Camera Array for Multi-Spectral Imaging

Nils Genser, *Graduate Student Member, IEEE*, Jürgen Seiler, *Senior Member, IEEE*, and André Kaup, *Fellow, IEEE*

***Abstract*— Recently, many new applications arose for multi-spectral and hyper-spectral imaging. Besides modern biometric systems for identity verification, also agricultural and medical applications came up, which measure the health condition of plants and humans. Despite the growing demand, the acquisition of multi-spectral data is up to the present complicated. Often, expensive, inflexible, or low resolution acquisition setups are only obtainable for specific professional applications. To overcome these limitations, a novel camera array for multi-spectral imaging is presented in this article for generating consistent multi-spectral videos. As differing spectral images are acquired at various viewpoints, a geometrically constrained multi-camera sensor layout is introduced, which enables the formulation of novel registration and reconstruction algorithms to globally set up robust models. On average, the novel acquisition approach achieves a gain of 2.5 dB PSNR compared to recently published multi-spectral filter array imaging systems. At the same time, the proposed acquisition system ensures not only a superior spatial, but also a high spectral, and temporal resolution, while filters are flexibly exchangeable by the user depending on the application. Moreover, depth information is generated, so that 3D imaging applications, e.g., for augmented or virtual reality, become possible. The proposed camera array for multi-spectral imaging can be set up using off-the-shelf hardware, which allows for a compact design and employment in, e.g., mobile devices or drones, while being cost-effective.**

***Index Terms*— Multi-spectral imaging, image acquisition.**

I. Introduction

THE measurement of light in a number of spectral bands is known as multi-spectral imaging. Traditional color imaging can be seen as a simple representative by recording red, green, and blue color data. Consequently, it is possible to cover the same information as the human visual system does. Nevertheless, one can also acquire more spectral bands to reveal further information humans can't experience. In the last decades, numerous fields of application have been studied, which benefit from the measurement of multi-spectral information. Some popular use-cases shall be shortly reviewed in the following to collate common challenges and to motivate the need of a novel multi-spectral imaging system.

A. Applications

In the past years, biometric identification systems have been intensively studied to include multi-spectral imaging as security feature [1]. There, images of palm-dorsa veins are recorded in the near-infrared or infrared range, which lead to a unique pattern for every human. In combination with other meaningful features, e.g., iris scans, highly secure authentication systems are realized [2]. In Fig. 1, an example for vein detection is depicted that was recorded utilizing the proposed Camera Array for Multi-Spectral Imaging (CAMSI). Recently, also the food industry is exploring solutions to enhance quality controls using multi-spectral acquisition systems [3] to identify foreign objects and substances, which contaminate the food during production. Other applications deal with the analysis of paintings using multi-spectral imaging [4]. Hence, visual and measurement-based quantitative scientific analysis can be conducted to evaluate the quality of art, or the time of origin. Moreover, in agriculture and geoinformatics, a lot of applications recently came up, e.g., classification of different types of crops [5], or monitoring agricultural field conditions [6]. Other methods estimate coastal water depth [7], do land cover classification in urban and rural landscape [8], or measure the temperature of different seashore sections [9]. In medical applications, multi-spectral imaging is applied to detect distressed persons, e.g., in swimming lakes [10]. Regarding the human body itself, one can evaluate the in vivo microcirculation by investigating the vessels [11]. It is possible to investigate humans for skin diseases and measure the overall health condition [12]. Other examples deal with the estimation of the heart rate [13], or the measurement of the extent of dermal perfusion [14].

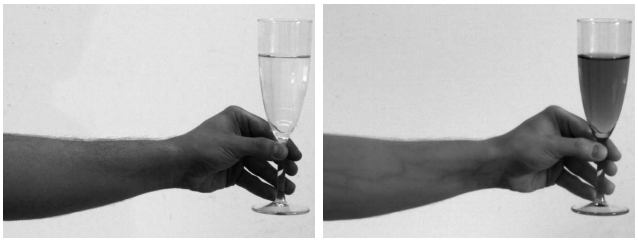

Fig. 1. Liquids detection and security feature tracking are typical applications for multi-spectral imaging. The pictures were recorded and processed using the proposed CAMSI acquisition system. On the left, the green component is shown, while a 950 nm bandpass image is depicted on the right. The veins of the arm become visible and the liquid changes its brightness.

Even though there are many applications, typical similarities regarding the requirements of the multi-spectral acquisition system can be formulated. For one thing, changing filters depending on the requirements is preferable, as the development of applications can be significantly fastened. Finding a suitable set of filters for the specific use case is not easy, so that most applications benefit from a dynamic filter adaption [15]. Secondly, many designs find their way to mass-market only

Manuscript received July 22, 2019; revised March 24, 2020; accepted September 5, 2020. Date of publication September 24, 2020; date of current version September 30, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Trac D. Tran. *(Corresponding author: Nils Genser.)*

The authors are with the Chair of Multimedia Communications and Signal Processing, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91058 Erlangen, Germany (e-mail: nils.genser@fau.de; juergen.seiler@fau.de; andre.kaup@fau.de).

Digital Object Identifier 10.1109/TIP.2020.3024738

if the size of the acquisition system is compact and handy. To reach a wide distribution of the system, also price plays a crucial role. Likewise, many health and agricultural applications require or benefit from the acquisition of image as well as video data. Of great importance is also the spatial resolution, which is a limiting factor for many systems as even a resolution of half a mega pixel is not reached for many acquisition principles [16].

B. Outline of the Paper and Contributions

The manuscript is organized as follows. In Sec. II, pros and cons of existing multi-spectral imaging techniques are examined. To overcome limitations, the novel CAMSI approach is introduced in Sec. III. Besides the hardware description, the problem statement as well as the versatility and extensibility of the novel approach are discussed. In Sec. IV calibration and rectification of the multi-camera setup are depicted, before the problem of cross-spectral registration is denoted in Sec. V. There, the relation between disparity and depth is briefly reviewed and the effect of cross-spectral extinction is presented to motivate the need of a novel registration approach. After introducing the global registration, a general formulation of the pixel transformation is given that demands for only one resampling. Next, recovery of occluded and mispredicted pixels is tackled by a novel cross-spectral reconstruction in Sec. VI. To demonstrate the performance of the overall CAMSI approach, an extensive evaluation is given in Sec. VII. For example, the registration performance of state-of-the-art methods and CAMSI as well as an analysis to other multi-spectral imaging techniques are given. In line with this, visual examples and an evaluation on computational complexity are depicted. The paper is briefly concluded in Sec. VIII.

In this paper, the following contributions are introduced:

- A novel geometrically constrained multi-camera sensor layout is presented together with all required algorithms to acquire high-quality multi-spectral images.
- State-of-the-art multi-modal stereo matching and the resulting problem of cross-spectral extinction is discussed.
- A novel global registration is formulated for geometrically constrained multi-spectral camera arrays.
- In line with this, CAMSI is compared to state-of-the-art cross-spectral stereo-matching methods highlighting the global cost aggregation as key feature.
- A fast cross-spectral reconstruction algorithm is introduced, which takes all spectral components into account for reconstructing occluded and mispredicted pixels.
- For CAMSI and several registration methods as well as imaging systems, an extensive evaluation is conducted.
- In contrast to state-of-the-art approaches, an overall processing chain is provided. Starting with the sensor layout, also calibration, registration, reconstruction, and direct mesh to grid resampling strategies are presented.

II. Related Multi-Spectral Imaging Techniques

More than 40 years ago, the rise of digital color imaging began with the publication of the Bayer filter [17], which brought color cameras to the mass-market by placing red, green, and blue color filter arrays in front of the camera sensor. Up to now, this principle was maintained, so that the Bayer filter can be found in professional equipment as well as in everyday life, e.g., smartphones, or cars. In line with this, multi-spectral imaging can be defined as measurement of 3 to 15 spectral bands. Recording more than three spectral bands is complicated and briefly summarized in the following.

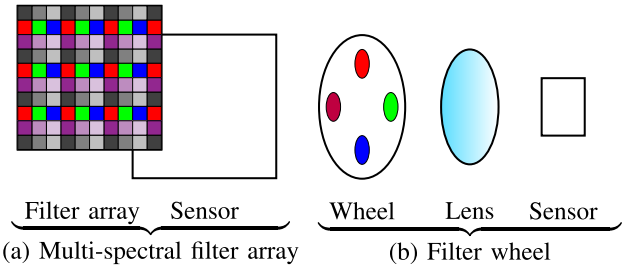

Fig. 2. On the left-hand side, a multi-spectral array is depicted. There, a 3×3 filter pattern is overlaid and repeated to cover the whole image sensor. The filter array is integrated in front of the image sensor. On the right-hand side, the structure of an exemplary filter wheel is shown.

State-of-the-art approaches use either Charge-Coupled Device (CCD) [18], [19] or CMOS Active Pixel (CAP) sensors [20], [21] to measure incident light. Consequently, either a CCD or CMOS sensor forms the backbone of a modern multi-spectral imaging system. To separate information into multiple color components, different strategies can be deployed.

Similar to Bayer color imaging, it is common to use a Multi-Spectral Filter Array (MSFA) in front of the sensor by including more filter elements into the mosaic [22]–[24]. For a better overview, Fig. 2a depicts the structure of an exemplary MSFA with nine different filters. In the last years, many efforts were made to improve MSFAs [25], especially regarding filter properties and technical feasibility. For example, silicon nanowires can be used to control spectral filter properties [26], or optical filters are integrated monolithically at wafer level to measure 16, 36, or even more spectral components [27]. Image and video data is recordable, but the spatial resolution of every spectral component is significantly reduced, which can lead to strong aliasing. Hence, MSFAs always have to be optimized by trading spatial versus spectral resolution, which makes the measurement of many subbands difficult. Moreover, the selection of filters and acquirable spectral bands is fixed and can't be changed in hindsight. As with widely-distributed Bayer sensors, a mass-production can be realized very cost-effective, while small custom-designed batch sizes become highly expensive. MSFAs are well-suited for applications that tackle the consumer market, but they are difficult to use for industrial as well as professional applications and experimental settings. Furthermore, MSFAs can't provide depth information, so that 3D imaging is not possible without using additional hardware. An advantage of MSFAs is the compact design, which makes them handy for mobile devices. Related to MSFAs, also pushbroom acquisition techniques [28], [29] that work as a line scanning camera are common. By integrating different spectral filters in front of each row of a monochrome sensor, the acquisition of multi-spectral data is achieved. Pushbroom imaging devices are especially suited for moving

camera scenarios, e.g., airborne recordings [30], as motion can be used to allow for spatial imaging.

Another intuitive and well investigated class of multi-spectral acquisition systems are Tunable Filters (TF). There, a complete measurement is obtained by recording one image of a spectral band after the other, yielding a sequence of exposures resulting in a multi-spectral image. One ipossible approach is to use a filter wheel (FW) setup [31], [32], where a wheel of filters is mounted in front of a monochromatic image sensor. By revolving the wheel, different subbands can be measured one after the other (see Fig. 2b). An advantage is that spatial resolution is not reduced, but it comes at the price that video acquisition is impeded. However, the setup is flexible, which means that the filters can be exchanged as required by the user. FW setups are suitable for professional applications as well as experimental settings and can be ranked in the mid-price segment. In the end, one has to consider that FW reach a certain size, which often restricts the usage to scientific communities or special commercial applications and prevents mobile usage, respectively. As for MSFAs, FW do not provide depth information without using additional hardware. Two further representatives of TFs are liquid crystal tunable filters (LCTF) and acousto-optical tunable filter arrays (AOTF) [33]. Both allow for electronically controlled spectral filter properties and the transition of filters can be implemented faster in comparison to filter wheels [34].

Besides these methods, a variety of more exotic approaches exists. For example, tunable illumination devices are common in cultural heritage imaging [35]. Beam splitters were used since the 1950s in television cameras [24], which divide light beams into different fractions similar to prisms. Interferometer based techniques [36], filtered lenslet arrays [37], division of focal plane polarimeters [38], spectrometers [39], multi-view multi-spectral mirror systems [40], and tunable sensors [41] can be named, too. These methods are tied to specify problems and are used in few industrial and professional applications.

