Bayesian Multispectral Videos Super Resolution

Abdelhamid N. FSIAN*, Jean-Baptiste THOMAS*, Jon Y. HARDEBERG†, Pierre GOUTON*
* IMVIA, Université de Bourgogne
†Colourlab, department of computer science, NTNU
{firstname.lastname}@u-bourgogne.fr* or @ntnu.no†

2023 11th European Workshop on Visual Information Processing (EUVIP) | 979-8-3503-4218-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/EUVIP58404.2023.10323068

***Abstract*—Due to hardware limitations, multispectral videos often exhibit significantly lower resolution compared to standard color videos. These videos capture images in multiple bands of the electromagnetic spectrum, providing valuable additional information that is not available in traditional RGB images. This paper proposes a Bayesian approach to estimate super resolved images from low-resolution spectral videos. We consider adjacent frames from a video sequence to provide a super-resolution image at a time. We include in our proposal the motion between adjacent frames and unlikely to the literature, we estimate the blur and noise while reconstructing the higher resolution image. Experimental results on spectral videos demonstrate the effectiveness of our approach in producing high-quality super resolved images.**

I. Introduction

Multispectral imaging has become a major asset in various fields such as remote sensing, medical imaging, and surveillance. However, the acquired images are often of low resolution, which limits their usefulness in applications that require high-quality images. Super resolution (SR) techniques aim to overcome this limitation by reconstructing high-resolution images from low-resolution ones. In their work, Nasrollahi and Moeslund [1] emphasized the importance of conducting a comprehensive literature review in this particular domain.

While significant progress has been made in the field of spectral image super resolution [2], achieving high-quality super resolution for multispectral video sequences remains a challenging task. Theses challenges arise from various factors, including arbitrary motion of objects and cameras, unknown noise levels, and the presence of motion blur and point spread functions that introduce unknown blur kernels. Prior work often relied on oversimplified assumptions, assuming simple parametric motion forms and known blur kernels and noise levels. However, these assumptions do not hold in practical scenarios, making the super resolution problem more intricate and demanding a more comprehensive approach. Therefore, to develop a practical super resolution system, it is necessary to simultaneously estimate the optical flow [3], noise level [4], and blur kernel [5] ,in addition to reconstructing the high-resolution frames. Since each of these sub-problems has been thoroughly investigated in the field of computer vision, it is natural to integrate them into a unified framework without making oversimplified assumptions.

In this study, we introduce a Bayesian framework using Maximum A Posteriori knowledge (MAP) [6] for reconstructing super resolved multispectral images from low-resolution multispectral videos. Our method takes advantage of the temporal, spatial and spectral correlation between adjacent frames to enhance the super resolution process. Using a sparsity prior for the high-resolution image, flow fields, and blur kernel. The MAP inference iterates between optical flow, noise estimation, blur estimation, and image reconstruction to estimate the optimal values for these parameters. This approach enables us to handle different types of blur kernels and noise models, making our method adaptable and robust in a range of scenarios. Despite different noise levels and blur kernels, our method successfully reconstructs both large-scale structures and small texture features in difficult real-world sequences. The remainder of this paper is organized as follows. In Section 2, we provide a review of related works in the field of spectral super resolution, highlighting the limitations of existing approaches. This review serves as a foundation for understanding the motivation behind our proposed Bayesian framework. Section 3 presents the core of our method, focusing on the image reconstruction process using the Bayesian MAP approach. We describe the key components and their interplay in achieving high-quality super resolution for multispectral video sequences. In Section 4, we present the results obtained from applying our Bayesian framework to real-world multispectral video sequences. We discuss the performance of our method and provide in-depth analysis and discussions on the reconstructed super-resolved images before to conclude.

II. Related Work

A variety of methods have been developed for multispectral super resolution, based on different mathematical models, deep learning architectures, or physical priors. In this section, we provide a review of some of the most representative methods in this field, organized by their underlying principles and approaches.

