min(s^{t-1} + \Delta s, s)$
  Propagate $\boldsymbol{x}^{t-1}$ using magnification factor s^t
  Compute σ^t_{noise} from $\boldsymbol{x}^{t-1}$ and $\boldsymbol{\beta}^{t-1}$ according to (23)
  Compute σ^t_{prior} from $\boldsymbol{x}^{t-1}$ and $\boldsymbol{\alpha}^{t-1}$ according to (24)
  Compute $\boldsymbol{\beta}^t$ from $\boldsymbol{r}^{t-1}$ and σ^t_{noise} according to (17)
  Compute $\boldsymbol{\alpha}^t$ from $\boldsymbol{S}\boldsymbol{x}^{t-1}$ and σ^t_{prior} from (21)
  Select λ^t in $T^t_{\text{cv}} = \lceil 0.5 \cdot T^{t-1}_{\text{cv}} \rceil$ steps according to (27)
  Set $\boldsymbol{x}^t = \boldsymbol{x}^{t-1}$ and $t_{\text{SCG}} = 1$
  while (32) not fulfilled and $t_{\text{SCG}} \leq T_{\text{SCG}}$ **do**
    Update $\boldsymbol{x}^t$ according to (30) using SCG iteration
    Set $t_{\text{SCG}} \leftarrow t_{\text{SCG}} + 1$ and proceed with next iteration
  end while
  Set $t \leftarrow t + 1$ and proceed with next iteration
end while

hyperparameter selection and image reconstruction until a stopping criterion is met or T_{max} iterations are performed. As a stopping criterion we choose the maximum absolute change of the pixel values between $\boldsymbol{x}^t$ and $\boldsymbol{x}^{t-1}$ according to:

$$\max_{i=1,\ldots,N} \left(\left| x_i^{t-1} - x_i^t \right| \right) < \eta, \tag{32}$$

where $\eta > 0$ denotes the termination tolerance. Our overall coarse-to-fine scheme is summarized in Algorithm 1.

IV. Algorithm Analysis

A. Formulation as Majorization-Minimization Algorithm

Our algorithm iteratively solves a sequence of convex minimization problems, which is steered by means of confidence weighting. Interestingly, using our weighting scheme, iteratively re-weighted minimization according to (30) coincides with the solution of the robust and sparse reconstruction problem:

$$\begin{aligned} \hat{\boldsymbol{x}} &= \arg\min_{\boldsymbol{x}} F(\boldsymbol{x}), \\ \text{where } F(\boldsymbol{x}) &= \sum_{i=1}^{KM} h_\delta \left([\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}]_i \right) + \lambda R(\boldsymbol{S}\boldsymbol{x}). \end{aligned} \tag{33}$$

In order to show the relation to our iteration scheme, the data fidelity term is defined by the Huber loss function:

$$h_\delta(z) = \begin{cases} z^2 & \text{if } |z| \leq \delta \\ 2\delta|z| - \delta^2, & \text{otherwise.} \end{cases}, \tag{34}$$

where δ denotes a non-negative threshold. The regularization term is defined via a mixed L_1/L_p norm according to:

$$R(\boldsymbol{z}) = \sum_{i \notin \mathcal{I}(\boldsymbol{z})} |z_i| + \sum_{i \in \mathcal{I}(\boldsymbol{z})} |z_i|^p, \tag{35}$$

where small elements are treated by the L_1 norm ($i \notin \mathcal{I}(\boldsymbol{z})$) and large elements are treated by the L_p norm ($i \in \mathcal{I}(\boldsymbol{z})$), where $0 < p < 1$. Without loss of generality, we define the index set for our analysis as $\mathcal{I}(\boldsymbol{z}) = \{i : z_i \geq 1\}$ since we can use a normalization of $\boldsymbol{z}$ to satisfy this condition.

Using the Huber loss in (34), small residuals are penalized quadratically while large residuals are penalized linearly making this formulation robust regarding outlier observations. Similarly, the mixed L_1/L_p norm with sparsity parameter p in (35) enforces sparsity of the reconstructed images under the transform $\boldsymbol{S}$. These relations to the optimization in (33) are useful to clarify the properties of our algorithm in terms of robustness. As the primary finding, our method is an MM algorithm to the non-convex optimization problem in (33) by reformulation via a sequence of weighted optimization problems. Notice that our algorithm extends related approaches for re-weighted L_2/L_p norm minimization [42].

B. Convergence Analysis

Now, let us investigate the convergence of the proposed iteratively re-weighted minimization scheme. In contrast to our numerical algorithm but to make our analysis tractable, we start from an initial guess $\boldsymbol{x}^0$ and limit ourselves to iteratively re-weighted minimization according to (30), where λ is assumed to be a constant regularization weight and the weights $\boldsymbol{A}^t$ and $\boldsymbol{B}^t$ are computed from (21) and (19) with constant scale parameters over the iterations. Using this iteration scheme, the objective value in (33) converges within a finite number of iterations. The convergence in terms of the objective value $F(\boldsymbol{x})$ is guaranteed by the following theorem.

Theorem 1. *Let $\{\boldsymbol{x}^t\}_{t=0}$ be an iteration sequence obtained by iteratively re-weighted minimization. Then, there exists a strict positive $\underline{\beta}$ such that:*

$$F(\boldsymbol{x}^{t-1}) - F(\boldsymbol{x}^t) \geq \underline{\beta} \left|\left| \boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t \right|\right|_2^2. \tag{36}$$

Proof. The proof is given in the appendix. □

This theorem shows that $F(\boldsymbol{x}^t)$ is monotonically decreasing over the iterations. In addition, $F(\boldsymbol{x})$ is a lower-bounded function. Hence, $F(\boldsymbol{x}^t)$ converges to an extreme value, and so does Algorithm 1. In addition to this theoretical finding, an experimental convergence analysis of our algorithm with adaptive parameter selection is presented in Section V-B.

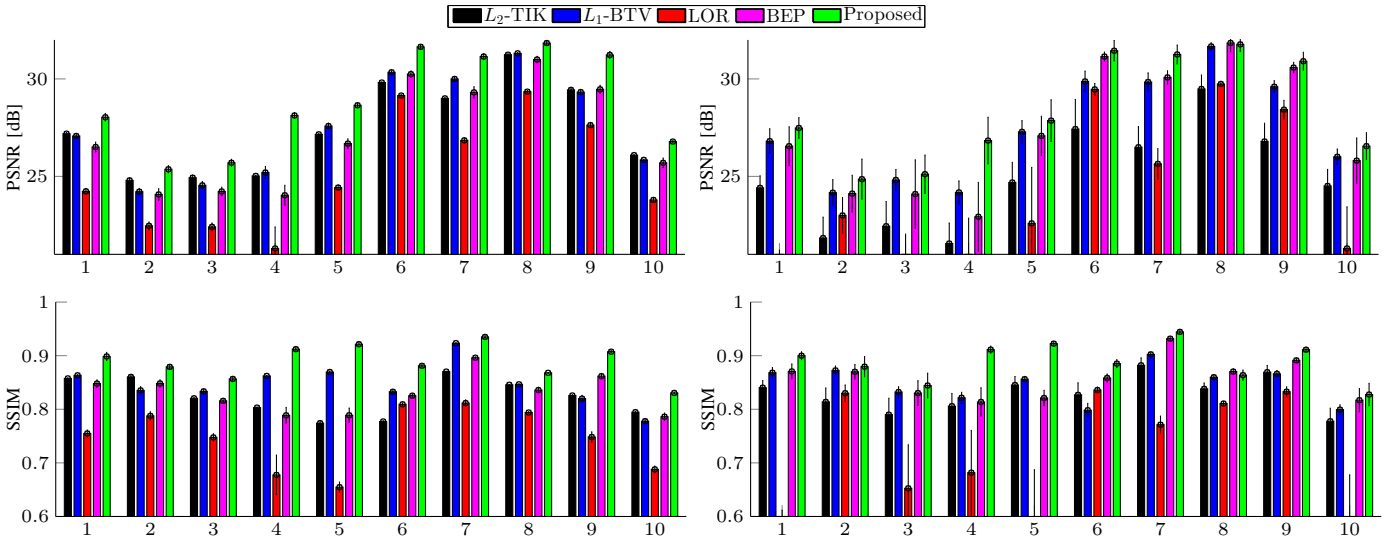
V. Experiments and Results

A. Experimental Setup

This section reports experimental results of our method on simulated and real data. The proposed method was compared to the following state-of-the-art super-resolution algorithms:

- L_2-TIK MAP super-resolution with L_2 norm data fidelity term and Tikhonov regularization [2].
- L_1-BTV The robust L_1 norm minimization with BTV regularization proposed by Farsiu et al. [4].
- LOR The robust Lorentzian M-estimator with Lorentzian prior proposed by Patanavijit and Jitapunkul [10].
- BEP The adaptive algorithm with bilateral edge preserving (BEP) regularization of Zeng and Yang [11].