Recently, also cross-spectral stereo camera approaches came up that combine, e.g., a color and an infrared camera for measuring multiple spectral components [42]. As image content is recorded at different spatial positions, the aim is to register heterogeneous content. Some algorithms focus only on the registration of objects in one depth plane, which results in a global transform and gives no depth information on the scene [43]. A more complicated task is the registration of objects in varying depth levels, which makes it necessary to estimate a 3D scene for mapping the recorded images onto a common viewpoint. For example, one can combine a structural template matching cost function, e.g. Census Transform, and a traditional cost aggregation algorithm like Semi-Global Matching (SGM) to estimate a disparity map that is used for pixel-wise registration [44]. Other approaches deal with the design of further cost metrics, e.g., Robust Selective Normalized Cross Correlation (RSNCC) [45], or Dense Adaptive Self-Correlation (DASC) [46], to enhance registration performance. Recently, also an unsupervised deep-learning based approach was presented for RGB and NIR content, while incorporating knowledge on fixed sets of materials [42]. To acquire multi-spectral data, the number of spectral components has to be increased [47]. However, cross-spectral stereo and multi-camera setups mainly focus on the design of cost metrics for registration. On this occasion, the cost fusion from all available cameras is mostly omitted as well as the problem of reconstructing occluded or mispredicted pixels, although a lot of spatial and spectral information is exploitable for registration and reconstruction.

TABLE I

PROS (+) AND CONS (−) OF MULTI-SPECTRAL IMAGING TECHNIQUES AND THE NOVEL CAMERA ARRAY FOR MULTI-SPECTRAL IMAGING. A MIXED PERFORMANCE IS INDICATED BY CIRCLES (o)

	MSFA	TF	Proposed CAMSI
High spatial resolution	-	+	+
Video capable	+	o/-	+
Exchangeable filters	-	+	+
Price (small batch sizes)	-	+/o	+/o
Price (high batch sizes)	+	-	+/o
Compact size	+	-	+/o
Depth information	-	-	+
Software postprocessing	o	+	o

The breakthrough of multi-spectral imaging is impeded by the different disadvantages of existing acquisition systems. A perfect setup would allow recording high-resolution image and video data, while filters are exchangeable according to the user's demands. Furthermore, the setup has to be operable independent of the filter configuration, e.g., color, bandpass, or polarization filters. Moreover, the system must be price-efficient and suited for the consumer market, as well as inexpensive for small batch sizes. Additionally, a compact construction size is asked to ensure the usage for mobile devices. For 3D imaging applications, e.g., virtual or augmented reality, the derivation of depth information is preferable. Likewise, detection and classification can be significantly simplified by using depth information. A brief summary of demanded features and available acquisition approaches is listed in Tab. I.

III. CAMERA ARRAY FOR MULTI-SPECTRAL IMAGING

Before the algorithmic challenges are discussed in the next sections, the hardware design of the proposed camera array for multi-spectral imaging setup shall be firstly described. Therefore, the placement of the different components together with the camera parameters and the choice of filters are shown. Afterwards, the problem statement is formulated, which arises due to the multi-spectral, multi-camera layout. Furthermore, the versatility and extensibility of CAMSI is examined.

A. Hardware Description

The proposed CAMSI system consists of $K = 9$ cameras, with K being selected flexibly depending on the required number of channels. The different camera positions are denoted as P_k allowing index k in the range 0 to $K - 1$ with center camera position P_c and index $c = 4$. Cameras are aligned on a 3×3 grid, so that the displacement from a peripheral camera to the center is always horizontal, vertical, or diagonal. As a constraint, the camera array's size must be chosen quadratic and the array length must be odd, e.g., 3×3 or 5×5, to align all peripheral views around the center camera. Introducing

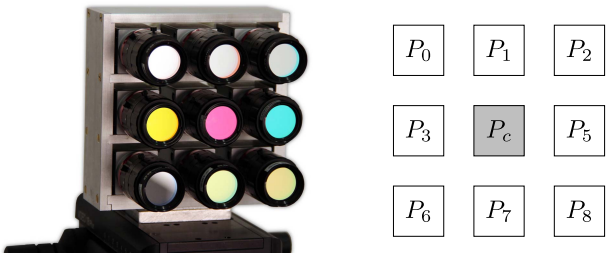

Fig. 3. The proposed CAMSI acquisition system together with the annotated camera views P_0 to P_8 with center view $P_c = 4$. Each camera is mounted with a different color filter.

TABLE II
PARAMETERS OF THE CAMSI SETUP

System Parameters	CAMSI
Camera type	Basler acA1600-60gm
Number of cameras K	9
Baseline B	4 cm / $4\sqrt{2}$ cm
Focal length f	16 mm
Image sensor	e2v EV76C570
Sensor pixel size p	4.5 µm
Resolution	1600×1200 pixels
Frame rate	60 fps
Bit depth	12 bit
Lens transmission	$300 - 1000$ nm

geometric restrictions in the hardware design is required in later software post-processing to reduce computational complexity and to significantly increase registration robustness, as described in Sec. V in detail. The novel CAMSI prototype and the annotated positions P_k are shown in Fig. 3.

The cameras are mounted in a solid aluminum enclosure to ensure a consistent calibration and a good thermal conduction. The baseline between each camera measures 4 cm, so that the distance between each peripheral and the center camera is 4 cm and $4\sqrt{2} \approx 5.7$ cm, respectively. The choice fell on professional monochrome cameras that record images up to a resolution of 1600×1200 pixels at 60 frames per second and a bit depth of 12, while the sensor pixel size measures 4.5 μm. The cameras are connected to a PC via Gigabit Ethernet while being synchronized using the camera trigger feature to ensure that all cameras take the pictures at the same time and still as well as moving content is recordable. Furthermore, the cameras allow for an easy mounting of different lenses, so that a wide range of problems can be tackled with this setup. For CAMSI, 16 mm prime lenses were installed that are convenient for all filters ranging from approx. 300 to 1000 nm. An overview of the parameters of CAMSI is depicted in Tab. II. Professional image processing color filters were used together with steep 50 nm band pass filters from stock to record color, ultra-violet, and near infrared images. The filter configurations used in this paper are depicted in Tab. III.

B. Problem Formulation

When acquiring images using the CAMSI setup, K recordings $\boldsymbol{I}_0$ to $\boldsymbol{I}_{K-1}$ result that measure different spectral components at positions P_0 to P_{K-1} at the same time instance.

TABLE III
TWO EXEMPLARY FILTER CONFIGURATIONS FOR THE EQUIPMENT OF CAMSI

Description	Position	Center frequency	Transmission
Filter configuration *NIR-RGB-UV*			
Steep bandpass	P_0	950 nm	$925 - 975$ nm
Steep bandpass	P_1	850 nm	$825 - 875$ nm
Steep bandpass	P_2	750 nm	$725 - 775$ nm
Blue color filter	P_3	470 nm	$385 - 555$ nm
Green color filter	P_c	524 nm	$432 - 616$ nm
Red color filter	P_5	660 nm	$594 - 726$ nm
Steep bandpass	P_6	400 nm	$375 - 425$ nm
Steep bandpass	P_7	450 nm	$425 - 475$ nm
Steep bandpass	P_8	500 nm	$475 - 525$ nm
Filter configuration *NIR-RGB-NIR*			
Steep bandpass	P_0	800 nm	$775 - 825$ nm
Steep bandpass	P_1	750 nm	$725 - 775$ nm
Steep bandpass	P_2	700 nm	$675 - 725$ nm
Blue color filter	P_3	470 nm	$385 - 555$ nm
Green color filter	P_c	524 nm	$432 - 616$ nm
Red color filter	P_5	660 nm	$594 - 726$ nm
Steep bandpass	P_6	950 nm	$925 - 975$ nm
Steep bandpass	P_7	900 nm	$875 - 925$ nm
Steep bandpass	P_8	850 nm	$825 - 875$ nm

In contrast to this, measurements of different subbands are required at the same spatial position for most applications. Consequently, the goal is to calculate a virtual center view of all recordings by applying a pixel-wise transformation to each image, which is possible as long as depth information is disposable and the distances between the cameras are known. Hence, depth must be calculated to warp the different peripheral positions onto the center P_c that serves as basis. In principle, deriving depth-information from multi-camera setups is a well-investigated problem statement [48], [49]. Typically, these methods require the same image content in the different views. Obviously, this is not the case for CAMSI as different spectral bands are measured at different camera positions. On the one side, the recording of multi-spectral data impedes the derivation of depth information. On the other side, depth information is needed to register the different subbands and to obtain a multi-spectral image. In contrast to state-of-the-art approaches that introduce further cost functions, this typical chicken-and-egg situation is solved by introducing a novel depth estimation method in Sec. V that combines information from all available cameras to improve accuracy. Moreover, loss areas that arise during the registration due to occlusions and mispredictions are estimated by a novel cross-spectral reconstruction algorithm (see Sec. VI), which also takes all available spectral components into account. An overview of the proposed CAMSI processing chain is given in Fig. 4 together with the required processing steps.

C. Versatility and Extensibility

A core feature of CAMSI is its versatility. Firstly, it is possible to capture a dynamic number of different spectral components by modifying the hardware layout. As long as the peripheral cameras are located around a center view, these geometric constraints are exploitable and result in a robust registration. The proposed sensor layout can be easily extended

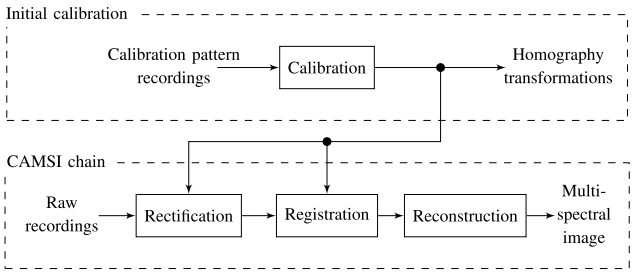

Fig. 4. Overview of the proposed CAMSI chain. The calibration of the camera array is carried out only once. Afterwards, the estimated homography transformations are used to calculate multi-spectral images.

to obtain hyper-spectral images by using 5×5, 7×7 or even larger array grid sizes. Secondly, the combination and arrangement of filters is independent of each other as a global depth estimation is conducted in Sec. V. Thus, the array can be set up without taking the filter transmissions or further a priori knowledge into account. Thirdly, the filter types can be arbitrarily chosen as long as there is an overlap of the spectrum of some filters. Hence, the design is highly adaptable to any specific user requirements.

IV. Calibration and Rectification

In practical implementation, all cameras exhibit arbitrary displacements to each other. On the one side, it is not possible to perfectly align the cameras onto the array grid. On the other side, the image sensor may not be installed precisely enough in the camera body itself. Consequently, all cameras in the array must be calibrated with respect to the central camera before a registered multi-spectral image can be calculated. In the following, it is assumed that the intrinsic camera parameters, e.g., lens distortions, are already compensated using methods like [50] or [51]. Thus, only the extrinsic calibration due to manufacturing tolerances is discussed.

After recording the different images, the corresponding pixels are misplaced both in horizontal and vertical direction at the same time. In a perfect setup, the displacement should be purely horizontal, vertical, or diagonal, which is not applicable in practice. The aim of the proposed calibration is to align the views back on the virtual CAMSI array grid and to resolve the arbitrary displacement to a one-dimensional disparity. Consequently, the disparity must be still determined on the epipolar lines, later. The remaining one-dimensional disparity remains due to the objects being placed in different depths, which yields pixel dependent displacements. An example for the middle row of the CAMSI array is given in Fig. 5, where the composition of $\{\boldsymbol{I}_5, \boldsymbol{I}_\mathrm{c}, \boldsymbol{I}_3\}$ results in a color image { red, green, blue }. On the left-hand side, the uncalibrated images are composed and the shift appears to be arbitrarily two dimensional. On the right-hand side, the calibrated image is depicted. Apparently, the objects are only displaced horizontally after rectification.

Briefly, the planar homography relates to the transformation between two planes. By recording a calibration pattern in front of all cameras, it is possible to estimate the homography of the planes which includes the pattern. Consequently, the displacement for the peripheral and the center camera can be corrected.

Fig. 5. Recordings of the images $\{\boldsymbol{I}_5, \boldsymbol{I}_\mathrm{c}, \boldsymbol{I}_3\}$ lead to a color image. On the left-hand side, the color components are not calibrated and suffer from a depth-dependent shift. On the right-hand side, the channels have been rectified. Consequently, the color channels are aligned for the calibrated depth plane in the background. The remaining shift in color channels results due to the horizontal disparity of the cameras.

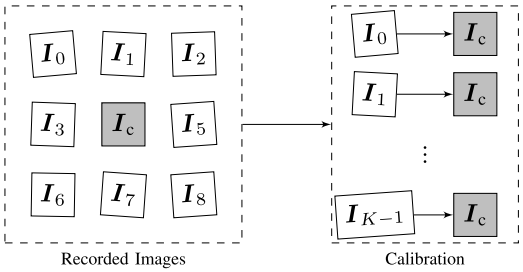

Fig. 6. A pairwise calibration approach is used in CAMSI. All peripheral camera views are calibrated with respect to the center view.

In contrast to traditional image rectification, which is typically used for stereo imaging [52], the transformation between all peripheral cameras and the center is estimated, while only the peripheral positions are warped onto the center. An overview of the proposed scheme is given in Fig. 6.