A. Single Frame Super Resolution

Super resolution from a single frame has been an active area of research in computer vision. Chang et al. [7] made significant contributions to the field, and their work serves as an important foundation. Early research in super resolution addressed the ill-posed problem of reconstructing high-resolution images from low-resolution frames [8]. Schultz et al. [9] utilized spatial priors to overcome the absence of constraints. Bascle et al. [10] considered motion blur using an affine motion model, while Hardie et al. [11] jointly estimated translational motion and the high-resolution image. However, these motion models have limitations in

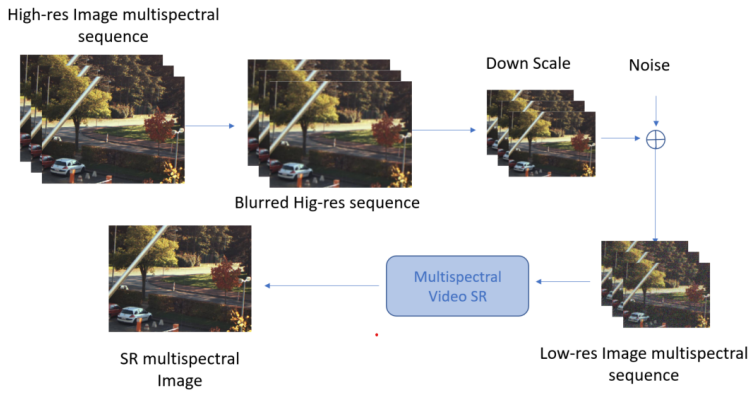


Fig. 1: **super resolution diagram**. To create the observed low-resolution multispectral sequence, the high-resolution sequence is downsampled, each frame is smoothed with a blur kernel, and corrupted with an uncorrelated noise.

capturing the complexity of real-world sequences. Baker and Kanade [12] proposed an approach using optical flow and a parametric motion model to handle motion in super resolution. Fransens et al. [13] introduced a probabilistic framework based on Expectation-Maximization algorithm, but their assumptions about blur kernels and Gaussian priors may affect edge preservation. Recent advancements in optical flow, such as those presented by Brox et al. [14], have provided more reliable techniques based on sparsity priors. On a different scoop, deep learning-based approaches have shown promising results in super resolution of images and videos. However, these methods require a large amount of training data and are computationally expensive [15]. Moreover, Multispectral database are still limited, compared to the RGB database [16].

B. Multi Frame Super Resolution

Super resolution techniques that utilize video sequences or multiple frames have also been explored. Irani et al. [8] proposed a method for enhancing the physical spatial resolution of multispectral images through logical reallocation of spectra. Takeda et al. [17] employed 3D kernel regression inspired by the non-local means technique for video denoising, exploiting spatiotemporal neighboring relationships for video up-sampling. While their technique still requires motion estimation in locations with significant motion, it offers a different approach to super resolution from video sequences. Liu and Freeman [18] developed a video denoising method with accurate motion estimation despite heavy noise. We aim to leverage these advancements in optical flow for more precise super resolution in our work.

In the domain of multispectral image super resolution, Vega et al. [19] proposed a Bayesian approach for super-resolution reconstruction of multispectral images using pansharpening. Pansharpening is the process of fusing high-resolution panchromatic and low-resolution multispectral images to create a single high-resolution color image. The proposed method incorporates prior knowledge on the expected characteristics of multispectral images, including smoothness within each band and correlation between bands. Zhi-Wei et al. [20] introduced an algorithm called SRIF (Multispectral Image Super-Resolution via RGB Image Fusion and Radiometric Calibration) that fuses low-resolution multispectral images with high-resolution RGB images to reconstruct high-resolution multispectral images. However, the linear relationship assumption between multispectral and RGB images may not always hold true, and the requirement of both low-resolution multispectral and high-resolution RGB images can be limiting. Lanaras et al. [21] developed a convex optimization method for improving the spatial resolution of lower-resolution bands in multispectral images. Their adaptive regularizer preserves edges and learns discontinuities, assuming these discontinuities are located in the same positions across all bands, which may not always hold true in practice.

III. Multispectral Super Resolution: Imaging Pipeline and Parameter Estimation

Our objective is to recover the high resolution sequence $\{I\}$ from the low resolution sequence $\{J\}$. In order to take advantage of multispectral video, we attempt to estimate the super resolved frames I_i using the neighboring low resolution frames J_{i-1}, J_i, J_{i+1}. We consider that the low resolution frame J_i is the result of down-sampling of I_i, smoothed with a blur kernel and corrupted with noise. Furthermore, we assume that the noise and kernel blur are consistent across all spectral bands. It simplifies the estimation process and reduces computational complexity. Thus the model of obtaining J_i is illustrated in Figure 1.