For fair comparison, all algorithms were implemented using the image formation model presented in Section II-A. In general, this model is parametrizable with different types of motion models ranging from a pure translation to non-rigid motion.



(a) PSNR and SSIM for exact motion estimation

(b) PSNR and SSIM for inaccurate motion estimation

Fig. 3: Comparison of mean ± standard deviation of the PSNR and the SSIM measures over ten simulated datasets with ten randomly generated image sequences per dataset. (a): error measures if exact subpixel motion is used. (b): error measures in case of inaccurate motion estimation.

Throughout our experiments, we limited ourselves to an affine motion model that was employed for all compared algorithms. As all methods are formulated via energy minimization but are independent of the underlying numerical optimization technique, we implemented them by means of SCG iterations. This might improve the convergence and avoids parameter tuning regarding the step size required for fixed step size steepest descent as proposed in the original publications. For L_2-TIK, L_1-BTV, LOR and BEP, the regularization weights were selected based on a grid search on one training sequence per dataset. The additional model parameters of the LOR and BEP approach were determined empirically.

Throughout all experiments, we used the following parameter settings for our method. For the WBTV prior, we set $P = 2$, $\alpha_0 = 0.7$ and $p = 0.5$. For coarse-to-fine optimization, we set $\Delta s = 1$ and $s^1 = 1$. Cross validation was performed with $\delta_{\text{cv}} = 0.95$, $\log \lambda_l = -12$, $\log \lambda_u = 0$ and $T_{\text{cv}}^1 = 20$ steps. In our iterative scheme, we used $T_{\max} = 10$ iterations, $T_{\text{SCG}} = 5$ iterations for SCG in the inner optimization loop and $\eta = 10^{-3}$ for the termination criterion.

B. Simulated Data Experiments

In order to conduct a quantitative evaluation, we generated synthetic data and measured the peak-signal-to-noise ratio (PSNR) in decibels (dB) of super-resolved data $\boldsymbol{x}$ with respect to a ground truth image $\hat{\boldsymbol{x}}$. Additionally, we measured the structural similarity (SSIM) to assess the fidelity of $\hat{\boldsymbol{x}}$ and $\boldsymbol{x}$. We used the reference images from the LIVE database [43] as ground truth given as grayscale images in the intensity range $[0, 1]$. For each dataset, we generated K low-resolution frames simulating a rigid motion. Thus, all frames were displaced by uniform distributed random translations (-3 to $+3$ pixels) and rotations with uniform distributed angles ($-1°$ to $+1°$) relative to the reference coordinate grid. Each frame was blurred by an isotropic Gaussian PSF with kernel size $6 \cdot \sigma_{\text{PSF}}$ low-resolution pixels ($\sigma_{\text{PSF}} = 0.5$) and subsampled according to the magnification factor s. Finally, each frame was affected by a fixed amount of Poisson noise as well as varying levels of Gaussian noise depending on the experiment[2]. All experiments reported in the subsequent paragraphs were repeated ten times for each dataset with randomly generated frames.

1) Effect of Image Noise: We generated ten datasets consisting of image sequences with $K = 8$ frames each from our ground truth images to reconstruct high-resolution images with magnification $s = 2$. Each frame was disturbed by AWGN with noise standard deviation $\sigma_{\text{noise}} = 0.02$. In order to consider the best-case for a baseline experiment, super-resolution was performed based on the exact subpixel motion. The corresponding statistics of the PSNR and SSIM measures over all datasets are summarized in Fig. 3a. In most of the datasets, the proposed method substantially outperformed the competing approaches. On average, compared to L_1-BTV, our approach improved the PSNR and SSIM by 1.2 dB and 0.03, respectively. A qualitative comparison of all methods is depicted in Fig. 4 (middle row) showing that our approach achieved smaller residual noise levels with accurate edge reconstruction. In order to assess the noise robustness of super-resolution reconstruction, the noise standard deviation was gradually increased from $\sigma_{\text{noise}} = 0$ to $\sigma_{\text{noise}} = 0.05$ on this dataset. The averaged error measures over ten realizations of the experiment versus the noise standard deviation are plotted in Fig. 5. In terms of both measures, our method consistently achieved the best reconstructions among the compared algorithms.

2) Effect of Motion Estimation Uncertainty: Next, we conducted experiments to investigate the influence of inaccurate motion estimation on our datasets. For this purpose, the rotation $\boldsymbol{R} \in \mathbb{R}^{2\times 2}$ of two randomly selected frames was corrupted by a small scaling factor c such that $c\boldsymbol{R}$ is a corrupted version of the rotation to simulate a zooming camera. For motion estimation, enhanced correlation coefficient (ECC) optimization [44] with rigid motion was employed. This

[2]Artificial noise was simulated using the Matlab `imnoise` function.

(a) Original (27.1 dB, 0.66)

(b) L_2-TIK [2] (29.9 dB, 0.76) (c) L_1-BTV [4] (30.4 dB, 0.83) (d) LOR [10] (29.2 dB, 0.80) (e) BEP [11] (30.3 dB, 0.82) (f) Proposed (31.7 dB, 0.88)

(g) L_2-TIK [2] (26.6 dB, 0.76) (h) L_1-BTV [4] (29.6 dB, 0.84) (i) LOR [10] (28.8 dB, 0.78) (j) BEP [11] (30.1 dB, 0.82) (k) Proposed (30.4 dB, 0.86)

Fig. 4: Super-resolution on the *Lighthouse* dataset with PSNR and SSIM measures in brackets ($K = 8$ frames, magnification $s = 2$). Top row: original, low-resolution frame. Middle row: super-resolution using exact subpixel motion. Bottom row: super-resolution under inaccurate subpixel motion.

resulted in outliers due to misregistrations of the two frames affected by scaling, while the accuracy for the remaining frames was limited by the uncertainty of ECC optimization. The scaling followed a normal distribution with mean equal to one and standard deviation σ_c. Fig. 3b shows the statistics of the error measures over ten datasets under inaccurate motion estimation for $\sigma_c = 0.05$. Form these results we can observe that the performance of the L_2-TIK method was severely degraded by motion estimation outliers. The proposed method achieved better robustness regarding inaccurate motion estimation and outperformed the competing methods on most of our datasets. A visual comparison is depicted in Fig. 4 (bottom row). Here, the effect of outliers is noticeable by ghosting artifacts and blurring of sharp edges in the outcome of L_2-TIK.

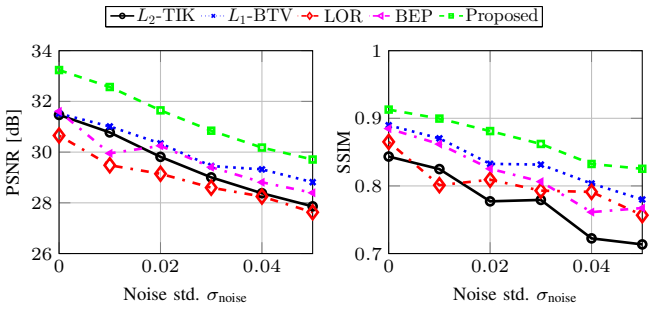


Fig. 5: Mean PSNR and SSIM versus the standard deviation of additive, white Gaussian noise on the *Lighthouse* dataset.

3) Effect of Invalid Pixels: In order to assess the effect of invalid pixels, the low-resolution frames were further degraded by salt-and-pepper noise. The noise level was gradually increased from $\nu = 0$ to $\nu = 0.15$, where ν denotes the fraction of invalid pixels. Super-resolution was performed for $K = 8$ frames and magnification $s = 2$ using the exact subpixel motion. Fig. 6 shows the averaged PSNR and SSIM measures for different amounts of invalid pixels for ten realizations of this experiment on the *Cemetery* dataset. We observe that the L_2-TIK method failed to reconstruct reliable high-resolution images in the presence of invalid pixels, while the methods based on robust observation models were insensitive to invalid pixels. In particular, our approach performed best and was only slightly affected by outliers. See Fig. 7 for a visual comparison.