The decision regarding the calibration pattern was made in favor of using a checkerboard as this is a well-investigated approach, which is proven to work robustly and accurately. Firstly, the calibration pattern must be detected in every recording. Therefore, the features are extracted using the approach in [53]. By providing prior knowledge about the pattern, e.g., number of boxes and geometry and by utilizing a pattern that is visible in all spectral components, the estimation of the features works reliable even for the different color and bandpass filters mounted in front of the cameras. Using two feature lists, which contain the chessboard corners for one peripheral and the center position, the transformation matrix $\boldsymbol{T}_k$ is estimated for every image $\boldsymbol{I}_k$

$$\boldsymbol{T}_k = \begin{pmatrix} a_1 & a_2 & b_1 \\ a_3 & a_4 & b_2 \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

using the methods from [54], [55]. Thereby, the point correspondences are robustly estimated using an extended random sample consensus algorithm to calculate the homography. As one can see in $\boldsymbol{T}_k$, the free choice of parameters a_1 to a_4 and b_1 to b_2 allows describing translation, rotation, scaling, shearing and tilting. Thus, the projective homography

transformation is formulated as

$$\begin{pmatrix} \hat{x}_k \\ \hat{y}_k \\ 1 \end{pmatrix} = \boldsymbol{T}_k \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \tag{2}$$

which describes the corrected pixel positions $(\hat{x}_k, \hat{y}_k)$ for every peripheral view P_k. By warping the peripherals view independently, all spectral components $\boldsymbol{J}_k$ are rectified to the central camera position, but only for the depth in which the calibrated chessboard pattern was located. For the compensation of the remaining one-dimensional disparity in other depths levels, it is necessary to resample the transformed pixels at floating-point positions back to a regular grid. Therefore, the fast triangulation-based cubic interpolation [56] is used.

V. Registration

After calibration and rectification, the images still yield a one-dimensional disparity. In the following, an appropriate disparity estimation method is presented, which is suitable for multi-spectral imaging in contrast to existing state-of-the-art algorithms. Firstly, the relationship between depth and disparity will be shortly summarized. Secondly, a novel fast one-dimensional disparity estimation for multi-spectral imaging is introduced, which is calculated globally for all cameras. In the end, a pixel transform is derived, which combines both calibration and registration information to avoid multiple concatenated image resamplings that would introduce blurring.

A. Disparity and Depth

Disparity information can be transferred to a depth map using the triangulation relationship

$$d = \frac{B \cdot f}{z \cdot p} \tag{3}$$

with baseline B, focal length f, sensor pixel size p and depth z. This relation holds if the optic axes of the cameras are parallel, which is assumed for the CAMSI setup. Consequently, the depth information is directly derivable by evaluating above equation for the pixels of interest. As already shown in Tab. II, the constants B, f and p are given for the CAMSI setup. Furthermore, the relationship between disparity and depth is inversely proportional. Hence, a high depth indicates a low disparity and vice versa.

The smaller the hardware implementation can be conducted, the smaller is the disparity due to the reduced baseline distance between the cameras. Consequently, it is advisable to shrink the hardware design as much as possible to achieve a faster and more robust disparity search.

B. Fast One-Dimensional Disparity Estimation

In contrast to traditional stereo camera setups that deal with the registration of the same image content seen from different positions, the CAMSI approach has to register images at various positions, which also contain different spectral content. Consequently, the difficulty significantly increases to achieve a registered image compared to traditional stereo imaging. In an initial exploration, several state-of-the-art algorithms have been discussed (see Sec. II) to perform pairwise registrations of the center view and the various peripheral views. Altogether, neither widely-distributed template matching-approaches, e.g., [58], nor deep learning methods [59] provide a satisfactory solution due to the large deviation of the image content. For deep learning algorithms, another problem is the limited amount of multi-spectral, multi-view training data, which restricts a reasonable application. The main challenge for the registration is to allow any combination of spectral filters in the proposed camera array. Cross-spectral registration methods rely on comparing structural features, e.g., gradients, or deviations, to find the pixel-wise relationship. However, image content often significantly changes its properties across the recorded spectra. Especially, the correlation between edges can flip or structure vanishes at all, which leads to cross-spectral extinction as shown in Fig. 7. The normalized gradient [47] of the flowers image from [57] is depicted for the green component as well as two 10 nm bandpass filters with CWLs 470 nm and 710 nm. Apparently, combining costs by multiplication [47] leads to extinction and should be avoided especially when combining costs of even more spectral components. Therefore, a novel global cross-correlation registration is proposed that estimates a central disparity map by taking many pairwisely estimated cost functions into account, while overcoming the cross-spectral extinction problem. In line with this, every camera should only contribute good matches to the central disparity map, so that the setup becomes independent of the chosen filters. For CAMSI, the disparity estimation is performed for each peripheral position with respect to the center view. Hence, for the proposed array size of $K = 9$, eight estimations have to be conducted.

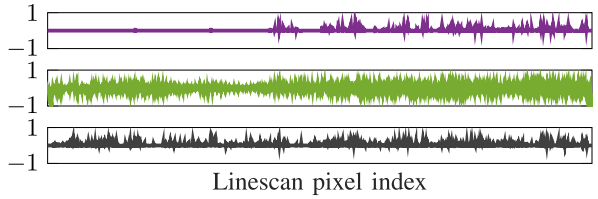

Fig. 7. The cross-spectral extinction effect demonstrated for three components of the flowers image [57]. From top to bottom, a 10 nm bandpass with CWL 470 nm, the green component, and a 10 nm bandpass at CWL 710 nm serve as example. Each row depicts the normalized gradient [47] for every pixel. Apparently, structure can be negatively correlated, or even vanished.

The disparity relationship between the different positions in the camera array is basically two-dimensional. However, due to the fixed arrangement of cameras in the array, the disparity search can be reduced to one-dimensional directions. A one-dimensional representation is chosen, such that the calibrated views have only a pure horizontal, vertical, or diagonal relationship. Therefore, it is proposed to search for candidates only horizontally, while all other cases are handled by rotating the images pairs by $+90$, $+45$ and -45 degrees, respectively. An overview of the disparity search dimensionality reduction is depicted in Fig. 8. As can be seen for the middle line of the camera array, no adjustment has to be made. In contrast, the second column must be rotated by 90 degrees, while the diagonal images are rotated by -45 and $+45$ degrees.

Then, the simplified one-dimensional registration between each center and peripheral view pair can be conducted. To give

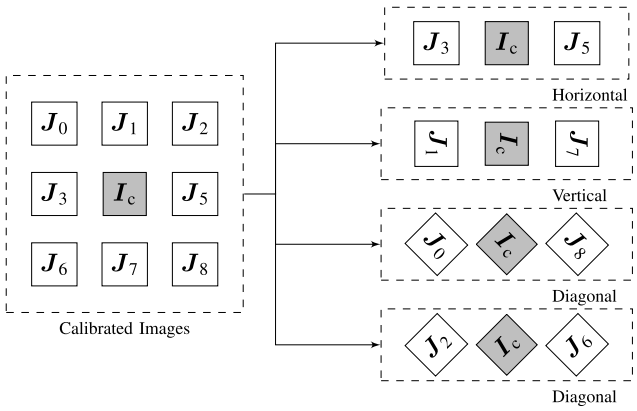

Fig. 8. The two-dimensional relationship of the different camera views is simplified by rotating the recordings in accordance with their position.

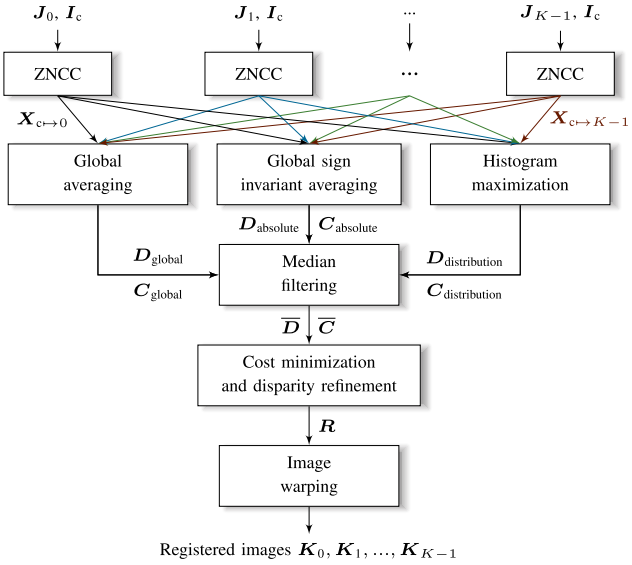

Fig. 9. Processing overview of the novel CAMSI registration approach. The calibrated recordings are used to estimate a global disparity map for the center view. Accordingly, the center disparity map is transformed to the peripheral views and used to register all recordings.

an overview of the procedure, a flowchart is shown in Fig. 9, which divides the registration problem into different subtasks. The first step is to calculate the matching costs using Zero Mean Normalized Cross-Correlation (ZNCC) [60], which compares the structure of two zero-mean signals. In contrast to other metrics, e.g., SAD, SSD, or Census [58], the ZNCC is substantially more robust for the CAMSI setup as the image content differs significantly for multi-spectral images. For two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, ZNCC is defined as

$$\mathrm{ZNCC}(\boldsymbol{a},\boldsymbol{b}) = -\frac{\langle \boldsymbol{a}-\overline{a}, \boldsymbol{b}-\overline{b}\rangle}{||\boldsymbol{a}-\overline{a}||_2 \cdot ||\boldsymbol{b}-\overline{b}||_2} \tag{4}$$

with $\overline{a}$ and $\overline{b}$ being the mean values of both input vectors. For ZNCC, costs can take values between -1 and $+1$, while -1 describes the maximum positive correlation, zero depicts no correlation and $+1$ denotes the maximum negative correlation between the signals $\boldsymbol{a}$ and $\boldsymbol{b}$. For an image of size $M \times N$ and D depth levels that shall be investigated, costs $\boldsymbol{M}_{\mathrm{c}\mapsto k} \in \mathbb{R}^{M\times N\times D}$ are obtained by evaluating the ZNCC between the center and the peripheral view k. By calculating

$$(\boldsymbol{M}_{\mathrm{c}\mapsto k})_{x,y,d} = \mathrm{ZNCC}(\boldsymbol{I}_{\mathrm{c}}(\boldsymbol{r}), \boldsymbol{J}_k(\boldsymbol{r}+d)) \tag{5}$$

for every pixel coordinate (x, y) and all investigated depth levels $d \in D$, one obtains costs for every matrix entry $(\boldsymbol{M}_{\mathrm{c}\mapsto k})_{x,y,d}$ of the cost matrix $\boldsymbol{M}_{\mathrm{c}\mapsto k}$. In the following, let $(\boldsymbol{Z})_{x,y,d}$ be the operator for accessing an element at index (x, y, d) in the exemplary matrix $\boldsymbol{Z}$, while the notation $\mathrm{c} \mapsto k$ states that the center recording is mapped to the camera position P_k. In above equation, $\boldsymbol{r}$ denotes the region of interest around the current position (x, y) using an experimentally determined support window of size $W = 7$. Varying W controls the sharpness and noise level of the disparity estimation. Thus, small window sizes lead to sharper disparity maps, but include noisy mispredictions and vice versa. For pixels at the image borders, missing entries are padded by repeating.

To assess the quality of the estimation, a search is also performed in the opposite direction from the peripheral view k to the center view, which leads to the cost matrix

$$(\boldsymbol{M}_{k\mapsto \mathrm{c}})_{x,y,d} = \mathrm{ZNCC}(\boldsymbol{J}_k(\boldsymbol{r}), \boldsymbol{I}_{\mathrm{c}}(\boldsymbol{r}+d)) \tag{6}$$

At this point, the differing baselines of the CAMSI setup must be taken into account. In fact, the disparity search range D must be increased by a factor $\sqrt{2}$ for the diagonal camera locations $k = \{0, 2, 6, 8\}$ as a wider search region has to be investigated. Thus, the resulting cost matrices $\boldsymbol{M}_{\mathrm{c}\mapsto\{0,2,6,8\}}$ and $\boldsymbol{M}_{\{0,2,6,8\}\mapsto \mathrm{c}} \in \mathbb{R}^{M\times N\times \lceil\sqrt{2}D\rceil}$ result. To obtain D disparity levels again, the cost matrices are filtered using a fast box filter and sampled according to the required dimensionality.