In order to estimate the high-resolution sequence and reverse the decay of resolution mentioned above, we need to estimate the noise level, the blur kernel and the motion. Among the unknown parameters in the generative models, the smoothing kernel K, which is equivalent to point spread functions in the imaging process or the smoothing filter when video is down scaled, parameter θ_i which controls the noise, and w_i which represent the motion information between consecutive frames. For this we will use a Bayesian method called MAP as defined in [22] and in the equation (1).

$$\{I', K', \{\theta_i\}', \{w_i\}'\} = \underset{I,K,\{\theta_i\},\{w_i\}}{\arg\max} \; p\left(I, K, \{\theta_i\}, \{w_i\} \,|\{J_i\}\right) \tag{1}$$

The model estimates the unknown parameters, such as the smoothing kernel and noise level, using adjacent frames and applies Bayesian MAP inference to find the optimal solution. The goal is to maximize the posterior probability, which is the product of prior and likelihood according to [23], and developed in equation (2).

$$p(I, K, \{\theta_i\}, \{w_i\}|\{J_i\}) \propto p(I)p(K)\prod_i p(w_i)\prod_i p(\theta_i) \tag{2}$$

$$p(J_0|I, K, \theta_0) \prod_{i=-1, i\neq 0}^{1} p(J_i|I, K, \theta_i, w_i)$$

Where i is the mutlispectral frame index. We assume an exponential distribution for the likelihood in order to handle outliers [24] (equation (3)).

$$p\left(J_i|I, K, \theta_i\right) = \frac{1}{Z\left(\theta_i\right)} exp\left\{-\theta_i||J_i - DBF_{w_i}I||\right\}, \tag{3}$$

where $Z\left(\theta_i\right) = \left(2\theta_i\right)^{-dim(I)}$ and especially the parameter θ_i represents the noise level of frame i. The matrices D and B stand for respectively, down sampling and filtering with kernel blur K. Moreover, F_{w_i} is the warping matrix that correspond to the flow w_i.

To model the priors of image I, optical flow field w_i and blur kernel K, we used sparsity on derivative filter responses. Sparsity on derivative filter responses is a technique used to model the priors of image, optical flow field, and blur kernel in the Bayesian model for super resolution. The sparsity constraint encourages the filter responses to be mostly zero, except for a few significant values, which helps to reduce noise and improve the accuracy of the estimation. This technique is commonly used in signal processing and computer vision applications [25] to promote efficient and robust representations of signals and images (See equations (4) , (5) , (6)). For more mathematical explanation, we refer the reader to this paper [26].

$$p\left(I\right) = \frac{1}{Z_I\left(\eta\right)} exp\left\{-\eta||\nabla I||\right\} \tag{4}$$

$$p\left(w_i\right) = \frac{1}{Z_w\left(\lambda\right)} exp\left\{-\lambda\left(||\nabla u_i||+||\nabla v_i||\right)\right\} \tag{5}$$

$$p\left(K\right) = \frac{1}{Z_K\left(\gamma\right)} exp\left\{-\gamma||\nabla K||\right\} \tag{6}$$

Where $Z_I(\eta)$ (equation (4)), $Z_w(\lambda)$ (equation (5)), and $Z_K(\gamma)$ (equation (6)) are normalization constants that depends only on η, λ and γ. Moreover, ∇ (equation (7)) is defined as the gradient, by extension

$$\begin{aligned} ||\nabla I|| &= \sum ||\nabla I(n)|| \\ &= \sum \left(|I_x(n)|+|I_y(n)|\right) \end{aligned} \tag{7}$$

where $I_x = \frac{\partial}{\partial x} I$, $I_y = \frac{\partial}{\partial y} I$ and n is the pixel index. The flow field's horizontal and vertical components, u_i and v_i, uses the same notation. As proposed by Liu et al. [4]. We assumed that the conjugate prior for θ_i is a Gamma Distribution (equation (8)) :

$$p\left(\theta_i; \alpha, \beta\right) = \frac{\beta^\alpha}{\Gamma\left(\alpha\right)} \theta_i^{\alpha-1} exp\left\{-\theta_i\beta\right\}. \tag{8}$$

Now that we have the prior and likelihood probability distributions, we can use coordinate descend to do the Bayesian MAP inference. Please note that the model have five free parameters which are $\eta,\ \lambda,\ \gamma,\ \alpha,\ \beta$.