4) Effect of the Sequence Length: Next, we evaluated the influence of the number of low-resolution frames to the quality of super-resolution. For this analysis, the AWGN level was set to $\sigma_{\text{noise}} = 0.02$, the fraction of invalid pixel was $\nu = 0.01$ and super-resolution was performed using the exact subpixel motion. This experiment considered $s = 3$ to analyze a larger magnification factor as this typically requires longer image sequences to achieve stable reconstructions. The

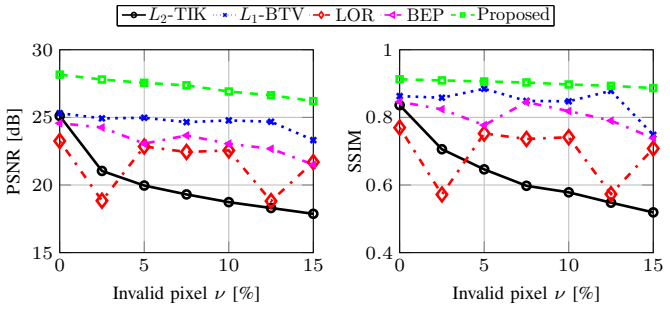


Fig. 6: Mean PSNR and SSIM versus the amount of salt-and-pepper noise (ν) evaluated on the *Cemetery* dataset.

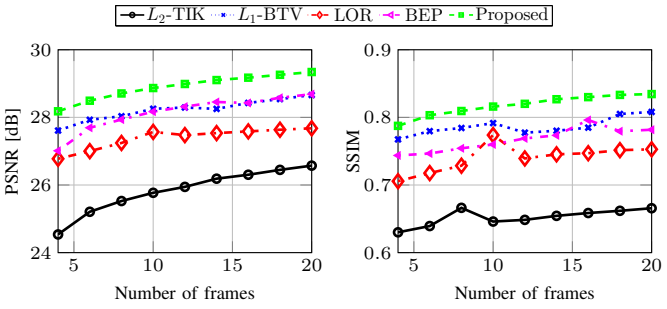


Fig. 8: Effect of the sequence length to the mean PSNR and SSIM on the *Lighthouse* dataset.

(a) Original (13.4 dB, 0.39) (b) L_2-TIK [2] (18.7 dB, 0.57)

(c) L_1-BTV [4] (24.7 dB, 0.84) (d) LOR [10] (23.3 dB, 0.74)

(e) BEP [11] (23.2 dB, 0.81) (f) Proposed (27.0 dB, 0.90)

Fig. 7: Super-resolution reconstructions with PSNR and SSIM measures in presence of salt-and-pepper noise (fraction of invalid pixels: $\nu = 0.10$) using the compared approaches ($K = 8$ frames, magnification $s = 2$).

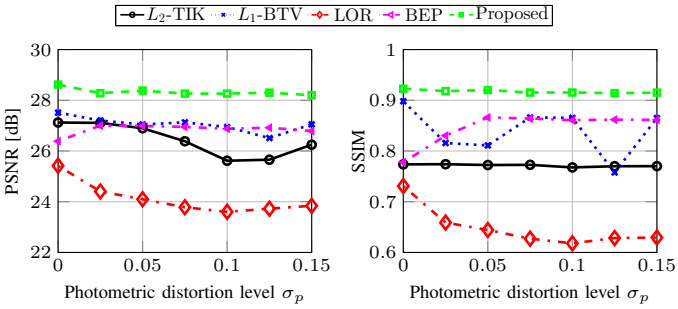
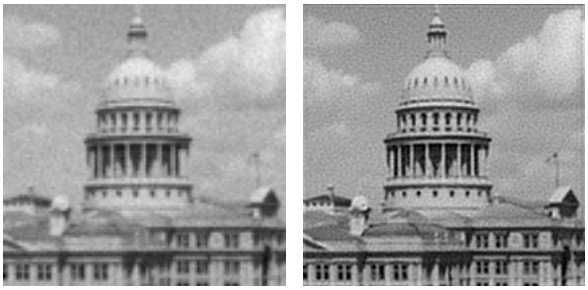


Fig. 9: Mean PSNR and SSIM for different error levels in photometric registration for the *Church-and-capitol* dataset.

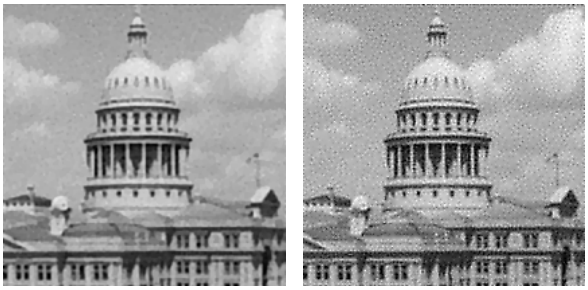
sequence length was varied between $K = 4$ to $K = 20$. Note that for $K < s^2$ the underlying super-resolution problem is underdetermined, which was the case in our experiment. The averaged error measures for ten realizations of this experiment on the *Lighthouse* dataset are depicted in Fig. 8. Super-resolution was more accurate for longer sequences and our method achieved the highest measures among the competing approaches. Notice that this was also the case for an underdetermined reconstruction problem ($K < 9$). Our approach based on $K = 8$ frames was competitive to L_1-BTV and BEP based on $K = 20$ frames and substantially outperformed the L_2-TIK as well as the LOR method.

5) Effect of Photometric Distortions: We also evaluated the effect of varying photometric conditions in low-resolution frames caused, e. g. by time variant lighting conditions. If such effects occur in real-world applications, photometric registration [45] needs to be considered. Outliers in photometric registration were simulated for two randomly selected frames that were corrupted by photometric distortions according to the model $\boldsymbol{z}^{(k)} = \gamma_m^{(k)} \boldsymbol{y}^{(k)} + \gamma_a^{(k)}$, where $\boldsymbol{y}^{(k)}$ is a frame without distortions and $\boldsymbol{z}^{(k)}$ is a corrupted frame. The parameters $\gamma_a^{(k)}$ and $\gamma_m^{(k)}$ were assumed to be uniform distributed in $[-\frac{1}{2}\sigma_p, +\frac{1}{2}\sigma_p]$ and $[1 - \frac{1}{2}\sigma_p, 1 + \frac{1}{2}\sigma_p]$, respectively. Super-resolution was performed with $K = 8$ frames and magnification $s = 2$ using exact subpixel motion. Fig. 9 shows the PSNR and SSIM measures averaged over ten realizations of the experiment on the *Church-and-capitol* dataset for different levels of photometric distortions. The L_2-TIK approach was sensitive to outliers in photometric registration resulting in a bias in the intensity values as indicated by a decreasing of the PSNR. With the exception of the LOR approach, the robust methods were less sensitive to photometric distortions. Note that even for severe distortions, the proposed approach compensated for this effect and achieved the best error measures. A visual comparison is depicted in Fig. 10.

6) Convergence and Parameter Study: Finally, we report results of a convergence and parameter study conducted for our method. For this purpose, we used the *Parrots* dataset and generated $K = 12$ low-resolution frames according to the



(a) Original (22.9 dB, 0.68) (b) L_2-TIK [2] (22.7 dB, 0.75)



(c) L_1-BTV [4] (26.6 dB, 0.85) (d) LOR [10] (23.4 dB, 0.59)

(e) BEP [11] (26.4 dB, 0.85) (f) Proposed (28.1 dB, 0.91)

Fig. 10: Super-resolution reconstructions with PSNR and SSIM measures under inaccurate photometric registration ($\sigma_p = 0.15$) using the compared approaches ($K = 8$ frames, magnification $s = 2$).

magnification factor $s = 3$. In order to analyze the convergence of our iterative scheme, we traced the PSNR and SSIM measures over the iterations. This was performed for different amounts of invalid pixels in low-resolution data. Moreover, we compared different initializations of our algorithm using the outlier-insensitive temporal median of the low-resolution frames as well as a simple bicubic interpolation of the reference frame. The averaged measures for ten realizations of this experiment are depicted in Fig. 11 (top row), which demonstrates that our algorithm typically converges after ≈ 5 iterations for different outlier fractions. The most substantial improvements were achieved in the first two iterations, which is also noticeable by visual comparison in Fig. 11 (bottom row). This finding confirms empirically that the proposed iterative scheme converges to accurate solutions even for a large amount of outlier. Moreover, despite the non-convexity of the underlying energy function solved by our algorithm, we observed that our iterative scheme converges to similar solutions even if the initial guess is different. In terms of the PSNR and SSIM measures, using a simple bicubic interpolation as a rough initial guess for the high-resolution image leads to reconstructions that are competitive to those obtained by the temporal median initialization. However, using the temporal