Given both cost estimations, the disparity maps

$$\boldsymbol{D}_{\mathrm{c}\mapsto k} = \arg\min_{(d)}\{\boldsymbol{M}_{\mathrm{c}\mapsto k}\} \tag{7}$$

and

$$\boldsymbol{D}_{k\mapsto \mathrm{c}} = \arg\min_{(d)}\{\boldsymbol{M}_{k\mapsto \mathrm{c}}\} \tag{8}$$

are obtained by total cost minimization, respectively. Afterwards, the cross-checked cost matrix

$$\boldsymbol{X}_{\mathrm{c}\mapsto k} = \frac{1}{|\boldsymbol{D}_{\mathrm{c}\mapsto k} - \boldsymbol{D}_{k\mapsto \mathrm{c}}| + 1}\boldsymbol{M}_{\mathrm{c}\mapsto k} \tag{9}$$

is formulated, which takes the discrepancy between both estimations into account. Hence, (9) gives untrustworthy estimations a lower weight, while good matches are unaffected.

After calculating the cost matrix $\boldsymbol{X}_{\mathrm{c}\mapsto k}$ for every peripheral camera view, a cost fusion is conducted to achieve a robust estimate, which makes use of all camera views. According to Fig. 9, this step consists of three different methods that are combined. Firstly, the matrices are averaged to obtain the global cost matrix

$$\boldsymbol{C}_{\mathrm{global}} = \frac{1}{K-1}\sum_{k\in\mathcal{K}} \boldsymbol{X}_{\mathrm{c}\mapsto k}. \tag{10}$$

In above equation, set $\mathcal{K} = \{0, \ldots, K-1\}\backslash\{c\}$ holds all camera view indices except the center index. Apparently, uncertainty gets averaged in the estimations due to multiple cost functions

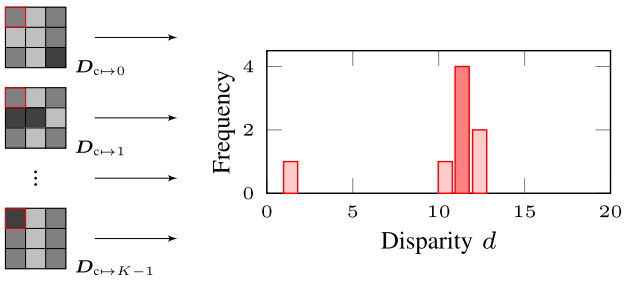

Fig. 10. Overview of the proposed histogram maximization. On the left-hand side, different sections of the various disparity maps are shown. On the right-hand side, a majority decision is conducted for every entry to find the most-likely disparity value (highlighted).

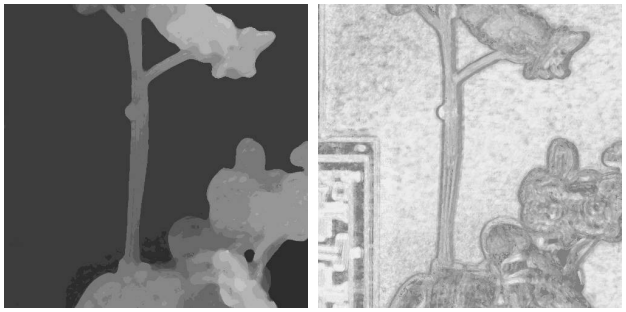

Fig. 11. A section of the disparity map and the respective ZNCC matching costs for the image set *NIR/RGB/UV Lab* (see Fig. 17). On the left, the disparity map calculated by the CAMSI method is shown. Due to the novel global registration, the map is smooth and accurate at the same time. On the right, the according ZNCC matching costs are given. Dark pixels indicate a trustworthy match, while bright regions depict more uncertain decisions.

that are combined. Typically, cross-spectral stereo matching algorithms include a cost aggregation step [44] to denoise the cost maps $\boldsymbol{X}_{c\mapsto k}$ [61]. In contrast, CAMSI skips an implicit cost aggregation step due to the high number of available cost matrices that are combined. This can be already interpreted as a high-quality aggregation.

Given a smooth cost matrix, a total cost minimization results in the disparity map

$$\boldsymbol{D}_{\text{global}} = \arg\min_{(d)}\{\boldsymbol{C}_{\text{global}}\}, \tag{11}$$

by selecting the disparity that stores the lowest costs per pixel.

Secondly, to increase the robustness of the registration even further, the correlation properties of ZNCC are taken into account. As recorded objects can flip brightness over the spectral components, it is advisable to investigate not only the positive but also the negative correlation. Hence, the global sign invariant cost matrix results in

$$\boldsymbol{C}_{\text{absolute}} = \frac{1}{K-1}\sum_{k\in\mathcal{K}} |\boldsymbol{X}_{c\mapsto k}| \tag{12}$$

similar to (10). In contrast to (11), the disparity map results in a total cost maximization

$$\boldsymbol{D}_{\text{absolute}} = \arg\max_{(d)}\{\boldsymbol{C}_{\text{absolute}}\} \tag{13}$$

as costs close to 0 indicate no correlation and costs close to $+1$ show a high positive or negative correlation.

Thirdly, the disparity distribution itself is taken into account. Therefore, the distribution matrix $\boldsymbol{C}_{\text{distribution}} \in \mathbb{R}^{M\times N\times D}$ is calculated, which stores the disparities that are estimated for the various camera views. $\boldsymbol{C}_{\text{distribution}}$ stores a histogram for every pixel position (x, y) that shows the absolute frequency of the disparity estimations per disparity level d for all maps $\boldsymbol{D}_k$. Consequently, the disparity map

$$\boldsymbol{D}_{\text{distribution}} = \arg\max_{(d)}\{\boldsymbol{C}_{\text{distribution}}\} \tag{14}$$

is calculated by choosing the most likely disparity. This third criterion can be interpreted as non-linear majority decision for every disparity entry and all camera recordings (see Fig. 10).

Given the three estimations, the combined disparity map

$$\overline{\boldsymbol{D}} = \text{median}\{\boldsymbol{D}_{\text{global}}, \boldsymbol{D}_{\text{absolute}}, \boldsymbol{D}_{\text{distribution}}\} \tag{15}$$

is calculated by applying a median filtering, so that the positive and negative correlations as well as the distribution over the cameras are taken into account. Furthermore, the combination of costs is defined similarly as

$$\overline{\boldsymbol{C}} = \text{median}\{|\boldsymbol{C}_{\text{global}}|, |\boldsymbol{C}_{\text{absolute}}|, |\boldsymbol{C}_{\text{distribution}}|\}. \tag{16}$$

Afterwards, the disparity map $\overline{\boldsymbol{D}}$ is refined by applying two filtering operations. Firstly, a peak filtering is conducted that identifies connected areas consisting of less than 1000 elements that are only allowed to differ one pixel. Instead of removing the identified pixels from the disparity map, the according costs are set to zero in the cost matrix $\overline{\boldsymbol{C}}$ to mark them for strong filtering. Then, a cost adaptive median filter is applied, which adjusts the filter size of the median filter according to the trustworthiness of the disparity estimation. The filter strength $\boldsymbol{F}$ is estimated for every position (x, y) as

$$(\boldsymbol{F})_{x,y} = \begin{cases} 0, & (\overline{\boldsymbol{C}})_{x,y} > F_{\text{th}} \\ \left\lfloor \dfrac{F_{\min}-F_{\max}}{F_{\text{th}}} \min\limits_{(d)}\{(\overline{\boldsymbol{C}})_{x,y}\} + F_{\max} \right\rfloor, & \text{else} \end{cases} \tag{17}$$

with minimal filter window size $F_{\min} = 3$, maximal window size $F_{\max} = 15$, and cost threshold $F_{\text{th}} = 0.5$. In (17), F_{th} ensures that trustworthy matches in the disparity map are not affected, while $F_{\max}$ defines the filter strength for the worst candidates and $F_{\min}$ for more convenient entries in the disparity map $\overline{\boldsymbol{D}}$. After filtering, the refined disparity map $\boldsymbol{R}$ results for the central camera position.

Finally, an exemplary disparity map section is shown in Fig. 11 together with the according ZNCC matching costs. Apparently, the proposed CAMSI registration achieves smooth and sharp disparity maps even for differing spectral images.

C. Pixel Transformation

After estimating the disparity, $\boldsymbol{R}$ is applied to warp the peripheral views to the center image. Since multiple mesh to grid resampling steps would have a negative effect on the quality, the estimated calibration matrices $\boldsymbol{T}_k$ from Sec. IV and the disparity map from Sec. V are combined instead of warping the already calibrated images for another time.

In order to achieve this, an according disparity map has to be calculated for the peripheral views. Therefore, the direction of the displacement must be taken into account (see also Fig. 8). Hence, the camera position dependent displacement sign

$$s_k = \begin{cases} -1, & \text{for } k = \{0, 1, 3, 6\} \\ +1, & \text{else} \end{cases} \tag{18}$$

is obtained. Parameter s_k is used to describe the direction where the peripheral camera is located with respect to the central camera. Due to the proposed one-dimensional disparity estimation, each rotated peripheral view is located either left (-1) or right $(+1)$ from the rotated center (see Fig. 8). Besides s_k, also the spatial position of disparity values has to be adapted for deriving the peripheral map. Thus, the peripheral disparity maps result in

$$(\boldsymbol{R}_k)_{x,y} = \begin{cases} s_k(\boldsymbol{R})_{x+s_k(\boldsymbol{R})_{x,y},\, y}, & k = \{3,5\} \\ s_k(\boldsymbol{R})_{x+s_k(\boldsymbol{R})_{x,y},\, y-s_k(\boldsymbol{R})_{x,y}}, & k = \{0,8\} \\ s_k(\boldsymbol{R})_{x+s_k(\boldsymbol{R})_{x,y},\, y+s_k(\boldsymbol{R})_{x,y}}, & k = \{2,6\} \\ s_k(\boldsymbol{R})_{x,y-s_k(\boldsymbol{R})_{x,y}}, & k = \{1,7\} \end{cases}, \tag{19}$$

by taking the epipolar geometry constraints between the peripheral camera views and the central position into account. As the displacement is known for every pixel from the center to the peripheral camera, one can also formulate the disparity map from the peripheral camera to the center, which is stated in (19). The resulting disparity maps $\boldsymbol{R}_k$ are defined for the integer grid camera positions, where the peripheral images would be expected. Unfortunately, the rectification moves the raw recordings to floating-point positions. Hence, the disparity maps $\boldsymbol{R}_k$ are resampled to the positions of the calibrated images $\boldsymbol{J}_k$ to obtain the corrected disparity maps $\boldsymbol{S}_k$ using triangulation-based cubic interpolation [56].

Afterwards, the position dependent displacement

$$\boldsymbol{v}_k = \begin{pmatrix} \tilde{\boldsymbol{x}}_k \\ \tilde{\boldsymbol{y}}_k \\ 1 \ \dots \ 1 \end{pmatrix} \in \mathbb{R}^{3 \times MN} \tag{20}$$

which contains the registration information for every pixel position (x, y), is determined and added to the mesh locations after calibration. As $\boldsymbol{v}_k$ is position dependent, the peripheral disparity maps $\boldsymbol{S}_k$ are reformulated as vector and inserted in

$$(\tilde{\boldsymbol{x}}_k)_j = \begin{cases} (\boldsymbol{S}_k)_{j/N,\, j-\lfloor j/N \rfloor N}, & \text{for } k = \{3, 5\} \\ (\boldsymbol{S}_k)_{j/N,\, j-\lfloor j/N \rfloor N}, & \text{for } k = \{0, 8\} \\ (\boldsymbol{S}_k)_{j/N,\, j-\lfloor j/N \rfloor N}, & \text{for } k = \{2, 6\} \\ 0, & \text{for } k = \{1, 7\} \end{cases} \tag{21}$$

and

$$(\tilde{\boldsymbol{y}}_k)_j = \begin{cases} 0, & \text{for } k = \{3, 5\} \\ -(\boldsymbol{S}_k)_{j/N,\, j-\lfloor j/N \rfloor N}, & \text{for } k = \{0, 8\} \\ (\boldsymbol{S}_k)_{j/N,\, j-\lfloor j/N \rfloor N}, & \text{for } k = \{2, 6\} \\ -(\boldsymbol{S}_k)_{j/N,\, j-\lfloor j/N \rfloor N}, & \text{for } k = \{1, 7\} \end{cases}, \tag{22}$$

respectively. The calculation of j/N and $j - \lfloor j/N \rfloor N$ transforms the vector index j to the row and columns indices x and y. To go more into detail, e.g., case $k = \{3, 5\}$ describes the horizontal displacement, which only requires $\tilde{\boldsymbol{x}}_k$ to be modified. To the contrary, $\tilde{\boldsymbol{y}}_k$ changes for vertical camera displacements. Moreover, the diagonal camera positions require the signs in (22) to be modified according to the rotation angle, which was presented in Fig. 8. Thus, the final mesh points

$$\begin{pmatrix} \check{\boldsymbol{x}}_k \\ \check{\boldsymbol{y}}_k \\ 1 \ \dots \ 1 \end{pmatrix} = \boldsymbol{T}_k \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ 1 \ \dots \ 1 \end{pmatrix} + \boldsymbol{v}_k \tag{23}$$

are formulated that both include calibration and disparity.