A. Reconstructing High-Resolution Images

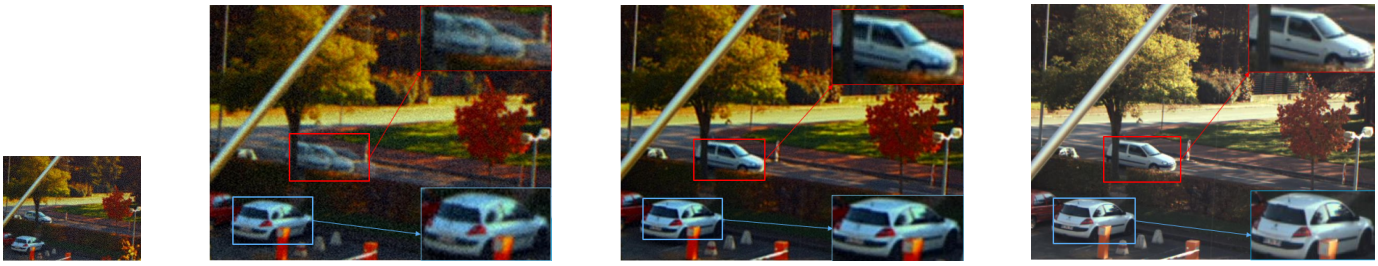
We estimate the high resolution multispectral image by calculating the equation (9), using the most recent estimations for the flow field w_i, the blur kernel K, and the noise level θ_i

$$\begin{aligned} I' = \arg\min_I \theta_0||DBI - J_0|| \\ +\eta||\nabla I||+ \sum_{i=-N, i\neq 0}^{N} \theta_i||DBF_{w_i}I - J_i|| \end{aligned} \tag{9}$$

The term $\theta_0||DBI - J_0||$ measures the data fidelity or the discrepancy between the down sampled and blurred low-resolution image DBI and the observed low-resolution multispectral image J_0. Moreover, the term $\eta||\nabla I||$, enforces sparsity on the gradient of the estimated high-resolution image I. This term promotes smoothness in the image while preserving edges and details. To use gradient-based methods, we replace the $L1$ norm with a differentiable approximation, which is $\xi\left(x^2\right) = \sqrt{x^2 + \epsilon^2}$ with $\epsilon = 0.001$. Moreover, the parameter η controls the strength of the sparsity regularization. A sum over adjacent frames with i indicating current frame index, where DBF_{w_i} denotes convolution operation using blur kernel estimate at ith frame index; this part encourages consistency between consecutive frames. The overall objective function aims to minimize differences between low resolution input and high resolution output while also incorporating constraints related to motion estimation. To solve this objective function, the iteratively reweighted least squares (IRLS) technique is employed. The IRLS [27] algorithm iteratively estimates the high-resolution image I by updating its estimate in each iteration.

B. Noise and Motion Estimation

We jointly estimate the flow field and the noise level on a Gaussian image pyramid knowing the high resolution image and the blur kernel. Therefore, the optical flow and noise level are iteratively evaluated for each pyramidal level. In the context of super-resolution, the Gaussian image pyramid is created by successively applying Gaussian blurring and down sampling operations to the original high-resolution image. Each level of the pyramid represents a different scale or resolution of the image. The coarsest level is the lowest resolution image, and as we move up the pyramid, the



(a) LR image (b) SR Bicubic (c) MAP (d) HR Image

Fig. 2: Our SR method is able to recover image details with a $\times 2$ upsacale.

resolution increases [28]. The following equation is the closed-form solution for the Bayesian MAP estimate for the noise parameter $\theta_i' = \frac{\alpha+N_q-1}{\beta+N_q\bar{x}}$.

