A Unified Bayesian Approach to Multi-Frame Super-Resolution and Single-Image Upsampling in Multi-Sensor Imaging

Thomas Köhler[12]
thomas.koehler@fau.de
Johannes Jordan[1]
johannes.jordan@fau.de
Andreas Maier[12]
andreas.maier@fau.de
Joachim Hornegger[12]
joachim.hornegger@fau.de

[1] Pattern Recognition Lab,
Department of Computer Science,
Friedrich-Alexander-Universität
Erlangen-Nürnberg
Erlangen, Germany

[2] Erlangen Graduate School in Advanced
Optical Technologies (SAOT)
Erlangen, Germany

Abstract

For a variety of multi-sensor imaging systems, there is a strong need for resolution enhancement. In this paper, we propose a unified method for single-image upsampling and multi-frame super-resolution of multi-channel images. We derive our algorithm from a Bayesian model that is formulated by a novel image prior to exploit sparsity of individual channels as well as a locally linear regression between the complementary channels. The reconstruction of high-resolution multi-channel images from low-resolution ones and the estimation of associated hyperparameters to define our prior model is formulated as a joint energy minimization. We introduce an alternating minimization scheme to solve this non-convex optimization problem efficiently. Our framework is applicable to various types of multi-sensor setups that are addressed in our experimental evaluation, including color, multispectral and 3-D range imaging. Comprehensive qualitative and quantitative comparisons demonstrate that our method outperforms state-of-the-art algorithms.

1 Introduction

The spatial resolution of a digital imaging system is one of its major quality indicators. If the resolution provided by the underlying hardware does not meet the desired application, software-based techniques to reconstruct high-resolution images are essential. Resolution enhancement can be tackled by single-image upsampling [29] or from a multi-frame perspective, e. g. by exploiting sub-pixel motion between multiple frames [22][1]. Most of the established techniques [3, 6, 8] consider resolution enhancement of single-channel images, e. g. monochromatic data. However, many imaging systems of practical interest are *multi-sensor* systems. Thus, they provide *multi-channel* images that may even combine different modalities. This can be achieved with a single camera equipped with a sensor array or by using different camera technologies that are combined by sensor data fusion. For many multi-sensor

[1] Super-Resolution refers to multi-frame approaches [6] to discriminate it from single-image upsampling [25].

Pages 143.1-143.12
DOI: https://dx.doi.org/10.5244/C.29.143

systems, there is a strong need for super-resolution or upsampling as the sensor resolution is often limited due to economical or technological constraints. From a resolution enhancement point of view, these systems pose both challenges and additional benefits. One issue is, that a straightforward application of resolution enhancement to each channel separately might not be optimal as this strategy ignores dependencies between the channels. This is obvious if image structures are visible in multiple channels. Moreover, these dependencies can be exploited as prior knowledge to overcome the ill-posedness of the image reconstruction problem. Our key idea is to exploit correlations between complementary channels as shown, e. g. in prior work on color representation [24], dehazing [11] or restoration [21]. Inspired by these concepts, we formulate a novel approach to super-resolution and upsampling in multi-sensor imaging. We address three applications that are of great interest in computer vision.

Related Work. The majority of commercially available cameras provide multi-channel images encoded in red (R), green (G) and blue (B) spectral components. In an early work, Zomet and Peleg [32] have proposed multi-sensor super-resolution for such setups. In their approach, the reconstruction of one color channel is guided by the set of the remaining channels to upsample the multi-channel image. Moreover, some existing multi-frame methods were extended to deal with color images. In this context, Farsiu *et al.* [9] as well as Gotoh and Okutomi [14] have proposed inter-channel regularization in the RGB space, which is driven by the assumption that color channels are correlated. These methods often share the property that color space transformations, e. g. the HSV space [9], are used to represent color images. This means a limitation towards multi-channel images with more spectral components. For instance, multispectral and hyperspectral sensors widely used in remote sensing capture the scene with a higher spectral resolution. Attempts have been made at extending super-resolution for an arbitrary number of channels [2]. In some of these applications, a panchromatic image needs to be used as guidance to super-resolve multispectral data [1].

In range imaging, one common setup is to combine 3-D range data with 2-D photometric data. Stereo vision is a passive technique to provide a disparity map computed from a pair of color images that is fused with one of these images. On the other hand, Time-of-Flight (ToF) and structured light are active sensor technologies realized, e. g. in Microsoft's Kinect. In addition to range data, color images or amplitude data are acquired. This sensor fusion can be interpreted as a multi-channel image. One emerging topic is to overcome the low spatial resolution of range sensors, which can be performed by super-resolution on the range channel [2, 30]. Over the past years, various single- and multi-frame methods have been introduced to enhance this task using color channels as guidance, including guided filtering [15], motion estimation techniques [19], spatially adaptive regularization [20, 25], anisotropic total generalized variation [12] and joint photometric and range co-sparse analysis [18]. Similar to multispectral imaging, the drawback of these concepts is the need for high-quality color data, which is hard to achieve with low-cost devices. To overcome this limitation, Ghesu *et al.* [13] have proposed guided super-resolution applied to joint resolution enhancement of range and photometric data. However, their formulation does only consider two channels.