Finally, the warped image has to be resampled to integer grid positions again. Due to the combination of calibration and disparity information, the mesh points are not equally distributed anymore. Consequently, it is important to define the maximum distance that is allowed for interpolation of the grid points. In the following, α-shapes [62] with a radius of one pixel are used to define the set of image points that can be interpolated. Hence, every grid entry that has at least one adjacent pixel in radius of one pixel is interpolated. Every excluded grid point has to be reconstructed according to Sec. VI. Regarding the interpolation, the triangulation-based cubic interpolation [56] is utilized to obtain the registered images $\boldsymbol{K}_k$. Thus, a registration like Fig. 12 is obtained, which has to be reconstructed in the following.

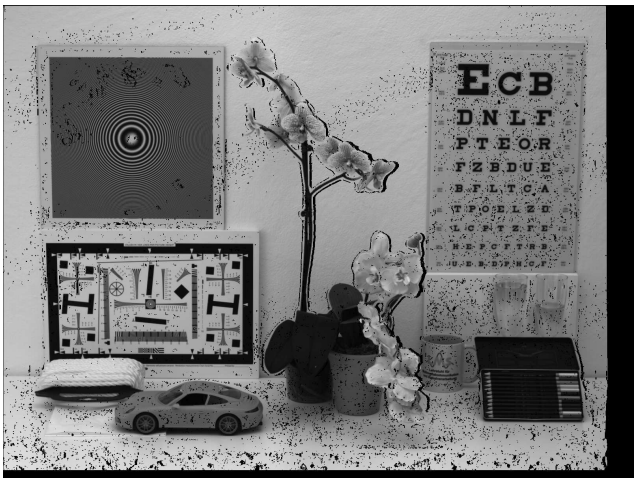

Fig. 12. Registered red channel $\tilde{\boldsymbol{I}}_5$ of image set *NIR-RGB-UV Lab* (see Fig. 17) using the proposed CAMSI registration method. The black areas indicate the missing and occluded pixels that have to be reconstructed.

VI. Reconstruction

The registered images $\boldsymbol{K}_k$ contain losses at various positions due to occlusions and mispredictions. The center image is fully preserved and can serve as reference for any distortion in every image. Moreover, it is very likely that losses are located at different positions for the various images as occlusions are dependent on the camera position. Consequently, a promising approach is to reconstruct the missing information by exploiting the spectral similarity as multiple references are available for every lost pixel. Hence, the reconstruction is interpreted as combined linear regression problem to estimate the distorted areas in the images $\boldsymbol{K}_k$.

Therefore, every image $\boldsymbol{K}_k$ is partitioned into square-shaped blocks $\mathcal{B} \in \mathbb{R}^{2 \times 2}$. For every block, the reconstruction area

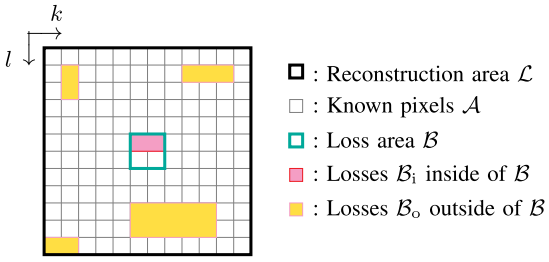

Fig. 13. Partitioning of the block-based reconstruction method for CAMSI.

$\mathcal{L} \in \mathbb{R}^{12\times 12}$ is defined, which surrounds $\mathcal{B}$ and contains the known pixels $\mathcal{A}$ and the inner and outer losses $\mathcal{B}_\mathrm{i}$ and $\mathcal{B}_\mathrm{o}$ (see. Fig. 13). Each reconstruction of a block $\mathcal{B}$ is treated as independent problem and an arbitrary processing order can be chosen. We use the optimized processing order from [63] for each image $\boldsymbol{K}_k$. The algorithm ensures that distortions are closed from the outer margin to the inside, so that the available support area is maximized. Moreover, the execution can be significantly fastened by processing unconnected losses in parallel, so that a higher computational efficiency is achieved.

After dividing the current distorted image $\boldsymbol{K}_k$ into a set of blocks $\mathcal{B}$, the block-based reconstruction is performed. Therefore, the current distorted block $\mathcal{B}$ is written as vector $\boldsymbol{s}$ and the according reference blocks from the images $\boldsymbol{K}_i$ with $i \in \{0,\ldots,K-1\} \setminus k$ are depicted as $\boldsymbol{r}_i$. The goal is to reconstruct the distorted vector by taking all undistorted references into account. For model generation, it is necessary to exclude the unknown samples from the distorted and the reference vectors. By allowing only undistorted entries of $\boldsymbol{s}$, vector $\tilde{\boldsymbol{s}}$ is obtained. The adjusted references $\tilde{\boldsymbol{r}}_i$ result after discarding samples, where $\boldsymbol{s}$ contains unknown entries. As reference views $\tilde{\boldsymbol{r}}_i$, which contain distortions themselves, are unsuitable to reconstruct $\boldsymbol{s}$, the set $\mathcal{U}$ is defined that only allows completely conserved reference views $\tilde{\boldsymbol{r}}_u$ with $u \in \mathcal{U}$.

We calculate a linear regression between every reference block and the distorted block to approximate

$$\tilde{\boldsymbol{s}} \approx a_u \cdot \tilde{\boldsymbol{r}}_u + b_u \tag{24}$$

with slope and offset scalars a_u and b_u. These are determined by minimizing the squared model error

$$\min_{a_u,b_u} ||a_u \cdot \tilde{\boldsymbol{r}}_u + b_u - \tilde{\boldsymbol{s}}||_2^2 \tag{25}$$

of the linear regression model.

The minimization problem is solved by applying a least squares fitting. Then, parameter a_u is calculated as

$$a_u = \frac{\langle \tilde{\boldsymbol{r}}_u - \overline{r}_u, \tilde{\boldsymbol{s}} - \overline{s}\rangle}{\langle \tilde{\boldsymbol{r}}_u - \overline{r}_u, \tilde{\boldsymbol{r}}_u - \overline{r}_u\rangle} \tag{26}$$

and parameter b_u results in

$$b_u = \overline{s} - a_u \cdot \overline{r}_u, \tag{27}$$

with the average values $\overline{s}$ and $\overline{r}_u$ of the distorted and the reference block. Thus, the prediction can be written as

$$\boldsymbol{p}_u = a_u \cdot \tilde{\boldsymbol{r}}_u + b_u. \tag{28}$$

By taking all suitable views into account, the combined prediction is formulated as linear combination of

$$\boldsymbol{p} = \sum_{u\in\mathcal{U}} \hat{w}_u \cdot \boldsymbol{p}_u \tag{29}$$

using weights $\hat{w}_u$. The weights are calculated with respect to the mean residual error

$$e_u = \sqrt{\frac{1}{|\mathcal{L} \setminus \{\mathcal{B}_\mathrm{u} \cup \mathcal{B}_\mathrm{o}\}|}\langle \tilde{\boldsymbol{s}} - \tilde{\boldsymbol{p}}_u, \tilde{\boldsymbol{s}} - \tilde{\boldsymbol{p}}_u\rangle}. \tag{30}$$

To penalize untrustworthy predictions further, w_u is formulated as

$$w_u = \rho^{e_u}. \tag{31}$$

An experimentally determined parameter $\rho = 0.8$ is used to control the decay of the weighting function and all weights are normalized by

$$\hat{w}_u = \frac{w_u}{\sum\limits_{u\in\mathcal{U}} w_u}. \tag{32}$$

Given the normalized weights, the prediction $\boldsymbol{p}$ of (29) can be calculated and the reconstruction

$$\hat{\boldsymbol{s}} = \begin{cases} \boldsymbol{p}, & \boldsymbol{p} \in \mathcal{B}_\mathrm{i} \\ \boldsymbol{s}, & \text{else} \end{cases} \tag{33}$$

results in estimated samples for the distortions in $\mathcal{B}_\mathrm{i}$, while original samples remain unchanged.

The block-based reconstruction is repeated until all distortions in every image are concealed. Thereby, already reconstructed pixels are used for the reconstruction of further pixels, but only in the same image. Thus, the reconstructions of different images are independent of each other and the reconstruction can be fully parallelized.

VII. Evaluation

To evaluate the performance of the proposed CAMSI system, several image sets and videos were recorded. Moreover, ground truth data was acquired for the images by sequentially mounting all utilized filters in front of the center camera. These references can be interpreted as Tunable Filter setups, e.g., filter wheels, which provide a high spatial, but a restricted temporal resolution. Consequently, ground truth data can be provided for the images and not for the video data. The first image set *NIR-RGB-UV Lab* shows different objects, e.g., plants, liquids, and test patterns, in different depths that are placed on a table. For the recordings, the filter configuration *NIR-RGB-UV* from Tab. III has been used. Furthermore, the second image set depicts different artificial plants, e.g., trees, orchids, which are captured under bad light conditions. For this, the filter set *NIR-RGB-NIR* from Tab. III was mounted. Both data sets consist of the images taken at the nine array positions as wells as ground truth images recorded from the center position. The distance between camera and objects was between 2.3 and 3.0 meters. Moreover, a third data set using the configuration *NIR-RGB-UV* shows an outdoor video with buildings, forested areas and a cloudy sky. Indeed, only CAMSI images can be provided as the scene is dynamic,

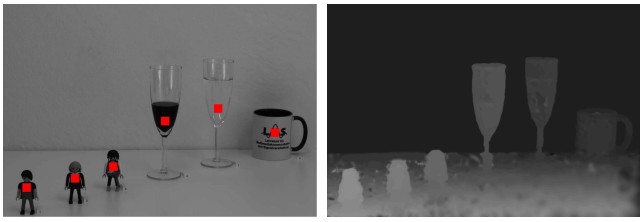

Fig. 14. Recorded test scene using the filter configuration *NIR-RGB-UV* for evaluating the registration accuracy. On the left, the green channel is depicted as an example together with the highlighted regions that are used to measure the accuracy. On the right, the according disparity map is shown, which was generated by applying CAMSI.

so that ground truth is not anymore acquirable. The distances between camera and the recorded objects range between several meters and a few kilometers. In the following, a small part of the data set is shown. All recordings together with the CAMSI reconstructions are publicly available[1] to invite researchers for participating in the proposed approach.

A. Registration Performance

At first, the registration performance of the novel CAMSI approach is compared to state-of-the-art cross-spectral disparity estimation techniques. Therefore, the depth resolution capability of two state-of-the-art methods and CAMSI is examined. As depth and disparity are inversely proportional, the depth resolution accuracy provides insight into the cross-spectral registration quality, as well. To measure the performance of disparity and depth estimation, an image set was recorded that contains six objects in varying depth levels while providing the ground truth camera distances for each object. Fig. 14 shows the recorded scene with highlighted marker regions and the estimated disparity map that is obtained after applying CAMSI. The highlighted regions are used to calculate an average depth μ_z as well as depth deviation σ_z using the relationship shown in (3). This procedure is repeated by generating further disparity maps by applying two state-of-the-art cross-spectral stereo matching algorithms, namely Census + SGM [44], [58], [64] and CCNG+SGM [47], [64]. The resulting disparity maps are averaged, so that a combined map is generated, which takes all camera view-points into account. Nevertheless, both approaches suffer from cross-spectral extinction which results in visually distorted disparity maps. Tab. IV depicts the depth resolution accuracy for the two state-of-the-art methods and the proposed CAMSI algorithm. Apparently, CAMSI achieves a significantly higher accuracy and a low deviation. Due to the global cost aggregation, a robust disparity map is estimated that does not suffer from cross-spectral extinction.

For further analysis, the selection of the utilized cost metric is evaluated. Therefore, (4) is exchanged to Census Transform [58] and CCNG [47], respectively. As all three approaches compare the structural similarity of image patches, no significant deviation regarding the overall performance of CAMSI is expected, which can be seen in Tab. V. Nevertheless, the default ZNCC cost metric performs best, which can be explained by taking the global cost aggregation of CAMSI into account. As CAMSI combines global averaging, sign invariant global averaging and a histogram analysis, it is very unlikely that structural information is extinct during combination. This advantage gets also apparent when analyzing the influence of the number of utilized cameras for generation of the global disparity map (see Fig. 15). A higher number of cameras are beneficial for the total reconstruction quality, which demonstrates the robustness of the proposed global CAMSI registration algorithm against cross-spectral extinction.