Where $\bar{x} = \frac{1}{N_q}\sum_{q=1}^{N_q}|(J_i - BDF_{w_i}I)(q)|$ is defined as a sufficient statistic used to estimate the noise level in the Bayesian MAP approach (following the convention in [28]). Once the noise is known, the flow field w_i is computed using MAP and IRLS technique as depicted in equation (10).

$$w_i' = \arg\min_{w_i} \theta_i ||BDF_{w_i}I - J_i|| + \lambda||\nabla u_i|| + \lambda||\nabla v_i|| \quad (10)$$

The objective function is a weighted sum of data fidelity and regularization terms, where the regularization term $\lambda(||\nabla u_i|| + ||\nabla v_i||)$ enforces smoothness of the motion field and the high-resolution image. The optimization problem is solved iteratively using the IRLS method, which alternates between solving a linear system and updating the weights based on the current estimate. Here again we approximate the norm $|x|$ by $\xi\left(x^2\right)$

C. Blur Kernel Estimation

Following the notation from [23], we only demonstrate how to estimate the x-component kernel K_x given I and J_0 without losing generality and assuming that the kernel K is $x-$ and $y-$separable : $K = K_x \otimes K_y$, where K_y probability distribution is the same as K_x. Therefore, we define each row of the matrix A as the concatenation of pixels that correspond to the filter K. Moreover, we define $M_y : M_y K_x = K_x \otimes K_y = K$. The estimation of K_x is depicted in equation (11).

$$K_x' = \arg\min_{K_x} \theta_0 ||AM_yK_x - J|| + \gamma||\nabla K_x|| \quad (11)$$

The method involves solving an optimization problem that minimizes the difference between the low-resolution observation and the convolution of the high-resolution image with the estimated kernel, subject to a regularization term that encourages spatial smoothness of the kernel. The optimization problem is solved using the IRLS method.

IV. Results and Discussions

In our study on multispectral SR, we utilized a publicly available database from Benezeth et al. [29], which contains a collection of five VNIR (Visible and Near-InfraRed) multispectral videos containing between 250 and 2300 frames

Fig. 3: Sample of three consecutive images from different videos. The video goes from left to right.

of spatial resolution 658×491. Each video in the database consists of a sequence of frames. Each frame contains seven spectral bands of which six are in the visible and one in the near infra-red (NIR), illustrations of the data-set are presented in Figure 3.

The inclusion of the NIR band allows for enhanced perception and analysis of various materials and phenomena that may exhibit distinct spectral characteristics in this region. The dataset provided a valuable resource for evaluating and benchmarking our proposed multispectral SR algorithm, enabling us to assess its performance across different spectral

TABLE I: Comparison of ESRGAN and MAP SR using average PSNR, SSIM, and RMSE from the Videezy4K dataset

Method	PSNR	SSIM	RMSE
ESRGAN $\times 2$	32.09	0.8793	0.00634
Ours $\times 2$	30.54	0.7467	0.076

bands and video sequences.

In our experimental analysis, we conducted a serie of evaluations to assess the performance and effectiveness of our proposed MAP for SR. To create a realistic simulation of real-world imaging conditions, we initially applied a blur operation followed by downscaling to the multispectral sequence. [1] This step aimed to mimic the inherent limitations and degradation commonly encountered in practical imaging scenarios. Furthermore, to account for the presence of noise in real data, we introduced a Gaussian noise into the blurred and downscaled images. Figure 2 shows respectively, the low resolution image, Bicubic SR image, MAP SR image (with a $\times 2$ upscale) and HR image.

In order to demonstrate the effectiveness of our Bayesian approach, we compared its results against those obtained using conventional interpolation methods, including Bicubic, nearest neighbor, and bilinear interpolation. By showcasing the comparative results, we highlight the distinct advantages and improvements achieved by our proposed Bayesian method in terms of both quantitative metrics and visual quality. We selected a set of 7 consecutive frames from each VNIR video containing scenes with moving objects. We considered factors such as object motion, scene complexity, and spectral diversity during the selection process. Specifically, we aimed to include scenes with diverse spectral content and varying degrees of motion to evaluate the performance of our SR MAP algorithm comprehensively. Furthermore, inspired by the versatility of our multispectral SR algorithm, we explore its applicability to RGB images. Unlike traditional SR algorithms that handle all channels simultaneously, our method treats each channel independently. To substantiate our findings, we conducted a comparison with a well-established state-of-the-art SR algorithm: ESRGAN [30] . We employed the RGB dataset from Videezy4K, a benchmark dataset renowned for its diverse and challenging image content. A dataset that contains 11 RGB videos and each videos contains 19 4K RGB images.

Comprehensive metrics including SSIM, PSNR and RMSE are provided in Table II for VNIR videos. The results from our experiments on the RGB dataset sourced from the Videezy 4K dataset are summarized in Table I.