Contribution. This paper examines multi-channel super-resolution and upsampling in a unified Bayesian framework. Unlike prior work in multispectral and range imaging, our approach does not rely on additional guidance data. Moreover, it is not limited to a fixed number of channels such as [13], making it relevant for a wide range of multi-sensor systems. In detail, our contribution is threefold: 1) We formulate resolution enhancement using a Bayesian multi-channel model and introduce a novel image prior that accounts for inter-channel dependencies. Our prior is spatially adaptive and tolerates image regions that violate

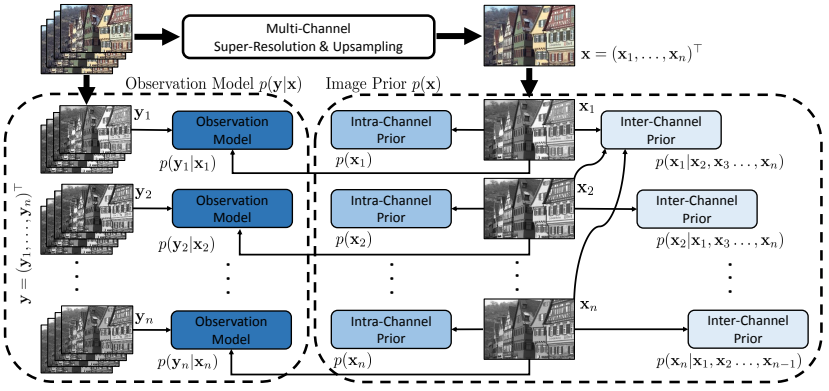


Figure 1: Proposed Bayesian model with a channel-wise observation model, an intra-channel prior per channel and an inter-channel prior to exploit dependencies between channels.

the assumption of dependencies. 2) We propose a joint energy minimization algorithm for single-image upsampling and multi-frame super-resolution. 3) The impact of our method is investigated for various applications including color-, multispectral and range imaging.

2 Bayesian Multi-Sensor Model

We model multi-channel images by $\boldsymbol{x} = (\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_n^\top)^\top$, where $\boldsymbol{x}_i$ $(1 \leq i \leq n)$ denotes the i-th channel. Each channel $\boldsymbol{x}_i$ is denoted by a $N_i \times 1$ vector and we assume uniform dimensionality, i. e. $N_i = N$. For each $\boldsymbol{x}_i$, a camera captures K low-resolution frames $\boldsymbol{y}_i = (\boldsymbol{y}_i^{(1)\top}, \ldots, \boldsymbol{y}_i^{(K)\top})^\top$ due to camera or object motion. Each frame $\boldsymbol{y}_i^{(k)}$ is encoded by a $M_i \times 1$ vector with $M_i = M$ and $M < N$. The set of low-resolution channels is given by $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_n^\top)^\top$. The goal is to reconstruct the multi-channel image $\boldsymbol{x}$ from $\boldsymbol{y}$. For $K > 1$, this results in a multi-frame super-resolution problem. For $K = 1$, we face the problem of single-image upsampling. In both cases, our algorithm is derived from the maximum a-posteriori (MAP) estimation:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n}{\arg\max} \left\{ p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) p(\boldsymbol{x}) \right\}, \tag{1}$$

where the probability density function $p(\boldsymbol{y}\,|\,\boldsymbol{x}) = p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ models the probability of observing $\boldsymbol{y}$ from $\boldsymbol{x}$ and $p(\boldsymbol{x})$ is the prior. If we would assume mutually independent channels, each channel $\boldsymbol{x}_i$ can be reconstructed separately using using single-channel approaches. However, this assumption would ignore correlations across the channels caused by the fact that structures in one channel might be also present in the other channels. Our approach is derived from the key idea to exploit these dependencies to guide the reconstruction of the entire multi-channel image $\boldsymbol{x}$. In doing so, we factorize $p(\boldsymbol{x})$ in Eq. (1) as $p(\boldsymbol{x}) = p(\boldsymbol{x}_1) p(\boldsymbol{x}_2 \,|\, \boldsymbol{x}_1) \ldots p(\boldsymbol{x}_n \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1})$ that encodes these dependencies. The basic structure of the Bayesian model that is employed for this purpose is shown in Fig. 1.

2.1 Observation Model

In this work, we limit ourselves to a general and very flexible observation model that describes the formation of a low-resolution channel $\boldsymbol{y}_i^{(k)}$ from the high-resolution channel $\boldsymbol{x}_i$ as follows. First, $\boldsymbol{y}_i^{(k)}$ is displaced with respect to $\boldsymbol{x}_i$ due to camera or object motion during image

acquisition. Second, each displaced frame is affected by a blur kernel given by the camera point spread function (PSF). Third, the spatial resolution in terms of pixel dimension is reduced due to sub-sampling on a sensor array. Finally, each frame is disturbed by additive noise $\boldsymbol{\varepsilon}_i$. This is described by the image formation model $\boldsymbol{y}_i = \boldsymbol{D}_i\boldsymbol{B}_i\boldsymbol{M}_i\boldsymbol{x}_i + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{D}_i$, $\boldsymbol{B}_i$ and $\boldsymbol{M}_i$ model sub-sampling, blur and motion, respectively [6]. These transforms are combined to the system matrix $\boldsymbol{W}_i = \boldsymbol{D}_i\boldsymbol{B}_i\boldsymbol{M}_i$ for the i-th channel. Under the assumption of additive Gaussian noise of standard deviation σ_i for channel i and the modeling of dependencies between $\boldsymbol{y}_i$ and the high-resolution channel $\boldsymbol{x}_i$, the observation model is given by:

$$p(\boldsymbol{y}\,|\,\boldsymbol{x}) \propto \exp\Big\{-\sum_{i=1}^{n}\frac{1}{2\sigma_i^2}\,||\boldsymbol{y}_i - \boldsymbol{W}_i\boldsymbol{x}_i||_2^2\Big\}. \tag{2}$$

This model generalizes fairly well to many imaging systems and can be easily extended by application-specific aspects, e. g. non-Gaussian noise [8], mosaicing in CCD devices [9], photometric distortions of color images [26] or out-of-plane motion in range imaging [19].