[1]CAMSI images and evaluation: https://gitlab.lms.tf.fau.de/lms/camsi

TABLE IV

ACCURACY OF DEPTH ESTIMATION FOR THE PROPOSED CAMSI APPROACH COMPARED TO STATE-OF-THE-ART CROSS-SPECTRAL STEREO MATCHING ALGORITHMS

Object		Distance	Census [58] + SGM [64]	CCNG [47] + SGM [64]	CAMSI
1	μ_z	1305 mm	1234 mm	1257 mm	**1305 mm**
	$\pm\sigma_z$		$\pm$ 53 mm	$\pm$ 10 mm	$\pm$ **0 mm**
2	μ_z	1355 mm	1279 mm	1318 mm	**1355 mm**
	$\pm\sigma_z$		$\pm$ 20 mm	$\pm$ 5 mm	$\pm$ **2 mm**
3	μ_z	1420 mm	1357 mm	1392 mm	**1422 mm**
	$\pm\sigma_z$		$\pm$ 12 mm	$\pm$ 8 mm	$\pm$ **0 mm**
4	μ_z	1495 mm	1499 mm	**1497 mm**	**1497 mm**
	$\pm\sigma_z$		$\pm$ 37 mm	$\pm$ 75 mm	$\pm$ **1 mm**
5	μ_z	1585 mm	1463 mm	1563 mm	**1587 mm**
	$\pm\sigma_z$		$\pm$ 67 mm	$\pm$ 86 mm	$\pm$ **22 mm**
6	μ_z	1615 mm	1553 mm	1641 mm	**1616 mm**
	$\pm\sigma_z$		$\pm$ 60 mm	$\pm$ 60 mm	$\pm$ **0 mm**

TABLE V

INFLUENCE OF THE APPLIED COST METRIC ON THE RECONSTRUCTION QUALITY WHEN USING THE PROPOSED GLOBAL COST AGGREGATION

	PSNR	SSIM	VMAF	BRISQUE
Image set *NIR-RGB-UV Lab*				
Census [58]	32.95 dB	0.92	96.32	19.54 [13.37]
CCNG [47]	32.20 dB	0.91	96.25	19.93 [13.37]
CAMSI ZNCC (4)	**33.58 dB**	**0.93**	**96.44**	**19.48 [13.37]**
Image set *NIR-RGB-NIR Low Light*				
Census [58]	41.86 dB	0.95	97.16	29.95 [25.37]
CCNG [47]	39.27 dB	0.95	97.13	27.96 [25.37]
CAMSI ZNCC (4)	**42.04 dB**	**0.96**	**97.26**	**27.66 [25.37]**

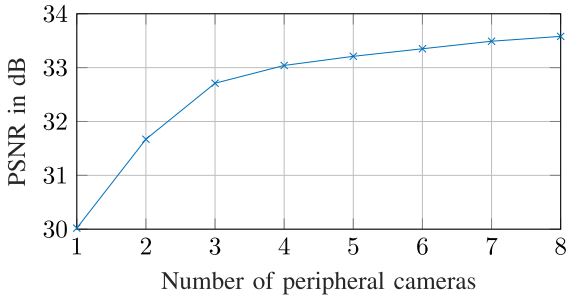

Fig. 15. Impact of the number of channels used in global cost aggregation to set up the central disparity map evaluated on the *NIR-RGB-UV Lab* images.

B. Comparison to Related Multi-Spectral Imaging Techniques

For analyzing the performance of the proposed CAMSI approach, a comparison to state-of-the-art MSFAs (see Fig. 2a)

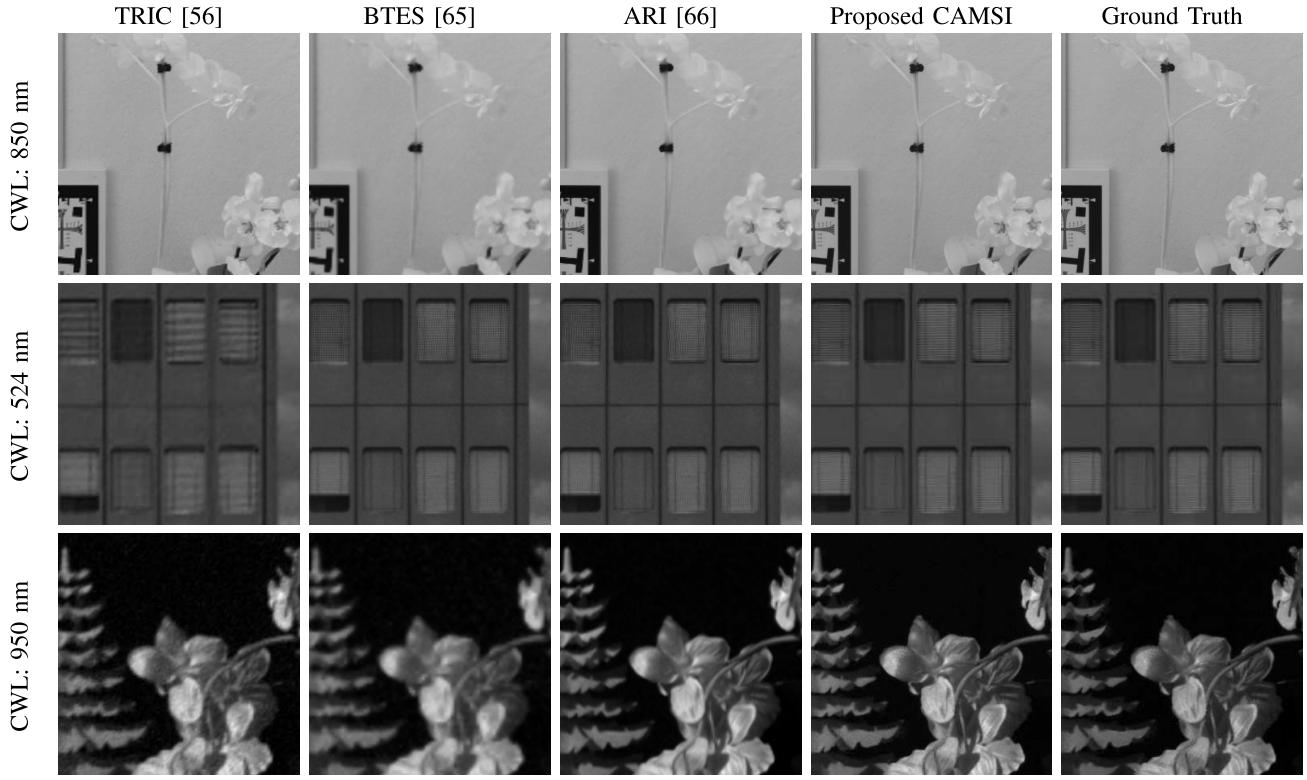

Fig. 16. Comparison between the simulated MSFAs and the proposed CAMSI method for different state-of-the-art demosaicing algorithms and varying data sets. The rows depict different image sections of the recorded data sets *NIR-RGB-UV Lab*, *NIR-RGB-UV Office View*, and *NIR-RGB-NIR Low Light*. The first three columns show the demosaicing results using TRIC [56], BTES [65], and ARI [66], respectively. Fourth column depicts the results using the proposed CAMSI approach, while last column gives the recorded ground truth data as reference.

is given. Therefore, the ground truth data of the image sets *NIR-RGB-UV Lab* and *NIR-RGB-NIR Low Light* was subsampled by a factor of three in horizontal and vertical direction. Many recent approaches, e.g., [66], [67], sample the green channel with half the sampling rate, while the remaining pixels are distributed equally to the other channels. Consequently, the green channel is reconstructed at first and further used as guidance for demosaicing all other components to achieve a high reconstruction quality. In line with this, the green channel of the ground truth data was sampled by 50 % and all other components with a rate of 6.25 %. For demosaicing, triangulation-based cubic interpolation (TRIC) [56], Binary Tree-Based Edge-Sensing (BTES) [65], and the recently published Adaptive Residual Interpolation (ARI) [66] are applied. While spatial interpolation, e.g. TRIC, is well suited for high sampling densities [68], BTES [65] and ARI [66] exploit both spatial and cross-spectral information. Afterwards, four objective metrics (PSNR, SSIM [69], VMAF [70], BRISQUE [71]) are calculated to evaluate the quality of the MSFA and CAMSI. PSNR, SSIM, and VMAF are full reference metrics and require the reconstructed as well as ground truth images. In contrast, BRISQUE is designed as no-reference metric that does not depend on ground truth data. Tab. VI, depicts the evaluation results for the simulated MSFA and the proposed CAMSI approach. The BRISQUE values in brackets depict the results for the unaltered center view. Hence, the influence of the CAMSI processing chain can be set in relation to the achievable quality of the recorded scene. Apparently, CAMSI is able to achieve a higher quality compared to the MSFA for all investigated data sets and for every applied demosaicing approach and metric. This behavior can be visually verified when looking at Fig. 16. There, a comparison between the simulated MSFA and CAMSI is given for different image sections of varying data sets and three spectral filters. Apparently, MSFAs suffers from aliasing and blurring that reduces the objective as well as the visual quality in contrast to CAMSI. However, this comes at the price that small artifacts can occur for CAMSI at depth inconsistencies, when parts of objects are occluded.

C. Visual Examples of Image and Video Data

In addition to the objective results, the visual quality of the CAMSI approach shall be analyzed. In Fig. 17, the computed CAMSI images are shown together with the ground truth data and difference images. A closer look reveals small artifacts especially at the border of the plants or at edges that are slightly misplaced. Additionally, one can see a small difference that is distributed over the complete image due to slight illumination differences of the ground truth data and the proposed CAMSI images. These differences are not introduced by the proposed CAMSI system, but are already contained in the recordings. For an enhanced visual experience, the full resolution images are included in the previously provided link[1] together with further CAMSI results, e.g., for the *NIR-RGB-NIR Low Light* filter configuration. In the additionally available data sets, it becomes apparent that CAMSI works

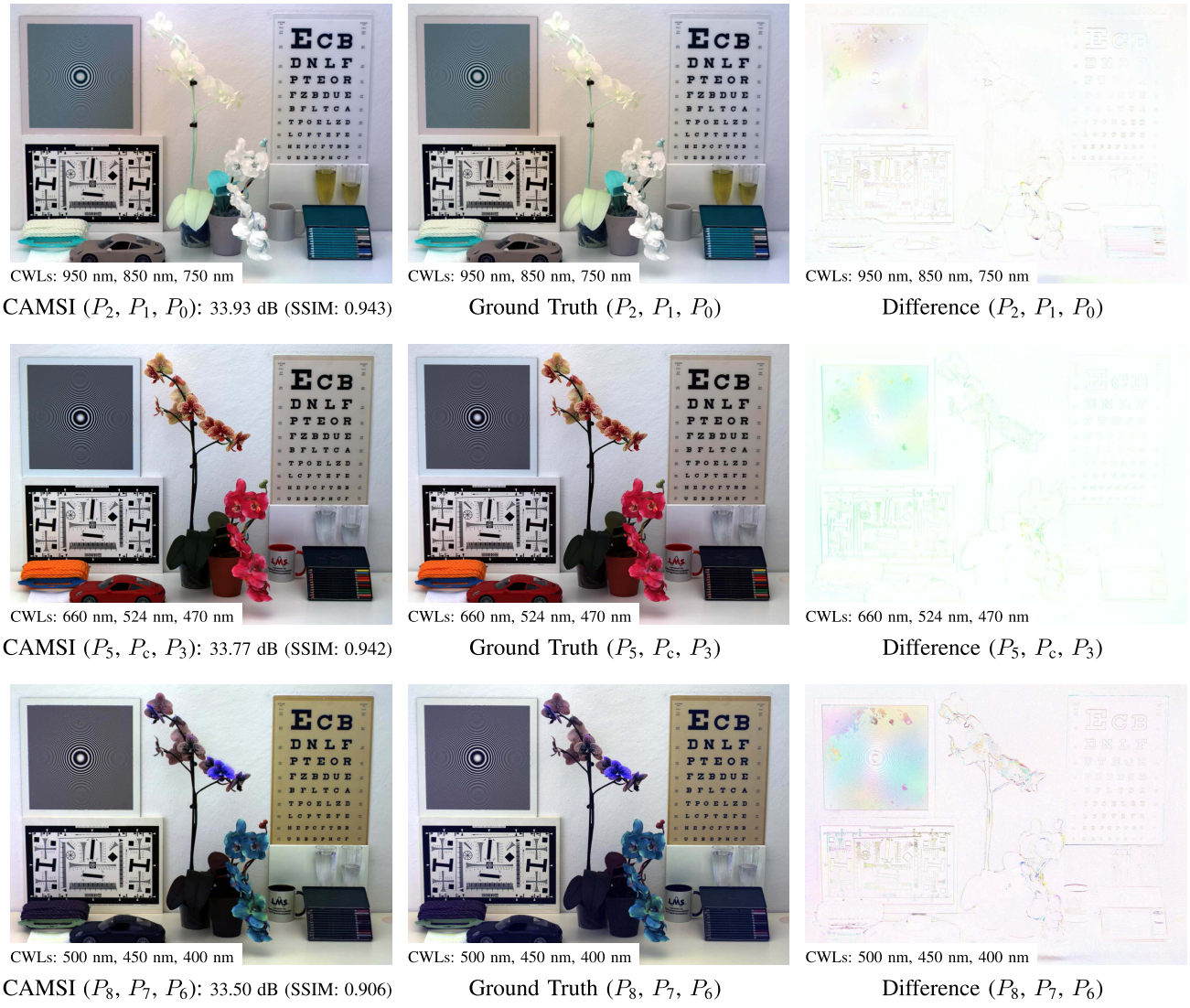

Fig. 17. Overview of the visual quality of the proposed CAMSI acquisitions. On the left, the recordings of CAMSI are shown together with the achievable PSNR and SSIM. In the middle and on the right, ground truth and difference images are given. The middle row shows the true color images, while first and third line depict false color representations. Central wavelengths of the utilized filters are added to every image. The examples are best to be viewed enlarged.

very well for arbitrary filter setups, homogeneous and heterogeneous image content, as well as different light conditions.