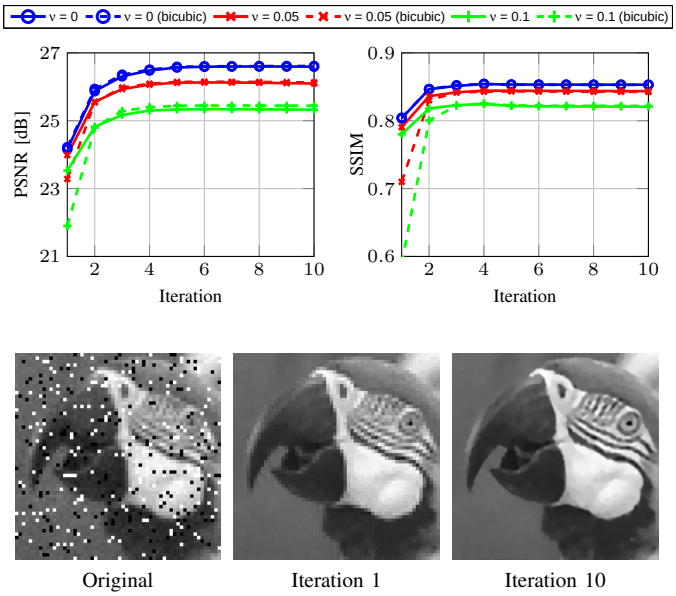


Fig. 11: Convergence analysis for our approach. Top row: mean PSNR and SSIM versus the iteration number for different amounts of invalid pixels using the outlier-insensitive temporal median of the low-resolution frames (solid lines) as well as the bicubic interpolation of a single frame (dashed lines) for initialization. Bottom row: super-resolved images depicted at different iterations for $\nu = 0.10$ using the temporal median as initial guess.

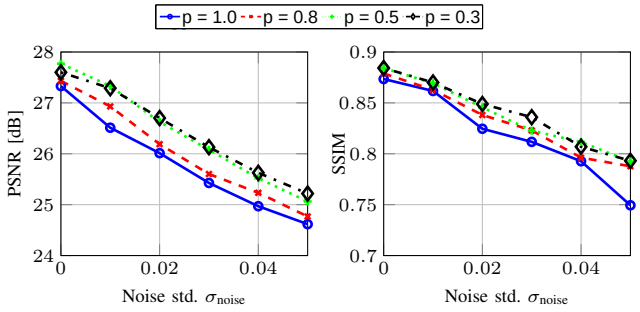
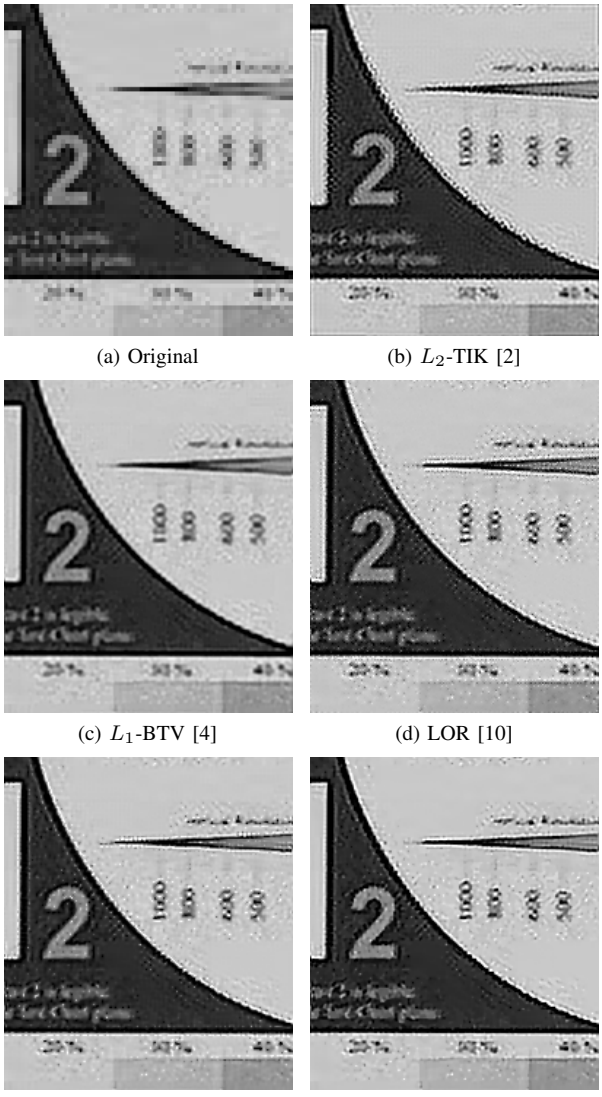


Fig. 12: Influence of the sparsity parameter p of weighted bilateral total variation to the accuracy of super-resolution on the *Parrots* dataset.

median leads to a faster convergence within the first iterations. This confirms the robustness of our approach regarding the initialization used for iteratively re-weighted minimization.

In addition to the convergence, we studied the benefit of our prior weighting scheme in (21) that is parametrized by the sparsity parameter p. We considered $p = 1$ corresponding to the unweighted BTV as well as different versions of WBTV with $p < 1$. For this experiment, the amount of invalid pixels was set to $\nu = 0.01$. Fig. 12 depicts the averaged error measures corresponding to the different p values for varying noise levels σ_{noise}. In case of $p < 1$, we consistently observed higher measures compared to $p = 1$. In particular, this is noticeable for higher noise levels. The major differences between these settings is that the smaller p, the stronger the sparsity in the transform domain is exploited by regularization. A higher degree of sparsity leads to improved edge reconstructions as sharp edges are penalized less severe, which encourages the use of our sparse regularization technique.

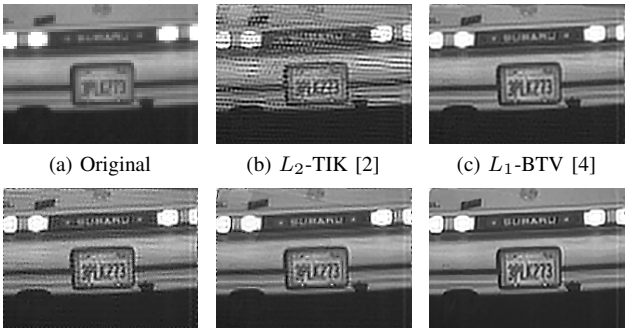


(a) Original (b) L_2-TIK [2]

(c) L_1-BTV [4] (d) LOR [10]

(e) BEP [11] (f) Proposed

Fig. 13: Super-resolution reconstruction on the *Adyoron* sequence ($K = 20$ frames, magnification factor $s = 3$).



(a) Original (b) L_2-TIK [2] (c) L_1-BTV [4]

(d) LOR [10] (e) BEP [11] (f) Proposed

Fig. 14: Super-resolution on the *Car* sequence to reconstruct the license plate ($K = 12$ frames, magnification factor $s = 3$).

C. Real Data Experiments

In addition to our experiments on simulated data, we demonstrate the performance of our method on real images. In the absence of a ground truth, we compared super-resolved images qualitatively and adjusted the model parameters of the competing methods empirically such that they produce the visually most appealing results.