2.2 Image Prior Model

The image prior $p(\boldsymbol{x})$ models two complementary aspects. First, we use an intra-channel prior defined for each channel $\boldsymbol{x}_i$ separately. This exploits the sparsity of each channel $\boldsymbol{x}_i$ as a common assumption in image restoration. Second, each channel $\boldsymbol{x}_i$ is assumed to be correlated with all other channels. This is modeled by an inter-channel prior defined for all pairs $(\boldsymbol{x}_i,\boldsymbol{x}_j)$ for $i \neq j$. In summary, for the channel $\boldsymbol{x}_i$, the prior model is given by:

$$p(\boldsymbol{x}_i\,|\,\mathcal{X}_i) \propto \exp\Big\{-\big(\lambda_i R_{\text{intra}}(\boldsymbol{x}_i) + \sum_{j=1,j\neq i}^{n}\mu_{ij}R_{\text{inter}}(\boldsymbol{x}_i,\boldsymbol{x}_j;\boldsymbol{\Phi}_{ij})\big)\Big\}, \tag{3}$$

where $\mathcal{X}_i = \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{i-1},\boldsymbol{x}_{i+1},\ldots,\boldsymbol{x}_n\}$, $R_{\text{intra}}(\boldsymbol{x}_i)$ is the intra-channel regularization term for $\boldsymbol{x}_i$ and $R_{\text{inter}}(\boldsymbol{x}_i,\boldsymbol{x}_j)$ is the inter-channel regularization term for $(\boldsymbol{x}_i,\boldsymbol{x}_j)$ parametrized by the hyperparameters $\boldsymbol{\Phi}_{ij}$. The weight $\lambda_i \geq 0$ specifies the strength of the intra-channel prior for the i-th channel. Similarly, $\mu_{ij} \geq 0$ specifies how strong dependencies between the i-th and the j-th channel are enforced. Note that the inter-channel prior can be employed in a non-symmetric manner, i. e. $\mu_{ij} \neq \mu_{ji}$. If $\mu_{ij} = 0$, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are treated as independent.

Intra-Channel Model. For the intra-channel model in Eq. (3), we employ bilateral total variation (BTV) [8] as an edge preserving image prior. This prior is defined by:

$$R_{\text{intra}}(\boldsymbol{x}_i) = \sum_{l=-L}^{L}\sum_{m=-L}^{L}\alpha^{|l|+|m|}\big|\big|\boldsymbol{S}_v^l\boldsymbol{S}_h^m\boldsymbol{x}_i - \boldsymbol{x}_i\big|\big|_1, \tag{4}$$

where $\boldsymbol{S}_v^l$ and $\boldsymbol{S}_h^m$ determines a shifted version of $\boldsymbol{x}_i$ by l pixel in vertical and m pixel in horizontal direction, respectively. The derivatives computed over different scales in an $2(L+1)\times 2(L+1)$ window are weighted by $\alpha > 0$ and the magnitude of the shift.

Inter-Channel Model. The inter-channel prior in Eq. (3) models pair-wise dependencies between complementary channels. To define this prior, we employ the *locally linear regression* (LLR) model between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$:

$$R_{\text{inter}}(\boldsymbol{x}_i,\boldsymbol{x}_j;\boldsymbol{\Phi}_{ij}) = \big|\big|\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j)\odot(\boldsymbol{A}_{ij}\boldsymbol{x}_i + \boldsymbol{b}_{ij} - \boldsymbol{x}_j)\big|\big|_2^2, \tag{5}$$

where $\boldsymbol{\Phi}_{ij} = (\boldsymbol{A}_{ij},\boldsymbol{b}_{ij})$ denotes the LLR hyperparameters given by the filter coefficients $\boldsymbol{A}_{ij} \in \mathbb{R}^{N\times N}$ and $\boldsymbol{b}_{ij} \in \mathbb{R}^N$ and $\odot$ is the Hadamard (element-wise) product. We restrict $\boldsymbol{A}_{ij}$ to

a diagonal matrix to define the LLR per pixel. To take a violation of the linear regression assumption into account and make this prior spatially adaptive, we employ confidence weighting $\boldsymbol{\kappa}: \mathbb{R}^N \times \mathbb{R}^N \to [0;1]^N$ to assign weights $\boldsymbol{\kappa}_{ij} = \boldsymbol{\kappa}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ to each pixel in $\boldsymbol{x}_i$. $\boldsymbol{\kappa}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is selected adaptively according to the following properties: 1) The weight $\kappa_{ij,l}$ associated with pixel l is set to $\kappa_{ij,l} > 0$ if the regression assumption holds true and a higher $\kappa_{ij,l}$ denotes a higher fidelity of this assumption. 2) Otherwise, $\kappa_{ij,l} = 0$ should indicate an outlier, which is the case, e. g. if a particular structure is present in only one of the channels.

The LLR model in Eq. (5) extends and generalizes the guided filter prior recently introduced in [13] and provides two desirable properties compared to prior work. On the one hand, our prior generalizes directly to an arbitrary number of image channels. On the other hand, due to its spatially adaptive definition, it tolerates outliers caused by a violation of the linear regression assumption between individual channels enhancing its robustness.

3 Optimization Algorithm

Our approach is derived from the Bayesian multi-channel model and formulated as joint energy minimization. The filter coefficients $\boldsymbol{\Phi}_{ij}$ and the confidence weights $\boldsymbol{\kappa}_{ij}$ are assumed to be unknown and are treated as latent variables. Taking the negative log-likelihood of Eq. (1), we reconstruct the high-resolution image $\boldsymbol{x}$ and the unknown $\boldsymbol{\Phi}_{ij}$ by the joint optimization:

$$\arg\min_{\boldsymbol{x},\boldsymbol{\Phi}} \sum_{i=1}^{n} ||\boldsymbol{y}_i - \boldsymbol{W}_i \boldsymbol{x}_i||_2^2 + \varepsilon\Gamma(\boldsymbol{\Phi}) + \sum_{i=1}^{n} \lambda_i R_{\text{intra}}(\boldsymbol{x}_i) + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mu_{ij} R_{\text{inter}}(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\Phi}_{ij}), \quad (6)$$

where $\Gamma(\boldsymbol{\Phi})$ denotes a prior for the set of unknown hyperparameters $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_{11}, \ldots, \boldsymbol{\Phi}_{nn})$ weighted by $\varepsilon \geq 0$. This is a non-convex optimization problem due to the non-convexity of the inter-channel regularization term $R_{\text{inter}}(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\Phi}_{ij})$. Moreover, Eq. (6) is underdetermined.