The experimental results clearly demonstrate a significant drop in performance when motion estimation is omitted from the super-resolution process. Without motion estimation, the algorithm fails to capture the temporal coherence between frames, resulting in reduced image quality and an inability to effectively compensate for motion-related artifacts. On the other hand, incorporating motion estimation enables the algorithm to align frames and accurately estimate motion, leading to improved reconstruction quality and better preservation of fine details. These findings highlight the crucial role of motion estimation in achieving enhanced multispectral super-resolution by leveraging temporal information and mitigating artifacts.

V. Conclusion

In this study, we conducted a comprehensive analysis of multispectral super resolution utilizing a publicly available VNIR video database. Our research focused on developing a novel Bayesian method for spectral image SR and compared its performance against conventional interpolation techniques, namely Bicubic, nearest neighbor, and bilinear interpolation. Additionally, we extended our investigation to include a comparison with a state-of-the-art method, ESRGAN, which employs RGB image sequences. The results of this comparative analysis showcased the promising performance of our proposed algorithm. By leveraging neighboring frames to enhance the reference frame without the need for any training. This approach allowed us to achieve improved super-resolution results and narrowed the performance gap between our method and ESRGAN, highlighting the potential of our approach in the field of multispectral super resolution.

In our experiments, incorporating motion estimation, demonstrated superior performance compared to conventional interpolation techniques. By effectively capturing temporal coherence and reducing motion-related artifacts, our algorithm achieved improved reconstruction quality and preservation of fine details. The experimental results underscored the significance of motion estimation in multispectral SR, as neglecting this crucial aspect significantly impacted the overall performance. Our research contributes to the understanding of multispectral SR techniques, highlighting the importance of leveraging temporal information to enhance the desired frame.

For future work, an interesting avenue would be to explore the integration of joint super-resolution and demosaicing. Combining these two tasks could lead to more comprehensive image enhancement, addressing both spatial and color resolution simultaneously. By leveraging the strengths of both techniques, it is possible to achieve further improvements in the visual quality and fidelity of spectral images.

References

[1] K. Nasrollahi and T. B. Moeslund, "Super-resolution: a comprehensive survey," *Machine vision and applications*, vol. 25, pp. 1423–1468, 2014.

[2] J. M. Amigo, "Practical issues of hyperspectral imaging analysis of solid dosage forms," *Analytical and bioanalytical chemistry*, vol. 398, pp. 93–109, 2010.

[3] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[4] C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, and W. T. Freeman, "Automatic estimation and removal of noise from a single image," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 2, pp. 299–314, 2007.

[5] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE signal processing magazine*, vol. 13, no. 3, pp. 43–64, 1996.

[6] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz, and R. Hanson, "Super-resolved surface reconstruction from multiple images," *Maximum Entropy and Bayesian Methods: Santa Barbara, California, USA, 1993*, pp. 293–308, 1996.

PSNR								
	MAP		Bicubic		Bilinear		Nearest	
	w motion	o motion	w motion	o motion	w motion	o motion	w motion	o motion
VS 1	**35,79**	34.96	27.33	26.21	27.24	2565	26.36	24.52
VS 2	**34.6**	33.51	31.87	30.79	30.65	29.54	30.49	29.44
VS 3	**33.44**	32.27	29.46	26.34	29.08	25.49	28.93	24.25
VS 4	**31.89**	31.76	27.58	25.74	27.77	26.50	26.61	24.47
VS 5	**30.60**	30.41	27.32	26.21	27.23	25.65	26.36	24.52
SSIM								
VS 1	**0.95**	0.934	0.7427	0.6426	0.727	0.572	0.632	0.482
VS 2	**0.83**	0.78	0.76	0.73	0.75	0.7	0.74	0.59
VS 3	**0.87**	0.86	0.79	0.61	0.769	0.52	0.75	0.42
VS 4	**0.88**	0.87	0.70	0.55	0.74	0.64	0.62	0.46
VS 5	**0.87**	0.84	0.74	0.64	0.70	0.57	0.63	0.48
RMSE								
VS 1	**0.0162**	0.018	0.042	0.048	0.043	0.052	0.048	0.0599
VS 2	**0.02**	**0.02**	0.025	0.029	0.026	0.029	0.028	0.03
VS 3	**0.026**	0.024	0.033	0.048	0.035	0.053	0.035	0.061
VS 4	**0.025**	0.0258	0.041	0.051	0.04	0.047	0.046	0.059
VS 5	**0.029**	0.03	0.043	0.048	0.043	0.052	0.048	0.0593

TABLE II: Performance Metrics(PSNR, SSIM, RMSE) for each method with and without motion estimation and for each of the 5 VNIR video.