Moreover, the video *NIR-RGB-UV Office View* shall be examined. To evaluate the visual quality, Fig. 18 shows the composed NIR and RGB images generated by CAMSI for three different points in time. As in the previous examples, the visual quality is also very pleasing for the recorded video. Apparently, the clouds drift past during the different points of time, while the landscape is very static. Regarding the choice of filters, especially the light beams of the sun are very well distinguishable in the top NIR images compared to the bottom RGB recordings and the hydrated clouds become yellow colored. As shown in Tab. VI, the BRISQUE metric correlates very well with PSNR, SSIM, VMAF, and the visual perception. Consequently, BRISQUE is used to evaluate the video recordings as ground truth data is not applicable. For the *NIR-RGB-UV Office View* data set, an averaged BRISQUE value of 23.17 is obtained for the complete video sequence, which also corresponds to the high subjective visual quality. When comparing the BRISQUE results of the CAMSI approach to the center camera recordings (see values in brackets, Tab. VI), only a small difference is measured. This indicates that the multi-spectral images acquired by CAMSI come very close to ground truth acquisitions.

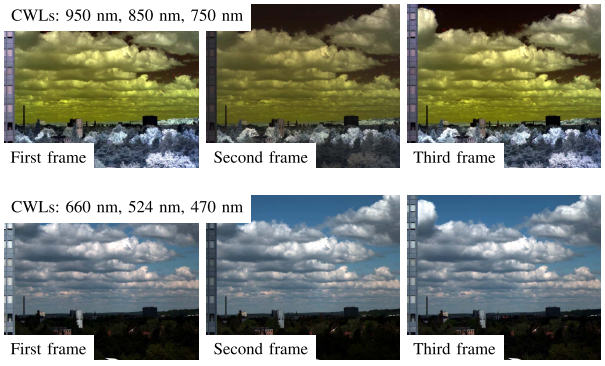

Fig. 18. Three sections from consecutive frames of the video *NIR-RGB-UV Office View* using the novel CAMSI setup and the filter set *NIR-RGB-UV* (see Tab. III). First row shows NIR false color images that were recorded at positions $\{P_2, P_1, P_0\}$. In the second row, color images were generated from the recordings at positions $\{P_5, P_4, P_3\}$.

TABLE VI

EVALUATION RESULTS FOR A DEMOSAICED, SIMULATED MSFA AND THE NOVEL CAMSI SETUP (PSNR, SSIM, VMAF: THE HIGHER THE BETTER; BRISQUE: THE LOWER THE BETTER). BRISQUE VALUES IN BRACKETS SHOW THE QUALITY OF THE RECORDINGS TAKEN FROM THE UNPROCESSED CENTER POSITION P_C AS REFERENCE

	PSNR	**SSIM**	**VMAF**	**BRISQUE**
Image set *NIR-RGB-UV Lab*				
MSFA (TRIC [56])	27.84 dB	0.84	96.04	33.72 [13.37]
MSFA (BTES [65])	28.67 dB	0.84	95.70	52.99 [13.37]
MSFA (ARI [66])	31.25 dB	0.90	96.37	31.48 [13.37]
Proposed CAMSI	**33.58 dB**	**0.93**	**96.44**	**19.48 [13.37]**
Image set *NIR-RGB-NIR Low Light*				
MSFA (TRIC [56])	37.37 dB	0.89	96.83	47.56 [25.37]
MSFA (BTES [65])	38.27 dB	0.89	96.96	44.06 [25.37]
MSFA (ARI [66])	39.44 dB	0.90	97.01	42.86 [25.37]
Proposed CAMSI	**42.04 dB**	**0.96**	**97.26**	**27.66 [25.37]**
Video *NIR-RGB-UV Office View*				
MSFA (TRIC [56])	-	-	-	43.85 [23.12]
MSFA (BTES [65])	-	-	-	38.22 [23.12]
MSFA (ARI [66])	-	-	-	36.75 [23.12]
Proposed CAMSI	-	-	-	**23.29 [23.12]**

TABLE VII

PROCESSING TIME OF THE CAMSI IMPLEMENTATION WRITTEN IN MATLAB ON A DESKTOP COMPUTER AND A MOBILE NOTEBOOK

Algorithm	***Desktop***	***Notebook***
Registration	67.4 s	211.6 s
Reconstruction	4.8 s	18.3 s
Total	72.2 s	229.9 s

D. Computational Complexity

Finally, the computational complexity of the proposed CAMSI framework shall be discussed. As camera calibration has to be conducted only once, the processing time of CAMSI is dominated by registration and reconstruction. The cost metric (4) is evaluated for every camera position P_k and all disparity candidates D, hence registration is more complex than solving regression tasks for the occluded and mispredicted pixels during reconstruction. Both, registration and reconstruction mainly calculate mean, variance, and covariance values of image patches, so that an efficient implementation could be derived by using integral images [72], [73]. Additionally, more advanced techniques can be applied to lower the computation time even further [74], or the algorithms can be implemented on GPU [75]. For the sake of an intuitive implementation, CAMSI was written in MATLAB without the usage of integral images, or any further optimization. To demonstrate the computational complexity, the CAMSI framework was evaluated on the test system *Desktop*, which encloses an Intel i9-7940X CPU and 64 GB RAM. Additionally, the test system *Notebook* was used that is equipped with a i7-6700HQ CPU and 16 GB RAM. The mean computation time averaged over 100 executions and the specified camera resolution in Tab. II is depicted in Tab. VII. The algorithms have been parallelized to exploit the multi-core computer architecture.

VIII. CONCLUSION

In this contribution, a novel multi-spectral imaging system is presented. This becomes necessary as a review on state-of-the-art multi-spectral imaging techniques has revealed that none of the existing approaches is capable to capture videos with a high spatial, temporal, and spectral resolution at the same time. Often, the dynamic application of filters is equally difficult as designing a handy setup, which is of compact size. A novel approach that can remedy the existing challenges is the proposed camera array for multi-spectral imaging, which provides additional depth information. However, this comes at the price of recording differing spectral images at various viewpoints, such that algorithmic post-processing is required after acquisition. By introducing a geometrically constrained multi-camera sensor layout, the formulation of novel registration and reconstruction algorithms becomes possible to globally set up robust models for generating consistent multi-spectral videos. The performance of the novel camera array for multi-spectral imaging is analyzed by giving an extensive visual and objective evaluation. In comparison to recently published multi-spectral filter array imaging systems, the novel acquisition approach achieves an average gain of 2.5 dB PSNR. Moreover, the recorded data sets are provided online together with the reconstructed images. In the future, it will be evaluated whether the novel concept is expandable to hyper-spectral imaging by extending the geometrically constrained multi-camera sensor layout.

REFERENCES

[1] A. Kumar and Y. Zhou, "Human ident. using finger images," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2228–2244, Apr. 2012.

[2] P. Gupta and P. Gupta, "Multibiometric authentication system using slap fingerprints, palm dorsal vein, and hand geometry," *IEEE Trans. Ind. Electron.*, vol. 65, no. 12, pp. 9777–9784, Dec. 2018.

[3] B. Garavelli, A. Mencarelli, and L. Zanotti, "XSpectra: The most advanced real time food contaminants detector," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2017, pp. 1–4.

[4] P. Colantoni, R. Pillay, C. Lahanier, and D. Pitzalis, "Analysis of multispectral images of paintings," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2006, pp. 1–5.

[5] Z. Zhou, S. Li, and Y. Shao, "Crops classification from sentinel-2A multi-spectral remote sensing images based on convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 5300–5303.

[6] A. Torres-Rua, M. A. Arab, L. Hassan-Esfahani, A. Jensen, and M. McKee, "Development of unmanned aerial systems for use in precision agriculture: The AggieAir experience," in *Proc. IEEE Conf. Technol. Sustainability (SusTech)*, Jul. 2015, pp. 77–82.

[7] S. Liu, J. Zhang, and Y. Ma, "Bathymetric ability of SPOT-5 multi-spectral image in shallow coastal water," in *Proc. 18th Int. Conf. Geoinformatics*, Jun. 2010, pp. 1–5.

[8] R. Duca, F. Del Frate, and F. G. Roca, "From multi-spectral to hyper-spectral imagery: A quantitative analysis of the improvements in terms of land cover classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 4, Jul. 2008, pp. 770–773.

[9] J. Zhang, E. Qu, J. Cao, Z. Fan, and H. Yang, "An automatic multi-spectral infrared sea surface temperature radiometer," in *Proc. Int. Conf. Electron. Optoelectronics*, vol. 2, Jul. 2011, pp. 339–343.

[10] X. Ran and B. Xiao, "Multi-spectral image analysis for rescue target detection," in *Proc. Int. Assoc. Inst. Navigat. World Congr. (IAIN)*, Oct. 2015, pp. 1–4.

[11] Q. Xu, J. Lei, and L. Zeng, "Evaluation of *in vivo* microcirculation via orthogonal polarization multi-spectral imaging and multi-spectral fusion," in *Proc. Int. Conf. Inf. Technol., Comput. Eng. Manage. Sci.*, vol. 3, Sep. 2011, pp. 164–168.

[12] S. Prigent, X. Descombes, D. Zugaj, P. Martel, and J. Zerubia, "Multi-spectral image analysis for skin pigmentation classification," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 3641–3644.

[13] M. Rapczynski, C. Zhang, A. Al-Hamadi, and G. Notni, "A multi-spectral database for NIR heart rate estimation," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2022–2026.

[14] N. M. Mohan and V. J. Kumar, "Contact-less, multi-spectral imaging of dermal perfusion," in *Proc. IEEE Instrum. Meas. Technol. Conf.*, May 2008, pp. 793–796.

[15] J. Brauers, N. Schulte, and T. Aach, "Modeling and compensation of geometric distortions of multispectral cameras with optical bandpass filter wheels," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2007, pp. 1902–1906.

[16] Y. Monno, S. Kikuchi, M. Tanaka, and M. Okutomi, "A practical one-shot multispectral imaging system using a single image sensor," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3048–3059, Oct. 2015.

[17] B. Bayer, "Color imaging array," U.S. America Patent US 3 971 065A, Jul. 20, 1975.

[18] W. Boyle and G. Smith, "Charge coupled semiconductor deadapted devices," *Bell Syst. Tech. J.*, vol. 49, pp. 587–593, Jul. 1970.

[19] H. Kumar Aggarwal and A. Majumdar, "Multi-spectral demosaicing technique for single-sensor imaging," in *Proc. 4th Nat. Conf. Comput. Vis., Pattern Recognit., Image Process. Graph. (NCVPRIPG)*, Dec. 2013, pp. 1–4.

[20] S. Mendis, S. E. Kemeny, and E. R. Fossum, "CMOS active pixel image sensor," *IEEE Trans. Electron Devices*, vol. 41, no. 3, pp. 452–453, Mar. 1994.

[21] M. M. El-Desouki, O. Marinov, M. J. Deen, and Q. Fang, "CMOS active-pixel sensor with *in-situ* memory for ultrahigh-speed imaging," *IEEE Sensors J.*, vol. 11, no. 6, pp. 1375–1379, Jun. 2011.

[22] D.-C. Sung and H.-W. Tsao, "Color filter array demosaicking by using subband synthesis scheme," *IEEE Sensors J.*, vol. 15, no. 11, pp. 6164–6172, Nov. 2015.

[23] Y. Yamashita and S. Sugawa, "Intercolor-filter crosstalk model for image sensors with color filter array," *IEEE Trans. Electron Devices*, vol. 65, no. 6, pp. 2531–2536, Jun. 2018.