3.1 Alternating Minimization

For an efficient solution of Eq. (6), we alternately minimize it either with respect to $\boldsymbol{x}$ or $\boldsymbol{\Phi}$ while keeping the remaining parameters constant. Each iteration t yields refined estimates $\boldsymbol{x}^{(t)}$ and $\boldsymbol{\Phi}^{(t)}$ of the multi-channel image and the hyperparameters, respectively. We also gradually update the confidence weights in our algorithm at each iteration.

Estimate Filter Coefficients. The filter coefficients $\boldsymbol{A}_{ij}^{(t)}$ and $\boldsymbol{b}_{ij}^{(t)}$ are determined by exploiting the LLR model for $\boldsymbol{x}^{(t-1)}$. Similar to ridge regression, we use $\Gamma(\boldsymbol{\Phi}) = ||\boldsymbol{A}||_F^2$ as prior for the filter coefficients $\boldsymbol{A} = (\boldsymbol{A}_{11}, \ldots, \boldsymbol{A}_{nn})^\top$, where $||\cdot||_F$ is the Frobenius norm. Then, for each pair of channels, we need to solve $\arg\min_{\boldsymbol{A}_{ij}, \boldsymbol{b}_{ij}} \{R_{\text{inter}}(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\Phi}_{ij}) + \varepsilon ||\boldsymbol{A}_{ij}||_F^2\}$. Since $\boldsymbol{A}_{ij}$ is diagonal, this optimization problem is separable with respect to the coefficients in $\boldsymbol{A}_{ij}$ and can be solved pixel-wise. However, since this problem is underdetermined, we minimize a relaxation to Eq. (6). In doing so, we solve a spatially smoothed version of Eq. (6) with respect to the filter coefficients, where smoothing is considered for a neighborhood $\boldsymbol{\omega}_k$ of radius $r \geq 1$ centered at each pixel k. The resulting sub-problem to determine $\boldsymbol{A}_{ij}^{(t)}$ and $\boldsymbol{b}_{ij}^{(t)}$ at pixel k is given by:

$$(\hat{A}_{ij,k}, \hat{b}_{ij,k}) = \arg\min_{A_{ij,k}, b_{ij,k}} \Big\{ \sum_{l \in \boldsymbol{\omega}_k} \big(A_{ij,k} x_{i,l} + b_{ij,k} - x_{j,l}\big)^2 + \varepsilon A_{ij,k}^2 \Big\}, \quad (7)$$

where $\hat{A}_{ij,k}$ and $\hat{b}_{ij,k}$ are estimates for $\boldsymbol{A}_{ij}^{(t)}$ and $\boldsymbol{b}_{ij}^{(t)}$ at pixel k and $x_{i,l}$ and $x_{j,l}$ denote the l-th pixel in the channels i and j extracted from $\boldsymbol{x}^{(t-1)}$. This ridge regression problem is equivalent

to the implicit formulation of unweighted guided filtering as shown by He *et al.* [15]. Hence, the unknown coefficients are computed in closed-form according to:

$$\hat{A}_{ij,k} = \frac{\mathrm{Cov}_{\omega_k}(\boldsymbol{x}_i,\boldsymbol{x}_j) - \mathrm{E}_{\omega_k}(\boldsymbol{x}_i)\mathrm{E}_{\omega_k}(\boldsymbol{x}_j)}{\mathrm{Var}_{\omega_k}(\boldsymbol{x}_i) + \varepsilon}\Big|_{\boldsymbol{x}:=\boldsymbol{x}^{(t-1)}} \quad \text{and} \quad \hat{b}_{ij,k} = \mathrm{E}_{\omega_k}(\boldsymbol{x}_j) - \hat{A}_{ij,k}\mathrm{E}_{\omega_k}(\boldsymbol{x}_i)\Big|_{\boldsymbol{x}:=\boldsymbol{x}^{(t-1)}}, \tag{8}$$

where $\mathrm{E}_{\omega_k}(\cdot)$, $\mathrm{Var}_{\omega_k}(\cdot)$, $\mathrm{Cov}_{\omega_k}(\cdot,\cdot)$ denote the mean, variance and covariance in the local neighborhood ω_k extracted from $\boldsymbol{x}^{(t-1)}$. Finally, $\boldsymbol{A}_{ij}^{(t)}$ and $\boldsymbol{b}_{ij}^{(t)}$ is composed element-wise by the mean of the estimates $\hat{A}_{ij,l}$ and $\hat{b}_{ij,l}$ in the neighborhood ω_k.

Estimate Confidence Map. To define the confidence $\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j)$, we analyze the residual error of the LLR model in Eq. (5) at iteration t given by $\tilde{\boldsymbol{r}}_{ij}^{(t)} = \boldsymbol{A}_{ij}^{(t)}\boldsymbol{x}_i^{(t-1)} + \boldsymbol{b}_{ij}^{(t)} - \boldsymbol{x}_j^{(t-1)}$. As the fidelity of the LLR model is assumed to be spatially smooth, a filtered version $\boldsymbol{r}_{ij}^{(t)}$ is determined from $\tilde{\boldsymbol{r}}_{ij}^{(t)}$ using 3×3 median filtering. Then, we compute $\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j)$ at iteration t for each pixel position k using the iteratively re-weighted least square scheme [13, 28]:

$$\kappa(x_{i,k},x_{j,k}) = \begin{cases} 1 & r_{ij,k}^{(t)} < \frac{1}{2}\sigma^{(t)} \\ \sigma^{(t)}/r_{ij,k}^{(t)} & \frac{1}{2}\sigma^{(t)} \le r_{ij,k}^{(t)} \le 2.5\sigma^{(t)} \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

where $\sigma^{(t)}$ denotes a threshold to discriminate inliers and outliers of the LLR model. This scheme assigns a higher confidence to smaller residuals $r_{ij,k}^{(t)}$ as the fidelity of LLR for $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ at pixel k can assumed to be higher. Otherwise, the confidence is decreased to make our algorithm robust to violations of the model. The threshold $\sigma^{(t)}$ is estimated adaptively and set to the standard deviation of the residual. For robust parameter selection, $\sigma^{(t)}$ is computed from the median absolute deviation (MAD) according to $\sigma^{(t)} = 1.4826\cdot\mathrm{MAD}(\boldsymbol{r}_{ij}^{(t)})$ [27].

Estimate Multi-Channel Image. Once the hyperparameters $\boldsymbol{\Phi}^{(t)}$ and the adaptive confidence weighting $\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j)$ for each pair (i,j) are determined, we reconstruct $\boldsymbol{x}^{(t)}$ according to:

$$\boldsymbol{x}^{(t)} = \underset{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}{\arg\min} \sum_{i=1}^{n} ||\boldsymbol{y}_i - \boldsymbol{W}_i\boldsymbol{x}_i||_2^2 + \sum_{i=1}^{n}\lambda_i R_{\text{intra}}(\boldsymbol{x}_i) + \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\mu_{ij}R_{\text{inter}}(\boldsymbol{x}_i,\boldsymbol{x}_j;\boldsymbol{\Phi}_{ij}^{(t)}). \tag{10}$$

This minimization is performed jointly with respect to the channels $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ while keeping $\boldsymbol{\Phi}^{(t)}$ fixed. For numerical optimization of Eq. (10), we employ Scaled Conjugate Gradient (SCG) iterations [23], where $\boldsymbol{x}^{(t-1)}$ is used as initial guess to obtain $\boldsymbol{x}^{(t)}$.

3.2 Implementation

In this work, image reconstruction is implemented in a non-blind, two-stage procedure:

1) In the first stage, we initialize the blur kernel $\boldsymbol{B}_i$ for each channel according to the PSF of the underlying imaging system. As we assume aligned channels, the sub-pixel motion $\boldsymbol{M}_i$ is the same for each channel, i. e. $\boldsymbol{M}_i = \boldsymbol{M}$. For multi-frame reconstruction, $\boldsymbol{M}$ is estimated using image registration, where the first frame $\boldsymbol{y}^{(1)}$ is used as reference. For single-image upsampling, $\boldsymbol{M}$ is set to identity. To initialize the multi-channel image $\boldsymbol{x}^{(1)}$, the average image [26] is computed and refined by means of a channel-wise reconstruction.

2) In the second stage, our algorithm alternates between the estimation of filter coefficients in Eq. (8) and the confidence weights in Eq. (9) for all pairs of channels (i,j) as well as the joint multi-channel reconstruction in Eq. (10). This stage infers the inter-channel hyperparameters across all channels (see Fig. 1) and refines the initial guess $\boldsymbol{x}^{(1)}$. Alternating minimization is performed until convergence or a maximum number of iterations t_{AM} is reached.

4 Applications and Experimental Evaluation

We examined our method for various multi-sensor imaging setups in computer vision including quantitative evaluations by the peak-signal-to-noise ratio (PSNR). Throughout all experiments, we limited ourselves to the application of our image prior in Eq. (3) with fixed regularization weights $\mu = \mu_{ij}$ for n channels $i, j \in [1;n]$. Numerical optimization was applied with $t_{\mathrm{AM}} = 10$ iterations of alternating minimization and $t_{\mathrm{SCG}} = 10$ SCG iterations.

4.1 Application to RGB and Multispectral Imaging

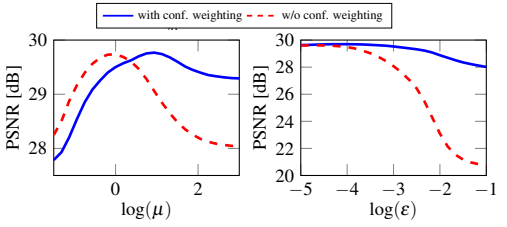
First, we studied single- and multi-frame reconstruction on color ($n = 3$ channels) and multispectral data ($n \gg 3$ channels). As a baseline approach, we processed the channels separately without inter-channel prior. Additionally, we evaluated the inter-channel prior proposed by Farsiu *et al.* [9] as a state-of-the-art technique in color super-resolution[2]. This prior penalizes mismatches of edges across the channels but is not spatially adaptive. BTV was used for intra-channel regularization with $L = 2$ and $\alpha = 0.5$. For our model, we set $r = 1$ and $\varepsilon = 10^{-4}$ with the same weight $\lambda = \lambda_i$ for each channel. The regularization weights μ and λ as well as the parameters of [9] were adjusted for the different applications on a training dataset by optimizing the PSNR or by visual comparison in the absence of a ground truth.

Color Image Super-Resolution. For quantitative evaluation, we simulated 29 low-resolution image sequences from the reference color images of the Live Database [31] used as ground truth. For each channel in the range $[0;1]$, $K = 10$ frames were generated simulating a randomly selected rigid motion, a Gaussian PSF of width $3 \cdot \sigma_{\mathrm{PSF}}$ low-resolution pixels where $\sigma_{\mathrm{PSF}} = 0.25$, a sub-sampling factor of 3 and additive Gaussian noise with standard deviation $\sigma_i = 0.05$. The sub-pixel motion was assumed to be known to explicitly evaluate the influence of the different priors and the PSNR was used for their quantitative comparison.

In this setup, we conducted a parameter sensitivity study regarding the regularization weights μ and ε. The influence of these parameters to the PSNR is plotted in Fig. 2 (left). Our method was applied with confidence weighting (κ_{ij} adaptive using Eq. (9)) and without adaptive weights ($\kappa_{ij} = 1$). We observed more stable results if confidence weighting in the LLR model was employed. This shows the merit of our adaptive prior. The statistics of the PSNR for all simulated datasets is summarized in Fig. 2 (right). Compared to channel-wise super-resolution, we observed substantial improvements by our multi-channel method of ≈ 1.3 dB on average. On simulated data, our LLR model achieved competitive results to the prior of Farsiu *et al.* [9] but shows an increased mean PSNR. A qualitative comparison of the different approaches for the *parrots* image is shown in Fig. 3. Our method achieved reduced noise levels in homogeneous regions as depicted by the color and composite images. Here, we also analyzed the convergence of our alternating minimization by depicting the PSNR of the super-resolved image over the iterations on two example sequences. We observed fast convergence of alternating minimization within ≈ 5 iterations.

We also conducted experiments with real images taken from the MDSP benchmark dataset [10] as depicted in Fig. 4. In this example, the first $K = 10$ frames of the *bookcase* sequence were used for super-resolution with a magnification factor of 3. Sub-pixel motion was estimated using the method in [7] based on an affine motion model. We used a Gaussian PSF ($\sigma_{\mathrm{PSF}} = 0.3$). In our experiments, inter-channel regularization got rid of color artifacts caused by a channel-wise super-resolution, e. g. jagged edges highlighted in Fig. 4. Compared to the regularization technique in [9], these artifacts were well suppressed by the proposed

[2] We omitted chrominance regularization and de-mosaicing as these concepts are only applicable for RGB data.



	Channel-wise	Inter-channel	
		Farsiu [9]	Proposed
Mean ± Std	26.9 ± 1.67	28.0 ± 1.88	**28.2 ± 1.93**
Median	26.8	28.2	**28.3**

Figure 2: Left: Influence of regularization weights μ and ε to the peak-signal-to-noise ratio (PSNR) evaluated with and without confidence weighting. Right: Statistics of PSNR of for color super-resolution on 29 simulated sequences obtained from the Live database [31].

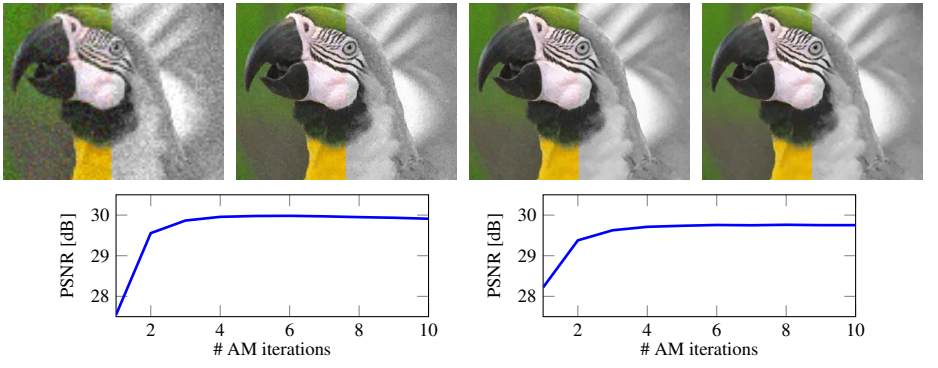


Figure 3: Color super-resolution on simulated data ($K = 10$ frames, $3\times$ magnification). Top row from left to right: low-resolution *parrots* frame, channel-wise super-resolution, multi-channel super-resolution using the prior in [9] and our multi-channel method visualized as composites showing colors and green channel intensities. Bottom row: convergence of our alternating minimization on two image sequences measured by the PSNR over the iterations.

method. In terms of run time, our method converged after 17 s on the *bookcase* sequence of size 90×60 px using our experimental Matlab code[3]. The run times of channel-wise super-resolution and the approach of Farsiu *et al.* [9] were 6 s and 23 s, respectively.

Multispectral Image Upsampling. We applied our method to multispectral images and performed experiments on real data taken from the Harvard dataset [5]. Each image consists of $n = 31$ bands (channels) with central wavelengths λ between 420 and 720 nm. We evaluated single-image upsampling with a magnification factor of 2. In order to compare to the inter-channel prior of Farsiu *et al.* [9], we also extended this approach for an arbitrary number of channels. Our approach generalizes to multispectral data without further modifications of the model. The original and upsampled images for an example dataset are visualized in Fig. 5 by a false-coloring method [17] and by depicting a single band at $\lambda = 600$ nm. We achieved smaller residual noise in homogeneous regions reconstructed by our approach compared to channel-wise upsampling. In the image region where the band at $\lambda = 600$ nm is depicted, we can observe that the regularization of Farsiu *et al.* [9] erroneously copied structure from other, original channels to the reconstructed channel, as highlighted in Fig. 5. This effect was avoided by our spatially adaptive confidence weighting.

[3] Experiments were performed on an Intel Xeon E3-1245 CPU with 3.4 GHz.

Figure 4: Color super-resolution ($K = 10$ frames, $3\times$ magnification) on the *bookcase* dataset [10]. From left to right: original frame, super-resolution applied channel-wise without inter-channel regularization, inter-channel regularization proposed in [9] and our method.

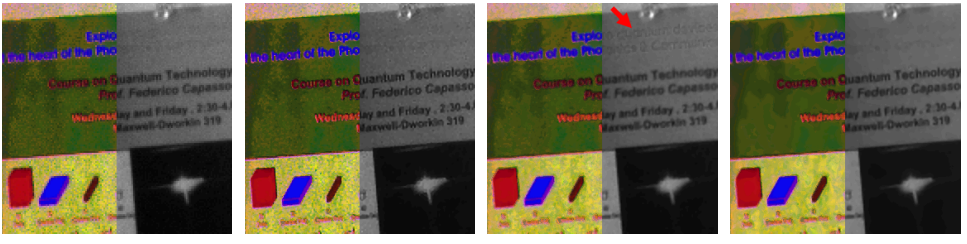


Figure 5: Upsampling ($2\times$ magnification) of multispectral data with $n = 31$ channels [5]. From left to right: original image, channel-wise upsampling, upsampling using inter-channel regularization in [9] and our method. Multispectral data is depicted by false-color visualization [17] (left), and by a single image band centered at wavelength $\lambda = 600\,nm$ (right).

4.2 Application to 3-D Range Imaging

We examined joint upsampling in RGB-D imaging using the Middlebury datasets [16] for a quantitative evaluation on synthetic data. For each ground truth range and color image, low-resolution versions were generated by simulating a Gaussian PSF ($\sigma_{\text{PSF}} = 0.5$) and sub-sampling by a factor of 3. To simulate conditions of low-cost devices, RGB and range images given in the range $[0;1]$ were corrupted by additive Gaussian noise with $\sigma_i = 0.015$ and $\sigma_i = 0.04$, respectively. Our method was compared to channel-wise upsampling without inter-channel prior. In addition, we evaluated guided upsampling proposed by He *et al.* [15] as a state-of-the-art technique, where range images are filtered under the guidance of the color image followed by bicubic interpolation. For RGB data, bicubic interpolation was directly applied to the low-resolution image. The BTV parameters were set to $L = 2$ and $\alpha = 0.7$ and for the LLR model we set $r = 2$ and $\varepsilon = 10^{-4}$. We used two intra-channel regularization weights λ_i for the color and the range channels and one inter-channel regularization weight μ across all channels that were adjusted on one training dataset. Upsampled data was compared for ten randomly generated images per dataset as depicted in Tab. 1.

Multi-channel upsampling outperformed both, guided as well as channel-wise upsampling. As shown in Tab. 1, the mean PSNR was enhanced by ≈ 1 dB compared to guided upsampling on range data. Here, our method takes advantage of the fact that color images are upsampled simultaneously, opposed to guided upsampling that directly employes low-resolution color images as guidance. The comparison in Fig. 6 shows that our multi-channel method achieved better reconstructions of smooth surfaces and depth edges. In particular, in comparison to guided upsampling, the LLR model in multi-channel upsampling was less prone to texture-copying artifacts. Our method also enhanced color images measured by the PSNR in Tab. 1.

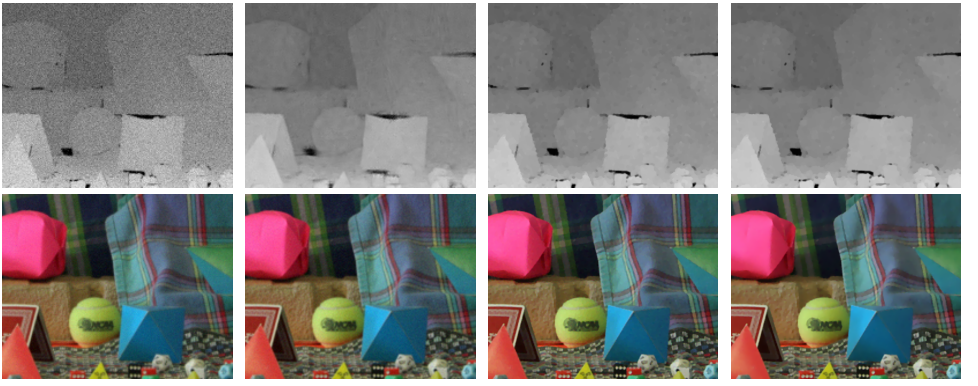


Figure 6: RGB-D upsampling (3× magnification) on the simulated *Moebius* dataset [16]. From left to right: original data, guided upsampling [15] using color images as guidance, channel-wise upsampling without inter-channel prior and our method with inter-channel prior.

Table 1: Peak-signal-to-noise ratio (PSNR) of different upsampling methods on simulated range data (in brackets: RGB data) obtained from the Middlebury 2005 database [16].

Data set	Guided upsampling [15]	Channel-wise	Proposed
Art	27.4 ± 0.01 (32.2 ± 0.01)	28.3 ± 0.03 (33.8 ± 0.02)	**28.5 ± 0.02 (34.1 ± 0.01)**
Books	29.6 ± 0.02 (27.3 ± 0.01)	30.1 ± 0.10 (28.4 ± 0.01)	**30.3 ± 0.03 (28.5 ± 0.01)**
Dolls	28.9 ± 0.02 (29.3 ± 0.01)	30.1 ± 0.11 (30.4 ± 0.01)	**30.3 ± 0.05 (30.6 ± 0.01)**
Laundry	31.3 ± 0.03 (31.3 ± 0.01)	32.1 ± 0.07 (33.2 ± 0.03)	**32.3 ± 0.02 (33.5 ± 0.03)**
Moebius	25.5 ± 0.01 (28.8 ± 0.01)	26.7 ± 0.03 (29.8 ± 0.01)	**26.8 ± 0.02 (29.9 ± 0.01)**
Reindeer	34.5 ± 0.04 (29.2 ± 0.01)	34.8 ± 0.14 (29.6 ± 0.01)	**35.3 ± 0.06 (29.7 ± 0.01)**

5 Conclusion

We proposed a unified approach to multi-frame super-resolution and single-image upsampling of multi-channel images. Our approach is derived from a novel Bayesian model that exploits sparsity of the individual channels as well as a local regression between them. We demonstrated the wide applicability of our method for several common multi-sensor imaging setups ranging from color and multispectral imaging to 3-D range imaging. The proposed method outperforms channel-wise image reconstruction that does not exploit local correlations as well as other state-of-the-art approaches. One of the main novelties is the underlying model being neither limited to a certain imaging setup nor to a fixed number of channels. Furthermore, unlike related work, our approach does not need reliable guidance data.

Future work should consider an adaption of our prior to blind super-resolution where the camera PSF is unknown or other image restoration problems, e. g. image deconvolution. To facilitate open research, Matlab code of our method is provided on our website[4].

Acknowledgments. 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.

[4] http://www5.cs.fau.de/research/software/

References

[1] M. Aguena and N. Mascarenhas. Multispectral image data fusion using POCS and super-resolution. *Computer Vision and Image Understanding*, 102(2):178–187, 2006.

[2] T. Akgun, Y. Altunbasak, and R. Mersereau. Super-resolution reconstruction of hyperspectral images. *IEEE Transactions on Image Processing*, 14(11):1860–1875, 2005.

[3] S. D. Babacan, R. Molina, and A. K. Katsaggelos. Variational bayesian super resolution. *IEEE Transactions on Image Processing*, 20(4):984–999, 2011.

[4] A. V. Bhavsar and A. N. Rajagopalan. Range map superresolution-inpainting, and reconstruction from sparse data. *Computer Vision and Image Understanding*, 116(4): 572–591, 2012.

[5] A. Chakrabarti and T. Zickler. Statistics of Real-World Hyperspectral Images. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 193–200, 2011.

[6] M. Elad and A. Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, 6 (12):1646–1658, 1997.

[7] G. D. Evangelidis and E. Z. Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1858–65, 2008.

[8] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing*, 13(10):1327–1344, 2004.

[9] S. Farsiu, M. Elad, and P. Milanfar. Multiframe demosaicing and super-resolution of color images. *IEEE Transactions on Image Processing*, 15(1):141–159, 2006.

[10] S. Farsiu, D. Robinson, and P. Milanfar. Multi-Dimensional Signal Processing Dataset, 2014. URL http://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html.

[11] R. Fattal. Dehazing using color-lines. *ACM Transactions on Graphics*, 34(1):13, 2014.

[12] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof. Image Guided Depth Upsampling Using Anisotropic Total Generalized Variation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 993–1000, 2013.

[13] 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*, pages 227–238. Springer, 2014.

[14] T. Gotoh and M. Okutomi. Direct super-resolution and registration using raw CFA images. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 600–607, 2004.

[15] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–409, 2013.

[16] H. Hirschmüller and D. Scharstein. Evaluation of cost functions for stereo matching. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[17] J. Jordan, E. Angelopoulou, and A. Robles-Kelly. An unsupervised material learning method for imaging spectroscopy. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 2428–2435, 2014.

[18] M. Kiechle, S. Hawe, and M. Kleinsteuber. A Joint Intensity and Depth Co-sparse Analysis Model for Depth Map Super-resolution. In *2013 IEEE International Conference on Computer Vision*, pages 1545–1552, 2013.

[19] T. Köhler, S. Haase, S. Bauer, J. Wasza, T. Kilgus, L. Maier-Hein, H. Feußner, and J. Hornegger. Tof meets rgb: Novel multi-sensor super-resolution for hybrid 3-d endoscopy. In *Proc. MICCAI 2013*, pages 139–146. Springer, 2013.

[20] T. Köhler, S. Haase, S. Bauer, J. Wasza, T. Kilgus, L. Maier-Hein, C. Stock, J. Hornegger, and H. Feußner. Multi-sensor super-resolution for hybrid range imaging with application to 3-d endoscopy and open surgery. *Medical Image Analysis*, 24(1):220–234, 2015.

[21] J. Mairal, M. Elad, and G. Sapiro. Sparse Representation for Color Image Restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, 2008.

[22] P. Milanfar. *Super-Resolution Imaging*. CRC Press, 2010.

[23] I. T. Nabney. *NETLAB: Algorithms for Pattern Recognition*. Springer, 1st edition, 2002.

[24] I. Omer and M. Werman. Color lines: image specific color representation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 946–953, 2004.

[25] J. Park, H. Kim, Y. Tai, M. S. Brown, and I. Kweon. High quality depth map upsampling for 3D-TOF cameras. In *2011 International Conference on Computer Vision*, pages 1623–1630, 2011.

[26] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman. Overcoming Registration Uncertainty in Image Super-Resolution: Maximize or Marginalize? *EURASIP Journal on Advances in Signal Processing*, 2007:1–15, 2007.

[27] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, 1987.

[28] J. A. Scales and A. Gersztenkorn. Robust methods in inverse theory. *Inverse Problems*, 4(4):1071–1091, 1988.

[29] R. R. Schultz and R. L. Stevenson. A Bayesian approach to image expansion for improved definition. *IEEE Transactions on Image Processing*, 3(3):233–42, 1994.

[30] S. Schuon, C. Theobalt, J. Davis, and S. Thrun. LidarBoost: Depth superresolution for ToF 3D shape scanning. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 343–350, 2009.

[31] H.R. Sheikh and L. Z. Wang. LIVE Image Quality Assessment Database Release 2, 2014. URL http://live.ece.utexas.edu/research/quality.

[32] A. Zomet and S. Peleg. Multi-sensor super-resolution. In *IEEE Workshop on Applications of Computer Vision, WACV 2002*, pages 27–31, 2002.