[7] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1. IEEE, 2004, pp. I–I.

[8] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical models and image processing*, vol. 53, no. 3, pp. 231–239, 1991.

[9] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE transactions on image processing*, vol. 5, no. 6, pp. 996–1011, 1996.

[10] B. Bascle, A. Blake, and A. Zisserman, "Motion deblurring and super-resolution from an image sequence," in *Computer Vision—ECCV'96: 4th European Conference on Computer Vision Cambridge, UK, April 15–18, 1996 Proceedings Volume II 4*. Springer, 1996, pp. 571–582.

[11] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, 1997.

[12] S. Baker and T. Kanade, "Super-resolution optical flow," Carnegie Mellon University, The Robotics Institute, Technical Report, 1999.

[13] R. Fransens, C. Strecha, and L. Van Gool, "Optical flow based super-resolution: A probabilistic approach," *Computer vision and image understanding*, vol. 106, no. 1, pp. 106–115, 2007.

[14] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part IV 8*. Springer, 2004, pp. 25–36.

[15] S. Liu, Y. Yang, Q. Li, H. Feng, Z. Xu, Y. Chen, and L. Liu, "Infrared image super resolution using gan with infrared image prior," in *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*. IEEE, 2019, pp. 1004–1009.

[16] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer, 2014, pp. 740–755.

[17] H. Takeda, P. Milanfar, M. Protter, and M. Elad, "Super-resolution without explicit subpixel motion estimation," *IEEE Transactions on Image Processing*, vol. 18, no. 9, pp. 1958–1975, 2009.

[18] C. Liu and W. T. Freeman, "A high-quality video denoising algorithm based on reliable motion estimation," in *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part III 11*. Springer, 2010, pp. 706–719.

[19] M. Vega, J. Mateos, R. Molina, and A. K. Katsaggelos, "Super resolution of multispectral images using l1 image models and interband correlations," *Journal of Signal Processing Systems*, vol. 65, no. 3, pp. 509–523, 2011.

[20] Z.-W. Pan and H.-L. Shen, "Multispectral image super-resolution via rgb image fusion and radiometric calibration," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1783–1797, 2018.

[21] C. Lanaras, J. Bioucas-Dias, E. Baltsavias, and K. Schindler, "Super-resolution of multispectral multiresolution images from a single sensor," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 20–28.

[22] M. Tipping and C. Bishop, "Bayesian image super-resolution," *Advances in neural information processing systems*, vol. 15, 2002.

[23] C. Liu and D. Sun, "A bayesian approach to adaptive video super resolution," in *CVPR 2011*. IEEE, 2011, pp. 209–216.

[24] N. Wang and D. Yeung, "Bayesian robust matrix factorization for image and video processing," in *IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, December 1-8, 2013*. IEEE Computer Society, 2013, pp. 1785–1792. [Online]. Available: https://doi.org/10.1109/ICCV.2013.224

[25] M. F. T. B. C. Russell and W. T. Freeman, "Exploiting the sparse derivative prior for super-resolution and image demosaicing," in *Proceedings of the Third International Workshop Statistical and Computational Theories of Vision*, 2003, pp. 1–28.

[26] H. S. Mousavi and V. Monga, "Sparsity-based color image super resolution via exploiting cross channel constraints," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5094–5106, 2017.

[27] C. Liu, "Beyond pixels: exploring new representations and applications for motion analysis," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[28] Y. Kwon, S. Kim, D. Yoo, and S.-E. Yoon, "Coarse-to-fine clothing image generation with progressively constructed conditional gan." in *VISIGRAPP (4: VISAPP)*, 2019, pp. 83–90.

[29] Y. Benezeth, D. Sidibé, and J.-B. Thomas, "Background subtraction with multispectral video sequences," in *IEEE International Conference on Robotics and Automation workshop on Non-classical Cameras, Camera Networks and Omnidirectional Vision (OMNIVIS)*, 2014, pp. 6–p.

[30] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.