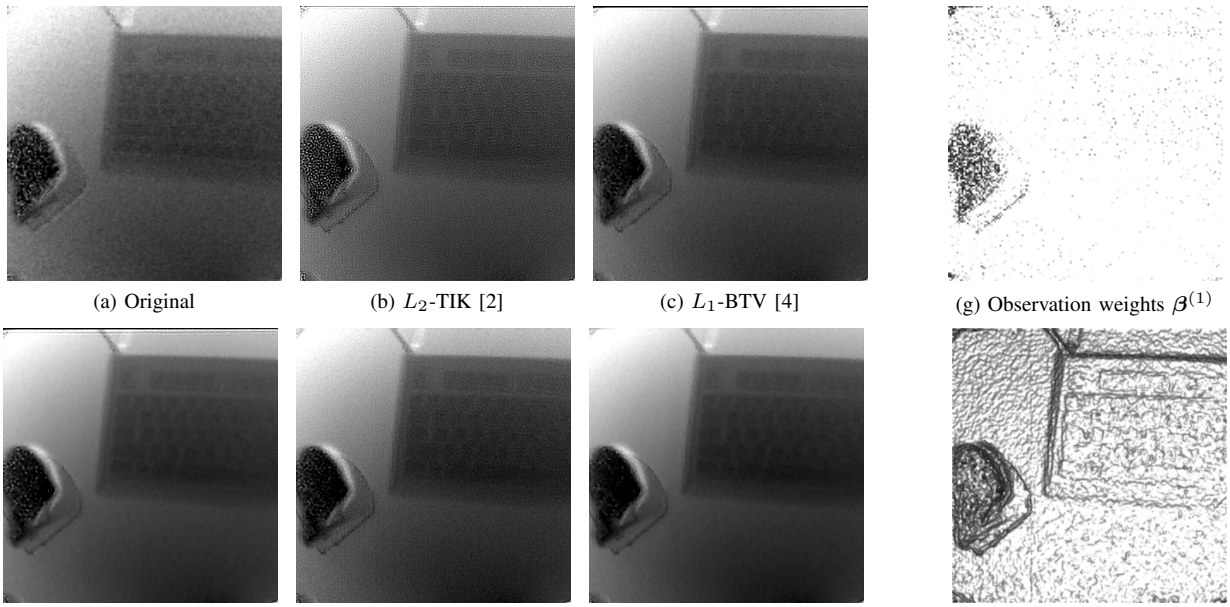
First, we conducted experiments on the MDSP benchmark dataset [46]. We approximated the PSF by an isotropic Gaussian kernel of size $6\sigma_{\text{PSF}}$ and used the *Adyoron* (frame 1 to 20, $\sigma_{\text{PSF}} = 0.4$, $s = 3$) as well as the *Car* sequence (frame 1 to 12, $\sigma_{\text{PSF}} = 0.5$, $s = 3$). Motion estimation was performed by the ECC method [44] using an affine motion model. However, individual frames were affected by a substantial amount of rotation and zoom, which resulted in misregistrations. The reconstructed images obtained by the different methods are depicted in Fig. 13 and Fig. 14. The *Adyoron* sequence follows a translational motion in the first subsequence and affine motion in the second subsequence. Here, the L_2-TIK method produced severe ghosting artifacts due to misregistrations. This was avoided by the algorithms that employ a robust observation model. In particular, the proposed method achieved the most accurate reconstruction of small details, e. g. the font, compared to the competing approaches. On the *Car* sequence, our goal was to super-resolve the car's license plate. This sequence follows a translational motion with zoom, which caused misregistrations for individual frames. Hence, the L_2-TIK method caused ghosting artifacts on the license plate. The most reliable reconstruction was achieved by the proposed method, whereas the images provided by the competing robust methods suffered from blurring.

We also examined super-resolution in the field of Time-of-Flight (ToF) imaging to reconstruct high-resolution surface data from low-resolution range images. We acquired range data from indoor scenes with a PMD CamCube 3.0 camera that provides range images of 200×200 pixels. Note that ToF imaging is typically prone to systematic errors, e. g. intensity-dependent noise. This results in space variant noise as distance measurements on dark surfaces are less reliable. Example reconstructions for the *Keyboard* sequence consisting of $K = 16$ frames with a Gaussian PSF ($\sigma_{\text{PSF}} = 0.5$) and magnification $s = 3$ are depicted in Fig. 15. Space variant noise is visible on the punch caused by its black surface. L_2-TIK, L_1-BTV, LOR and BEP failed to suppress space variant noise resulting in an inaccurate reconstruction of this surface. More reliable surface reconstructions were provided by the proposed algorithm. Fig. 15g shows the confidence map $\boldsymbol{\beta}^{(1)}$ associated with the first frame that was estimated by our method. This clearly highlights the identification of space variant noise, which is indicated by lower confidence weights in regions with higher noise levels. In addition, Fig. 15h visualizes the prior weights $\boldsymbol{\alpha}$ after convergence of our algorithm.

D. Computational Complexity

We also performed a detailed analysis of the computational complexity of our method in terms of run time[3]. In addition, we evaluated the number of energy function evaluations for numerical optimization in the different algorithms. Table I compares these measurements for the *Adyoron* and the *Car*

[3]The run time evaluation was performed on an Intel Xeon E3-1245 CPU with 3.4 GHz and 16 GB RAM using our experimental Matlab implementation.



(a) Original (b) L_2-TIK [2] (c) L_1-BTV [4] (g) Observation weights $\boldsymbol{\beta}^{(1)}$

(d) LOR [10] (e) BEP [11] (f) Proposed (h) Prior weights $\boldsymbol{\alpha}$

Fig. 15: Super-resolution on Time-of-Flight range data with space variant noise ($K = 16$ frames, magnification factor $s = 3$). Super-resolved images obtained by the different methods are depicted in (b) - (f). The observation confidence weights $\boldsymbol{\beta}^{(1)}$ associated with the first frame as well as the prior weights $\boldsymbol{\alpha}$ after convergence of our algorithm are visualized in gray-scale in (g) and (h), respectively (brighter regions denote higher weights).

sequence and summarizes the performance of iteratively re-weighted minimization with different configurations. We observed that L_2-TIK converged quite fast and resulted in the lowest run time among the compared methods while the different robust methods typically required more iterations.

In order to provide insights into the complexity of the different computational steps of our algorithm, we evaluated iteratively re-weighted minimization using the proposed hyperparameter selection (adaptive λ) and with a bypass of this step (constant λ). As expected, the adaptive version resulted in an increased run time and a higher number of function evaluations relative to the competing algorithms that do not provide automatic parameter selection. However, if we bypass the hyperparameter selection, the complexity of iteratively re-weighted minimization is comparable to those of the other algorithms. In addition, we compared our algorithm with coarse-to-fine optimization to a straightforward single-scale implementation. The by-passing of the coarse-to-fine iterations resulted in substantially higher computational costs. This proves the significance of the proposed iteration scheme that enables robust super-resolution without manual parameter tuning coupled with a reasonable computational effort.

TABLE I: Computational complexity in terms of run time and the number of function evaluations for numerical optimization. The measures in brackets denote the relative increase with respect to L_2-TIK considered as the baseline.

Method	**Adyoron sequence** (66×76, $K = 20$ frames)		**Car sequence** (70×50, $K = 12$ frames)	
	Time [s]	# Fun. eval.	Time [s]	# Fun. eval.
L_2-TIK [2]	4.7 ($\times$ 1.0)	31 ($\times$ 1.0)	1.9 ($\times$1.0)	50 ($\times$ 1.0)
L_1-BTV [4]	9.9 ($\times$ 2.1)	50 ($\times$ 1.6)	3.4 ($\times$ 1.8)	50 ($\times$ 1.0)
LOR [10]	10.8 ($\times$ 2.3)	50 ($\times$ 1.6)	3.6 ($\times$ 1.9)	50 ($\times$ 1.0)
BEP [11]	32.5 ($\times$ 6.9)	22 ($\times$ 0.7)	18.3 ($\times$ 9.6)	50 ($\times$ 1.0)
Proposed				
coarse-to-fine (adaptive λ)	**36.2** ($\times$ 7.7)	**250** ($\times$ 8.1)	**16.0** ($\times$ 8.4)	**250** ($\times$ 5.0)
coarse-to-fine (constant λ)	19.4 ($\times$ 4.1)	50 ($\times$ 1.6)	7.3 ($\times$ 3.8)	50 ($\times$ 1.0)
w/o coarse-to-fine (adaptive λ)	66.4 ($\times$ 14.1)	246 ($\times$ 7.9)	24.4 ($\times$ 12.8)	250 ($\times$ 5.0)

VI. Conclusions and Future Work

In this paper, we have introduced a new robust multiframe super-resolution algorithm. Unlike prior work, our method is derived from a spatially adaptive Bayesian model to consider space variant noise and outliers as well as sparse regularization. The proposed iteratively re-weighted minimization algorithm provides a joint estimation of model confidence weights, regularization parameters and the super-resolved image. We examined super-resolution under challenging conditions of practical relevance including inaccurate geometric and photometric registration, invalid pixels and space variant noise. Our evaluation confirms the performance of our algorithm that outperforms other robust methods based on L_1 norm and M-estimator models. We also examined super-resolution for ToF imaging, where the properties of our algorithm are essential to reconstruct accurate 3-D surface data. Our method combines the benefits of being robust, computationally efficient and easy to implement. Moreover, it provides automatic parameter selection in contrast to methods that require manual parameter tuning. This is profitable in practical applications as parameter settings might have substantial influence to the algorithms robustness.

This framework might be further extended in several ways. In this work, we limited ourselves to non-blind super-resolution, where the PSF is assumed to be known. However, iteratively re-weighted minimization could be augmented by blur estimation as also used in [21]. Another promising extension is joint

motion estimation and super-resolution, e. g. by using the non-linear least squares algorithm in [19]. Conversely, blur and motion estimation can also benefit when using it in combination with our spatially adaptive model. One further direction of our future work is to make our approach adaptive to the scene content, e. g. by a local selection of the sparsity parameter p.

APPENDIX
PROOF OF THEOREM 1

In this appendix, we summarize our proof for Theorem 1 to analyze the convergence of our method. Our proof is a modified version of the derivation of Chen and Zhou [42] that addresses L_p norm regularized least-squares optimization. Let us first consider the sparse regularization term defined via the L_1/L_p norm in (35) that fulfills the following lemma:

Lemma 1. *For all index sets $\mathcal{I}'$, $\mathcal{I}(\boldsymbol{z}) = \{i : z_i > 1\}$ and the parameter $0 < p < 1$, there is:*

$$\sum_{i \notin \mathcal{I}(\boldsymbol{z})} |z_i| + \sum_{i \in \mathcal{I}(\boldsymbol{z})} |z_i|^p \leq \sum_{i \notin \mathcal{I}'} |z_i| + \sum_{i \in \mathcal{I}'} |z_i|^p. \tag{37}$$

Proof. From (37) and $\mathcal{I} = \mathcal{I}(\boldsymbol{z})$ we have:

$$\begin{aligned}
&\sum_{i \notin \mathcal{I}} |z_i| - \sum_{i \notin \mathcal{I}'} |z_i| + \sum_{i \in \mathcal{I}} |z_i|^p - \sum_{i \in \mathcal{I}'} |z_i|^p \\
&\quad = \sum_{i \in \mathcal{I}' \setminus \mathcal{I}} |z_i| - \sum_{i \in \mathcal{I} \setminus \mathcal{I}'} |z_i| + \sum_{i \in \mathcal{I} \setminus \mathcal{I}'} |z_i|^p - \sum_{i \in \mathcal{I}' \setminus \mathcal{I}} |z_i|^p \\
&\quad = \sum_{i \in \mathcal{I}' \setminus \mathcal{I}} \left(|z_i| - |z_i|^p\right) - \sum_{i \in \mathcal{I} \setminus \mathcal{I}'} \left(|z_i| - |z_i|^p\right).
\end{aligned}$$

Notice that $|z_i| \geq |z_i|^p$ if and only if $z_i \geq 1$, i.e., $i \in \mathcal{I}(\boldsymbol{z})$. Thus, $|z_i| - |z_i|^p < 0, \forall i \in \mathcal{I}' \setminus \mathcal{I}(\boldsymbol{z})$ and $|z_i| - |z_i|^p \geq 0, \forall i \in \mathcal{I}(\boldsymbol{z}) \setminus \mathcal{I}'$. From these inequalities, it follows that:

$$\sum_{i \in \mathcal{I}' \setminus \mathcal{I}(\boldsymbol{z})} \left(|z_i| - |z_i|^p\right) - \sum_{i \in \mathcal{I}(\boldsymbol{z}) \setminus \mathcal{I}'} \left(|z_i| - |z_i|^p\right) \leq 0.$$

Hence, (37) is true for any index set $\mathcal{I}'$. □

Next, let us now examine the Huber loss function defined in (34). The following lemma shows how this loss function can be expressed in terms of weighted optimization.

Lemma 2. *The Huber loss function in* (34) *can be written as the weighted minimization problem:*

$$\begin{aligned}
h_\delta(z) &= \underset{\beta \in \mathbb{R}_0^+}{\arg\min} \left\{ \beta z^2 + \delta^2 \rho(\beta) \right\}, \\
\text{where } \rho(\beta) &= \begin{cases} \frac{1}{\beta} - 1 & \text{if } \ 0 \leq \beta < 1 \\ 0, & \text{if } \ \beta \geq 1. \end{cases}
\end{aligned} \tag{38}$$

Proof. Obviously, $h_\delta(z)$ is a convex function and when $\beta \geq 1$, it is monotonically increasing. Thus, the optimal weight is $\beta^* = 1$ in case of $z^2 \leq \delta^2$ or $\beta^* = \frac{\delta}{|z|}$ in case of $z^2 > \delta^2$, where the later comes from the first order optimality condition. Comparing the objective values the optimal weight is:

$$\beta^* = \begin{cases} 1 & \text{if } \ |z| \leq \delta \\ \frac{\delta}{|z|} & \text{if } \ |z| > \delta \end{cases}. \tag{39}$$

Therefore, the solution of the weighted minimization yields:

$$\min_{\beta \in \mathbb{R}_0^+} \beta z^2 + \delta^2 \rho(\beta) = \begin{cases} z^2 & \text{if } \ |z| \leq \delta \\ 2\delta|z| - \delta^2 & \text{if } \ |z| > \delta \end{cases}, \tag{40}$$

which coincides with the Huber loss in (34). □

Finally, we use these lemmas for the proof of Theorem 1 to show the convergence of iteratively re-weighted minimization.

Proof. For the sake of notational brevity, we prove this theorem by assuming the identity for the sparsifying transform ($\boldsymbol{S} = \boldsymbol{I}$)[4]. Then, according to Lemma 2, we can reformulate $F(\boldsymbol{x}^t)$ to:

$$\begin{aligned}
F(\boldsymbol{x}^t) &= \lambda R(\boldsymbol{x}^t) + \sum_{i=1}^{KM} \min_{\beta \in \mathbb{R}^+} \beta \left[\boldsymbol{W}\boldsymbol{x}^t - \boldsymbol{y}\right]_i^2 + \delta^2 \rho(\beta) \\
&\leq \lambda R(\boldsymbol{x}^t) + \sum_{i=1}^{KM} \beta_i^t \left[\boldsymbol{W}\boldsymbol{x}^t - \boldsymbol{y}\right]_i^2 + \delta^2 \rho(\beta_i^t),
\end{aligned}$$

where β_i^t is computed according to (19). Comparing the weights given by (19) and (39), one can verify that:

$$F(\boldsymbol{x}^t) = \lambda R(\boldsymbol{x}^t) + \sum_{i=1}^{KM} \beta_i^t \left[\boldsymbol{W}\boldsymbol{x}^t - \boldsymbol{y}\right]_i^2 + c_0,$$

where $c_0 \geq 0$ is a constant not related to $\boldsymbol{x}^t$. Therefore, we can derive the inequality condition for the objective values:

$$\begin{aligned}
F(\boldsymbol{x}^{t-1}) - F(\boldsymbol{x}^t) \geq\ & (\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{y})^\top \boldsymbol{B}^t (\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{y}) \\
& - (\boldsymbol{W}\boldsymbol{x}^t - \boldsymbol{y})^\top \boldsymbol{B}^t (\boldsymbol{W}\boldsymbol{x}^t - \boldsymbol{y}) \\
& + \lambda \big(R(\boldsymbol{x}^{t-1}) - R(\boldsymbol{x}^t)\big),
\end{aligned}$$

where $\boldsymbol{B}^t$ is constructed as $\boldsymbol{B}^t = \operatorname{diag}(\beta_1^t, \ldots, \beta_{KM}^t)$. This inequality condition can be rearranged according to:

$$\begin{aligned}
F(\boldsymbol{x}^{t-1}) - F(\boldsymbol{x}^t) &\geq (\boldsymbol{x}^{t-1} - \boldsymbol{x}^t)^\top \boldsymbol{W}^\top \boldsymbol{B}^t \boldsymbol{W} (\boldsymbol{x}^{t-1} - \boldsymbol{x}^t) \\
&\quad + 2(\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t)^\top \boldsymbol{B}^t (\boldsymbol{W}\boldsymbol{x}^t - \boldsymbol{y}) \\
&\quad + \lambda \big(R(\boldsymbol{x}^{t-1}) - R(\boldsymbol{x}^t)\big) \\
&\geq \underline{\beta}^t \|\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t\|_2^2 \\
&\quad + 2(\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t)^\top \boldsymbol{B}^t (\boldsymbol{W}\boldsymbol{x}^t - \boldsymbol{y}) \\
&\quad + \lambda \big(R(\boldsymbol{x}^{t-1}) - R(\boldsymbol{x}^t)\big),
\end{aligned}$$

where the last inequality is based on Lemma 1 and $\underline{\beta}^t = \min_i \underline{\beta_i}^t$. Then, $\underline{\beta} = \min_k \underline{\beta}^t$ is strictly positive and:

$$\begin{aligned}
F(\boldsymbol{x}^{t-1}) - F(\boldsymbol{x}^t) \geq\ & \underline{\beta} \|\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t\|_2^2 \\
& + 2(\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t)^\top \boldsymbol{B}^t (\boldsymbol{W}\boldsymbol{x}^t - \boldsymbol{y}) \\
& + \lambda \big(R(\boldsymbol{x}^{t-1}) - R(\boldsymbol{x}^t)\big).
\end{aligned} \tag{41}$$

Since $\boldsymbol{x}^t$ is the solution of iteratively re-weighted minimization at iteration t, it follows:

$$\boldsymbol{0} \in \frac{\partial}{\partial \boldsymbol{x}} \Big\{ (\boldsymbol{W}\boldsymbol{x} - \boldsymbol{y})^\top \boldsymbol{B}^t (\boldsymbol{W}\boldsymbol{x} - \boldsymbol{y}) + \lambda \sum_{i=1}^{N'} \alpha_i^t |x_i| \Big\},$$

for $\boldsymbol{x} = \boldsymbol{x}^t$ and α_i^t computed from the weighting function (21).

[4]Note that for $\boldsymbol{S} \neq \boldsymbol{I}$, we can include the transform $\boldsymbol{S}$, by re-formulation of (33) as a constrained problem.

Thus, it follows for the subgradient:

$$2\boldsymbol{W}^\top \boldsymbol{B}^t(\boldsymbol{W}\boldsymbol{x}^t - \boldsymbol{y}) + \lambda c_i^t \alpha_i^t = 0, \forall i$$
$$\text{where } c_i^t \in \begin{cases} \{1\}, & \text{if } x_i^t > 0 \\ [-1,1] & \text{if } x_i^t = 0 \\ \{-1\}, & \text{if } x_i^t < 0 \end{cases}. \tag{42}$$

Substituting the condition in (42) into (41) and using the fact that $c_i^t x_i^t = |x_i^t|$ and $|c_i^t| \leq 1$, leads to:

$$\begin{aligned}
&F(\boldsymbol{x}^{t-1}) - F(\boldsymbol{x}^t) \\
&\geq \underline{\beta}\|\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t\|_2^2 \\
&\quad + \lambda \sum_{i \notin \mathcal{I}(\boldsymbol{x}^{t-1})} \left(|x_i^{t-1}| - |x_i^t| + c_i^t(x_i^t - x_i^{t-1})\right) \\
&\quad + \lambda \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t-1})} \Big(|x_i^{t-1}|^p - |x_i^t|^p \\
&\quad + p|x_i^{t-1}|^{p-1} c_i^t (x_i^t - x_i^{t-1})\Big) \\
&\geq \underline{\beta}\|\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t\|_2^2 \\
&\quad + \lambda \sum_{i \notin \mathcal{I}(\boldsymbol{x}^{t-1})} \Big(|x_i^{t-1}| - |x_i^t| + (|x_i^t| - x_i^{t-1})\Big) \\
&\quad + \lambda \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t-1})} \Big(|x_i^{t-1}|^p - |x_i^t|^p \\
&\quad + p|x_i^{t-1}|^{p-1}(|x_i^t| - x_i^{t-1})\Big) \\
&\geq \underline{\beta}\|\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t\|_2^2 \\
&\quad + \lambda \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t-1})} |x_i^{t-1}|^{p-1} \Big((1-p)|x_i^{t-1}| \\
&\quad + p|x_i^t| - |x_i^{t-1}|^{1-p}|x_i^t|^p)\Big) \\
&\geq \underline{\beta}\|\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t\|_2^2,
\end{aligned}$$

where the last inequality is according to Lemma 1 in [42] as corollary of Young's inequality, which completes the proof. □

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers for their valuable comments to improve the quality of this paper.

REFERENCES

[1] P. Milanfar, *Super-resolution imaging*. CRC Press, 2010.

[2] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans Image Process*, vol. 6, no. 12, pp. 1646–1658, 1997.

[3] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution," in *Proc. IEEE Works. on Appl. of Comput. Vis.*, 2001, pp. 645–650.

[4] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution." *IEEE Trans Image Process*, vol. 13, no. 10, pp. 1327–1344, 2004.

[5] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231–239, 1991.

[6] T. Köhler, J. Jordan, A. Maier, and J. Hornegger, "A Unified Bayesian Approach to Multi-Frame Super-Resolution and Single-Image Upsampling in Multi Sensor Imaging," in *Proc. British Machine Vision Conference*. BMVA Press, 2015, pp. 143.1–143.12.

[7] Z. Lin and H. Shum, "Fundamental limits of reconstruction-based superresolution algorithms under local translation," *IEEE Trans Pattern Anal Mach Intell*, vol. 26, no. 1, pp. 83–97, 2004.

[8] W. Zhao and H. S. Sawhney, "Is super-resolution with optical flow feasible?" in *Proc. Europ. Conf. on Comput. Vis.*, 2002, pp. 599–613.

[9] S. Cho, J. Wang, and S. Lee, "Handling outliers in non-blind image deconvolution," in *Proc. IEEE Int. Conf. on Comput. Vis.*, 2011, pp. 495–502.

[10] V. Patanavijit and S. Jitapunkul, "A Lorentzian Stochastic Estimation for a Robust Iterative Multiframe Super-Resolution Reconstruction with Lorentzian-Tikhonov Regularization," *EURASIP J Adv Signal Process*, vol. 2007, no. 1, pp. 21–21, 2007.

[11] X. Zeng and L. Yang, "A robust multiframe super-resolution algorithm based on half-quadratic estimation with modified BTV regularization," *Digit Signal Process*, vol. 23, no. 1, pp. 98–109, 2013.

[12] H. Song, L. Zhang, P. Wang, K. Zhang, and X. Li, "An adaptive L1–L2 hybrid error model to super-resolution," in *Proc. IEEE Int. Conf. Image Process.*, 2010, pp. 2821–2824.

[13] L. Yue, H. Shen, Q. Yuan, and L. Zhang, "A locally adaptive L1-L2 norm for multi-frame super-resolution of images with mixed noise and outliers," *Signal Processing*, vol. 105, pp. 156–174, 2014.

[14] C. Liu and D. Sun, "A Bayesian approach to adaptive video super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 209–216.

[15] T. Köhler, S. Haase, S. Bauer, J. Wasza, T. Kilgus, L. Maier-Hein, H. Feußner, and J. Hornegger, "Outlier detection for multi-sensor super-resolution in hybrid 3d endoscopy," in *Proc. BVM Workshop*. Springer Berlin Heidelberg, 2014, pp. 84–89.

[16] F. C. Ghesu, T. Köhler, S. Haase, and J. Hornegger, "Guided Image Super-Resolution: A New Technique for Photogeometric Super-Resolution in Hybrid 3-D Range Imaging," in *Pattern Recognition*. Springer Berlin Heidelberg, 2014, pp. 227–238.

[17] H. He and L. Kondi, "An image super-resolution algorithm for different error levels per frame," *IEEE Trans Image Process*, vol. 15, no. 3, pp. 592–603, 2006.

[18] M. Vrigkas, C. Nikou, and L. Kondi, "Robust maximum a posteriori image super-resolution," *Journal of Electronic Imaging*, vol. 23, no. 4, pp. 043 016–043 016, 2014.

[19] Y. He, K.-H. Yap, L. Chen, and L.-P. Chau, "A nonlinear least square technique for simultaneous image registration and super-resolution." *IEEE Trans Image Process*, vol. 16, no. 11, pp. 2830–2841, 2007.

[20] X. Zhang, J. Jiang, and S. Peng, "Commutability of blur and affine warping in super-resolution with application to joint estimation of triple-coupled variables." *IEEE Trans Image Process*, vol. 21, no. 4, pp. 1796–808, 2012.

[21] E. Faramarzi, D. Rajan, and M. P. Christensen, "Unified blind method for multi-image super-resolution and single/multi-image blur deconvolution." *IEEE Trans Image Process*, vol. 22, no. 6, pp. 2101–14, 2013.

[22] S. Babacan, R. Molina, and A. Katsaggelos, "Variational bayesian super resolution." *IEEE Trans Image Process*, vol. 20, no. 4, pp. 984–999, 2011.

[23] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Overcoming registration uncertainty in image super-resolution: maximize or marginalize?" *EURASIP J Adv Signal Process*, vol. 2007, no. 2, pp. 20–20, 2007.

[24] M. Ng, H. Shen, E. Lam, and L. Zhang, "A Total Variation Regularization Based Super-Resolution Reconstruction Algorithm for Digital Video," *EURASIP J Adv Signal Process*, vol. 2007, pp. 1–17, 2007.

[25] Q. Yuan, L. Zhang, and H. Shen, "Multiframe Super-Resolution Employing a Spatially Weighted Total Variation Model," *IEEE Trans Circuits Syst Video Technol*, vol. 22, no. 3, pp. 379–392, 2012.

[26] X. Li, Y. Hu, X. Gao, D. Tao, and B. Ning, "A multi-frame image super-resolution method," *Signal Processing*, vol. 90, no. 2, pp. 405–414, 2010.

[27] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing Sparsity by Reweighted L1 Minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[28] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun Pure Appl Math*, vol. 63, no. 1, pp. 1–38, 2010.

[29] A. Levin, Y. Weiss, F. Durand, and W. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1964–1971.

[30] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, "An Iterated L1 Algorithm for Non-smooth Non-convex Optimization in Computer Vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 1759–1766.

[31] Y. Xiao, T. Zeng, J. Yu, and M. K. Ng, "Restoration of images corrupted by mixed Gaussian-impulse noise via l1–l0 minimization," *Pattern Recognition*, vol. 44, no. 8, pp. 1708–1720, 2011.

[32] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans Pattern Anal Mach Intell*, vol. 24, no. 9, pp. 1167–1183, 2002.

[33] K. Hamada, R. Nakashima, K. Takahashi, and T. Naemura, "Super-resolution with Adaptive Pixel Weighting Scheme and Its Application to Super-resolved Free-Viewpoint Image Synthesis," in *IEEE Int. Conf. on Signal-Image Technology & Internet-Based Systems*, 2013, pp. 757–764.
[34] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J R Stat Soc Series B Stat Methodol*, vol. 39, no. 1, pp. 1–38, 1977.
[35] D. Ba, B. Babadi, P. L. Purdon, and E. N. Brown, "Convergence and stability of iteratively re-weighted least squares algorithms," *IEEE Trans Signal Process*, vol. 62, no. 1-4, pp. 183–195, 2014.
[36] J. A. Scales and A. Gersztenkorn, "Robust methods in inverse theory," *Inverse Problems*, vol. 4, no. 4, pp. 1071–1091, 1988.
[37] Y.-W. Wen and R. H. Chan, "Parameter selection for total-variation-based image restoration using discrepancy principle." *IEEE Trans Image Process*, vol. 21, no. 4, pp. 1770–81, 2012.
[38] N. Nguyen, P. Milanfar, and G. Golub, "Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement." *IEEE Trans Image Process*, vol. 10, no. 9, pp. 1299–308, 2001.
[39] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. Figueiredo, "Adaptive total variation image deblurring: A majorization–minimization approach," *Signal Processing*, vol. 89, no. 9, pp. 1683–1693, 2009.
[40] I. T. Nabney, *NETLAB: Algorithms for Pattern Recognition*, 1st ed. Springer, 2002.
[41] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Robust shift and add approach to superresolution," in *Proc. SPIE Appl. of Dig. Image Process.*, 2003, pp. 121–130.
[42] X. Chen and W. Zhou, "Convergence of the reweighted L1 minimization algorithm for L2-Lp minimization," *Computational Optimization and Applications*, vol. 59, no. 1-2, pp. 47–61, 2014.
[43] H. Sheikh and L. Z. Wang, "LIVE Image Quality Assessment Database," 2015. [Online]. Available: http://live.ece.utexas.edu/research/quality
[44] G. D. Evangelidis and E. Z. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization." *IEEE Trans Pattern Anal Mach Intell*, vol. 30, no. 10, pp. 1858–65, 2008.
[45] D. Capel and A. Zisserman, "Computer vision applied to super resolution," *IEEE Signal Process Mag*, vol. 20, no. 3, pp. 75–86, 2003.
[46] S. Farsiu, D. Robinson, and P. Milanfar, "Multi-Dimensional Signal Processing Dataset," 2015. [Online]. Available: http://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html

Thomas Köhler received the diploma in computer science in engineering at the University of Applied Sciences Würzburg-Schweinfurt, Schweinfurt, Germany in 2008 and the M.Sc. degree in computer science at the University of Applied Sciences Munich, Munich, Germany in 2010.

From 2010 to 2011 he worked as a Research Assistant at the Fraunhofer Institute for Integrated Circuits, Erlangen, Germany in the field of image processing for contactless test and measuring systems. He is currently pursuing the Ph.D. degree in computer science at the Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. His current research areas include inverse problems in image and video processing, especially blind deconvolution and super-resolution.

Xiaolin Huang (S'10-M'12) received the B.S. degree in control science and engineering, and the B.S. degree in applied mathematics from Xi'an Jiaotong University, Xi'an, China in 2006. In 2012, he received the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China.

From 2012 to 2015, he worked as a Postdoctoral Researcher in ESAT-STADIUS, KU Leuven, Leuven, Belgium. After that he has been selected as an Alexander von Humboldt Fellow and being working at the Pattern Recognition Lab of the Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. His current research areas include machine learning methods, optimization algorithms, and piecewise linear system identification.

Frank Schebesch was born in Erlangen, Germany, in 1985. He studied mathematics and physics at the Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, where he received his diploma in mathematics in 2013.

Currently, he is pursuing the Ph.D. degree in computer science at the Friedrich-Alexander-Universität Erlangen-Nürnberg. His research interests include image quality assessment in medical images, model observer for digital mammography and digital breast tomosynthesis, computer aided detection algorithms, and statistical evaluation of reading studies.

André Aichert received the B.Sc. and M.Sc. degrees in computer science in 2010 and 2012 at the Technische Universität München, Munich, Germany.

In 2008 and 2010, he served two research internships at Siemens Corporate Research in Princeton, NJ, in the field of medical image registration and medical augmented reality. In 2012, he joined the Department of Neuroradiology at the Universitätsklinikum Erlangen, Erlangen, Germany. Currently he is pursuing the Ph.D. in computer science at the Friedrich Alexander Universität Erlangen-Nürnberg, Erlangen, Germany. His research interests are medical image processing and computer vision, notably image registration and data consistency in X-ray imaging.

Andreas Maier (M'05) graduated in computer science in 2005 and received the Ph.D. degree in computer science at the Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, in 2009.

From 2005 to 2009 he was working at the Pattern Recognition Lab at the Friedrich-Alexander-Universität Erlangen-Nürnberg, where he developed the first online speech intelligibility assessment tool (PEAKS). From 2009 to 2010, he started working on flat-panel C-arm CT as Postdoctoral Fellow at the Radiological Sciences Laboratory of the Stanford University, Stanford, CA. From 2011 to 2012 he joined Siemens Healthcare as Innovation Project Manager and was responsible for reconstruction topics in the angiography and X-ray business unit. In 2012, he returned to the Friedrich-Alexander-Universität Erlangen-Nürnberg as Postdoctoral Researcher and Head of the Medical Image Reconstruction Group at the Pattern Recognition Lab. In 2015, he became Professor and Head of the Pattern Recognition Lab.

Joachim Hornegger graduated in theoretical computer science/mathematics in 1992 and received the Ph.D. degree in applied computer science at the Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, in 1996.

He was an Assistant Professor at the Friedrich-Alexander-Universität Erlangen-Nürnberg (1996/97), a Visiting Scientist at the Technion, Haifa, Israel, and a Visiting Scholar at the Massachusetts Institute of Technology, Cambridge, MA, and the Stanford University, Stanford, CA (1997/98). In 1998, he joined Siemens Medical Solutions, Inc., where he was working on 3-D angiography. In parallel, he was a Lecturer at the Universities of Erlangen (1998 to 1999), Eichstaett-Ingolstadt (2000), and Mannheim (2000 to 2003). In 2003, he became a Professor of Medical Imaging Processing, from 2005 to 2015 he was Head of the Pattern Recognition Lab, and in 2015 he was elected President of the Friedrich-Alexander-Universität Erlangen-Nürnberg. He is author and co-author of more than 50 scientific publications including a monography on applied pattern recognition and a book on statistical object recognition. His main research topics are medical image processing and computer vision, image registration and sensor data fusion.