[24] P. Lapray, X. Wang, J. Thomas, and P. Gouton, "Multispectral filter arrays: Recent advances and practical implementation," *Sensors*, vol. 14, pp. 21626–21659, Nov. 2014.

[25] J. Thomas, P. Lapray, P. Gouton, and C. Clerc, "Spectral characterization of a prototype SFA camera for joint visible and NIR acquisition," *Sensors*, vol. 16, no. 7, pp. 993–1012, Jun. 2016.

[26] H. Park and K. B. Crozier, "Multispectral imaging with vertical silicon nanowires," *Sci. Rep.*, vol. 3, no. 1, pp. 1–2, Aug. 2013.

[27] B. Geelen, N. Tack, and A. Lambrechts, "A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic," in *Proc. SPIE, 7th Int. Soc. Opt. Photon., Adv. Fabr. Technol. Micro/Nano Opt. Photon.*, vol. 8974, Mar. 2014, pp. 80–87.

[28] Z. Liu, S. Ma, Y. Ji, L. Liu, J. Guo, and Y. He, "Parallel scan hyperspectral fluorescence imaging system and biomedical application for microarrays," *J. Phys.*, vol. 277, pp. 1–10, Jan. 2011.

[29] N. Tack, A. Lambrechts, P. Soussan, and L. Haspeslagh, "A compact, high-speed, and low-cost hyperspectral imager," in *Proc. SPIE, 7th Int. Soc. Opt. Photon., Silicon Photon.*, vol. 8266, Feb. 2012, pp. 126–138. [Online]. Available: https://doi.org/10.1117/12.908172

[30] F. Li, C. Li, L. Tang, and Y. Guo, "Elastic registration for airborne multispectral line scanners," *J. Appl. Remote Sens.*, vol. 8, no. 1, pp. 1–13, Jun. 2014.

[31] J. Brauers, N. Schulte, and T. Aach, "Multispectral filter-wheel cameras: Geometric distortion model and compensation algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 12, pp. 2368–2380, Dec. 2008.

[32] J. Brauers and T. Aach, "Geometric calibration of lens and filter distortions for multispectral filter-wheel cameras," *IEEE Trans. Image Process.*, vol. 20, no. 2, pp. 496–505, Feb. 2011.

[33] D. N. Stratis, K. L. Eland, J. C. Carter, S. J. Tomlinson, and S. M. Angel, "Comparison of acousto-optic and liquid crystal tunable filters for laser-induced breakdown spectroscopy," *Appl. Spectrosc.*, vol. 55, no. 8, pp. 999–1004, Aug. 2001.

[34] O. Khait, S. Smirnov, and C. D. Tran, "Time-resolved multispectral imaging spectrometer," *Appl. Spectrosc.*, vol. 54, no. 12, pp. 1734–1742, Dec. 2000.

[35] K. Martinez, J. Cupitt, and D. Saunders, "High-resolution colorimetric imaging of paintings," *Proc. SPIE, Cameras, Scanners, Image Acquisition Syst.*, vol. 1901, pp. 1–12, May 1993.

[36] M. Kudenov, M. Jungwirth, E. Dereniak, and G. Gerhart, "White-light sagnac interferometer for snapshot polarimetric and multispectral imaging," *Proc. SPIE, Polarization, Meas., Anal., Remote Sens.*, vol. 7672, pp. 1–10, Apr. 2010.

[37] K. Kagawa and S. Kawahito, "Endoscopic application of a compact compound-eye camera," *Makara J. Technol.*, vol. 18, no. 3, pp. 128–132, Jan. 2015.

[38] R. Marinov, N. Cui, M. Garcia, S. B. Powell, and V. Gruev, "A 4-Megapixel cooled CCD division of focal plane polarimeter for celestial imaging," *IEEE Sensors J.*, vol. 17, no. 9, pp. 2725–2733, May 2017.

[39] D. Knipp, H. Stiebig, S. R. Bhalotra, E. Bunte, H. L. Kung, and D. A. B. Miller, "Silicon-based micro-Fourier spectrometer," *IEEE Trans. Electron Devices*, vol. 52, no. 3, pp. 419–426, Mar. 2005.

[40] F. Kazemzadeh, S. A. Haider, C. Scharfenberger, A. Wong, and D. A. Clausi, "Multispectral stereoscopic imaging device: Simultaneous multiview imaging from the visible to the near-infrared," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 7, pp. 1871–1873, Jul. 2014.

[41] A. Longoni, F. Zaraga, G. Langfelder, and L. Bombelli, "The transverse field detector (TFD): A novel color-sensitive CMOS device," *IEEE Electron Device Lett.*, vol. 29, no. 12, pp. 1306–1308, Dec. 2008.

[42] T. Zhi, B. R. Pires, M. Hebert, and S. G. Narasimhan, "Deep material-aware cross-spectral stereo matching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1916–1925.

[43] R. Shrestha, A. Mansouri, and J. Y. Hardeberg, "Multispectral imaging using a stereo camera: Concept, design and assessment," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 1, pp. 1–15, Sep. 2011.

[44] P. Pinggera, T. Breckon, and H. Bischof, "On cross-spectral stereo matching using dense gradient features," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Sep. 2012, pp. 1–12.

[45] X. Shen, L. Xu, Q. Zhang, and J. Jia, "Multi-modal and multi-spectral registration for natural images," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2014, pp. 309–324.

[46] S. Kim, D. Min, B. Ham, M. N. Do, and K. Sohn, "DASC: Robust dense descriptor for multi-modal and multi-spectral correspondence estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 9, pp. 1712–1729, Sep. 2017.

[47] J. Holloway, K. Mitra, S. J. Koppal, and A. N. Veeraraghavan, "Generalized assorted camera arrays: Robust cross-channel registration and applications," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 823–835, Mar. 2015.

[48] J. Navarro and A. Buades, "Robust and dense depth estimation for light field images," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1873–1886, Apr. 2017.

[49] G. S. Muralidhar, A. C. Bovik, and M. K. Markey, "Disparity estimation on stereo mammograms," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2851–2863, Sep. 2015.

[50] T. Dang, C. Hoffmann, and C. Stiller, "Continuous stereo self-calibration by camera parameter tracking," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1536–1550, Jul. 2009.

[51] T. Rahman and N. Krouglicof, "An efficient camera calibration technique offering robustness and accuracy over a wide range of lens distortion," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 626–637, Feb. 2012.

[52] D. V. Papadimitriou and T. J. Dennis, "Epipolar line estimation and rectification for stereo image pairs," *IEEE Trans. Image Process.*, vol. 5, no. 4, pp. 672–676, Apr. 1996.

[53] A. Geiger, F. Moosmann, O. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2012, pp. 3936–3943.

[54] P. H. S. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understand.*, vol. 78, no. 1, pp. 138–156, Apr. 2000.

[55] R. Hartley and A. Zisserman, *Multiple View Geometry CV*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2003.

[56] I. Amidror, "Scattered data interpolation methods for electronic imaging systems: A survey," *J. Electron. Imag.*, vol. 11, no. 2, pp. 157–176, 2002.

[57] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.

[58] V. Q. Dinh, V. D. Nguyen, and J. W. Jeon, "Robust matching cost function for stereo correspondence using matching by tone mapping and adaptive orthogonal integral image," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5416–5431, Dec. 2015.

[59] E. Ilg, T. Saikia, M. Keuper, and T. Brox, "Occlusions, motion and depth boundaries with a generic network for disparity, optical flow or scene flow estimation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 1–26.

[60] S. Chambon and A. Crouzil, "Towards correlation-based matching algorithms that are robust near occlusions," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, vol. 3, 2004, pp. 20–23.

[61] O. Choi and H. S. Chang, "Yet another cost aggregation over models," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5397–5410, Nov. 2016.

[62] S. K. Lodha and R. Franke, "Scattered data techniques for surfaces," in *Proc. Sci. Vis. Conf. (DAGSTUHL)*, Jun. 1997, pp. 181–222.
[63] J. Seiler and A. Kaup, "Optimized and parallelized processing order for improved frequency selective signal extrapolation," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2011, pp. 269–273.
[64] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2005, pp. 807–814.
[65] L. Miao, H. Qi, R. Ramanath, and W. E. Snyder, "Binary tree-based generic demosaicking algorithm for multispectral filter arrays," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3550–3558, Nov. 2006.
[66] Y. Monno, D. Kiku, M. Tanaka, and M. Okutomi, "Adaptive residual interpolation for color and multispectral image demosaicking," *Sensors*, vol. 17, no. 12, pp. 2787–2811, Dec. 2017.
[67] Y. Monno, D. Kiku, S. Kikuchi, M. Tanaka, and M. Okutomi, "Multispectral demosaicking with novel guide image generation and residual interpolation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 645–649.
[68] S. Mihoubi, O. Losson, B. Mathon, and L. Macaire, "Multispectral demosaicing using pseudo-panchromatic image," *IEEE Trans. Comput. Imag.*, vol. 3, no. 4, pp. 982–995, Dec. 2017.
[69] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[70] A. Aaron, Z. Li, M. Manohara, J. Y. Lin, E. C.-H. Wu, and C.-C.-J. Kuo, "Challenges in cloud based ingest and encoding for high quality streaming media," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 1732–1736.
[71] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[72] F. C. Crow, "Summed-area tables for texture mapping," *ACM SIGGRAPH Comput. Graph.*, vol. 18, no. 3, pp. 207–212, Jul. 1984.
[73] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Dec. 2001, p. 1.
[74] J. Luo and E. E. Konofagou, "A fast normalized cross-correlation calculation method for motion estimation," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 57, no. 6, pp. 1347–1357, Jun. 2010.
[75] R. Fan and N. Dahnoun, "Real-time implementation of stereo vision based on optimised normalised cross-correlation and propagated search range on a GPU," in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, Oct. 2017, pp. 1–6.

Nils Genser (Graduate Student Member, IEEE) received the master's degree in information and communication technology from Friedrich-Alexander University Erlangen-Nürnberg (FAU), Germany, in 2016.

During his master's degree, he worked on image reconstruction and conducted his thesis at the Chair of Multimedia Communications and Signal Processing, FAU, where he continued as a Researcher. His research interests include image and video signal processing, reconstruction, and coding. Moreover, he conducts research on spectral imaging, especially on regression-based and deep-learning-based registration and reconstruction algorithms. Among other things, he received a Best Master Thesis Award at FAU and a Best Student Paper Award at IWSSIP, in 2017.

Jürgen Seiler (Senior Member, IEEE) received the diploma degree in electrical engineering, electronics and information technology and the Ph.D. and Habilitation degrees from the Chair of Multimedia Communications and Signal Processing, Friedrich-Alexander Universität Erlangen-Nürnberg, Germany, in 2006, 2011, and 2018, respectively.

He is currently a Senior Scientist and a Lecturer with the Chair of Multimedia Communications and Signal Processing, Friedrich-Alexander Universität Erlangen-Nürnberg. He has authored or coauthored more than 90 technical publications. His research interests include image and video signal processing, signal reconstruction and coding, signal transforms, and linear systems theory. He received the Dissertation Award of the Information Technology Society of the German Electrical Engineering Association as well as the Dissertation Award of the Staedtler-Foundation, both in 2012. In 2007, he received diploma awards from the Institute of Electrical Engineering, Electronics and Information Technology, Erlangen, as well as from the German Electrical Engineering Association. He also received scholarships from the German National Academic Foundation and the Lucent Technologies Foundation. He was a co-recipient of four best paper awards.

André Kaup (Fellow, IEEE) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from RWTH Aachen University, Aachen, Germany, in 1989 and 1995, respectively.

He joined Siemens Corporate Technology, Munich, Germany, in 1995, and became the Head of the Mobile Applications and Services Group in 1999. Since 2001, he has been a Full Professor and the Head of the Chair of Multimedia Communications and Signal Processing, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Germany. From 1997 to 2001, he was the Head of the German MPEG delegation. From 2005 to 2007, he was a Vice Speaker of the DFG Collaborative Research Center 603. From 2015 to 2017, he has served as the Head of the Department of Electrical Engineering and the Vice Dean of the Faculty of Engineering, FAU. He has authored around 350 journal and conference papers and has over 120 patents granted or pending. His research interests include image and video signal processing and coding, and multimedia communication. He is a member of the IEEE Multimedia Signal Processing Technical Committee and the scientific advisory board of the German VDE/ITG. In 2018, he was elected a Full Member of the Bavarian Academy of Sciences. He was a Siemens Inventor of the Year 1998 and obtained the 1999 ITG Award. He received several best paper awards including the Paul Dan Cristea Special Award in 2013, and his group won the Grand Video Compression Challenge at the Picture Coding Symposium 2013. The Faculty of Engineering at FAU honored him with the Teaching Award in 2015. He has also served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He was a Guest Editor for IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING.