Fast Motion Estimation for Field Sequential Imaging: Survey and Benchmark

Holger Steiner[a], Hendrik Sommerhoff[b], David Bulczak[b], Norbert Jung[a], Martin Lambers[b], Andreas Kolb[b]

[a] *Hochschule Bonn-Rhein-Sieg, St. Augustin, Germany*
[b] *University of Siegen, Germany*

Abstract

Field sequential (FS) imaging comprises image acquisition systems that capture image channels in temporal sequence in order to provide the final image. A classical application is multispectral imaging. In case of dynamic scenes, the sequential nature of the acquisition imposes motion artifacts, i.e., spatially misaligned images channels. Compensating motion artifacts for this kind of imagery is non-trivial, as common methods for motion estimation rely on the intensity consistency constraint that is violated in FS imaging.

This paper surveys approaches to motion compensation in the context of FS imaging. We focus on accuracy in handling intensity inconsistent data and, secondarily, speed, as FS imaging is commonly done in real-time. We introduce a conceptual classification for algorithmic approaches for motion estimation for FS imagery and discuss known and modified approaches to tackle the intensity inconsistencies between adjacent image channels using image transformation and intensity correction methods. As result, we get a set of 379 variants of motion estimation methods applicable to FS data streams. We evaluate these methods using our benchmark database, which comprises data sets from the Middlebury and the MPI Sintel databases, modified to emulate FS imagery, as well as additionally captured multispectral short wave infrared (SWIR) and sRGB image sequences, as well as simulated Time-of-Flight (ToF) image sequences that consist of four channels (called phase images). In order to quantify the motion estimation techniques, we use a ranking scheme similar to Middlebury and combine it with a run-time evaluation.

Keywords: Field sequential imaging, motion estimation, optical flow

1. Introduction

Field sequential (FS)[1] imaging systems acquire several *channel* images sequentially at full spatial resolution of the final image. These kind of image acquisition systems mainly appear in

Email address: andreas.kolb@uni-siegen.de (Andreas Kolb)
URL: www.cg.informatik.uni-siegen.de (Andreas Kolb)
[1] derived from Field Sequential Color Capturing for color imaging (Daly and Feng, 2004)

multispectral imaging, but also in range imaging, e.g. in *Time-of-Flight (ToF)* range imaging. In the case of multispectral imaging, the final image is composed of this set of spectral channels, while in ToF range imaging, the final depth image is computed from the channels (called *phase images* in the ToF context).

Multispectral FS imaging system are capable of capturing high-density spectral information of object surfaces and thus offer several advantages over grayscale or RGB cameras in applications such as remote sensing, astronomy, agriculture, medicine or food quality control (Gowen et al., 2007), as well as high quality color image reproduction and conservation of art (Brauers et al., 2009). While *simultaneous* multispectral image acquisition use, e.g., static filters such as the Bayer pattern or beam splitters, multispectral FS imaging systems are realized using, e.g., broad band imagers combined with interchangeable band pass filters mounted on a filter wheel (Helling et al., 2004; Brauers et al., 2009; Bourlai et al., 2012), electronically tunable filters (Gat, 2000), or active (narrow band) illumination setups (Steiner et al., 2016). The multispectral FS imaging approaches are more flexible in selecting the spectral bands and allow for the acquisition of a much larger number of spectral channels than simultaneous approaches.

ToF cameras calculate the camera-object distance by estimating the time delay that actively emitted light takes to travel from the light source to the object surface and back to the sensor's pixel. Therefore, the amplitude of the emitted light signal is modulated and the backscattered light signal is correlated at pixel-level in the sensor. It takes at least three different *phase images* (channels in our notation) in order to reconstruct a distance image (Lange and Seitz, 2001; Kolb et al., 2010; Lambers et al., 2015).

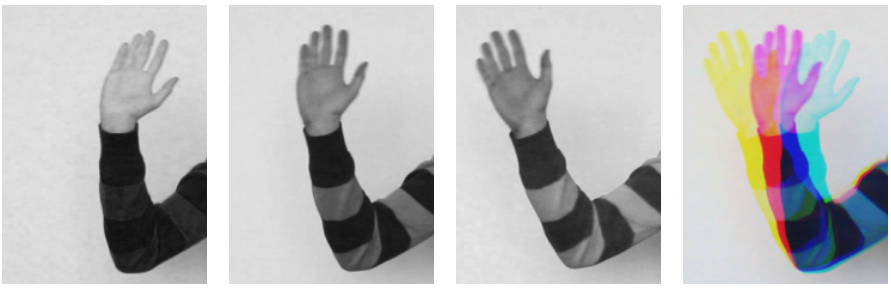


Figure 1: A waving hand recorded using field-sequential color capturing, with channels captured at subsequent times t. When the channels are combined in a multispectral image, the color breakup effect occurs.

In case of dynamic scenes, all FS image capturing systems suffer from motion artifacts, as moving objects will not match between the different channels; see Fig. 1. Depending on the amount of motion and the application requirements, raw FS imagery cannot be used without compensating the motion artifacts. Although motion estimation has a long and successful history

in computer vision, existing motion estimation techniques cannot handle FS imagery properly, as it *strongly violates the intensity consistency assumption* between adjacent channels, which most state of the art motion estimation techniques rely upon (Baker et al., 2011).

In this paper we describe simple and generic approaches in order to apply existing motion estimation approaches to FS imagery that, in most cases, incorporate up to three components:

Motion Estimation Scheme: There are various basic concepts on how to estimate motion within an FS image stream (see Sec. 3). *Corresponding channel matching (CCM)* methods estimate motion fields between corresponding channels of adjacent images, thus preventing the intensity inconsistency problem at the cost of larger temporal gaps that need to be bridged. In contrast to this, *neighboring channel matching (NCM)* approaches estimate temporally dense motion fields between neighboring channels within an image or across neighboring images, which requires the handling of intensity inconsistency; see also Fig. 2.

Image Transformation & Intensity Correction: Any NCM method needs to handle the intensity inconsistency. This can be either done by transforming the image in another domain (e.g. gradients) or by correcting the intensity by some preprocessing procedure.

Intensity Consistent Motion Estimation: Finally, the motion between several channels within or across the FS images are estimated using a state-of-the-art method (see Sec. 2).

We present a first thorough analysis and discussion of motion estimation approaches that are applicable to FS imaging systems. As motion compensation for FS imaging primarily makes sense for real-time image capturing, we mainly focus on online estimation methods. Thus, the performance of any FS motion estimation method is defined by both, *high motion estimation accuracy* and *low processing time*.

This paper provides the following methodological and technical contributions:

- A set of *general concepts for motion estimation schemes* that are applicable to FS imagery, refining the basic principles of *corresponding channel matching (CCM)* and *neighboring channel matching (NCM)* (see Sec. 3).
- A *benchmark dataset* including different test scenarios from both domains, FS multispectral imagery as well as phase image sequences from ToF cameras. These data sets include translational and rotational movements and partially comprise ground truth data. Regarding multispectral imagery, we also include existing data sets such as Middlebury or MPI Sintel (Sec. 6).
- An *in-depth evaluation* with respect to compensation accuracy and processing time of a large set of FS motion estimation methods comprising the components listed above (Sec. 7).

As an overall contribution, this paper provides simplified and quantified means to select the most promising methods for motion estimation on FS data depending on sample-based application scenarios.

The remainder of the paper is structured as follows. Sec. 2 gives an overview on existing motion estimation algorithms based on optical flow and block matching. Sec. 3 discusses the general approaches applicable to field sequential motion estimation. In sections 4 and 5 we describe the image transformation and intensity correction schemes that we use in order to compensate for the intensity inconsistency of FS imagery. Sec. 6 presents the evaluation of all 379 algorithm combinations.

2. State of the Art in Motion Estimation

Optical Flow (OF), *Block Matching* (BM), and *Deep Neural Network* based methods are the most prominent classes of approaches to estimate dense motion fields between consecutive images. As a fully comprehensive survey of motion estimation techniques is beyond the scope of this work, interested readers are referred to the work of Fortun et al. (2015), to get deeper insight into optical flow computation methods, and to the survey on block-based methods by Jakubowski and Pastuszak (2012).

In this paper, we focus on methods that allow for sufficiently fast motion estimation, i.e. for which fast implementations are available or which have the potential to be implemented in a near-to-realtime fashion. Similar as the Middlebury (Baker et al., 2011) and MPI Sintel (Butler et al., 2012) benchmarks, this survey and benchmark paper is open to be extended to any motion estimation technique, e.g. for more accurate (and potentially slower) approaches in the future.

The original approaches on the calculation of *optical flow* have been proposed by Horn and Schunck (1980) and Lucas and Kanade (1981). They assumed that every change in a pixel's brightness is due to motion. They compute the flow field using brightness gradients and a constraint on motion smoothness. Brox et al. (2004) extended this assumption by a gradient constancy constraint to deal with slight changes in brightness and an enhanced smoothness assumption. Another approach by Zach et al. (2007) is based on total variation (TV) regularization, using the L^1 norm (TV-L^1) and claims to be very robust against illumination changes and occlusions. Both, Brox et al. (2004) and Zach et al. (2007) are available as real-time GPU-based implementation. Werlberger et al. (2009) proposed to replace the TV regularization with the Huber norm (Huber-L^1). They presented a library called *FlowLib*, which contains GPU accelerated implementations of their algorithm in different variations. Additionally, Werlberger (2012) proposed alternative data terms, representing the structure of the image rather than intensities. Their normalized cross-correlation (NCC), census transform and consistency of gradients approaches provide better compensation for intensity variations.

More modern variants that surpass the accuracy of the aforementioned approaches and could be better suited for FS imagery include the *Large Displacement Optical Flow* (LDOF) presented by Brox and Malik (2011) as well as the *EpicFlow* described by Revaud et al. (2015). Both methods deal with the common problem of variational optical flow methods, which tend to select the local minimum closest to the initialization, i.e., a well matching point with the smallest motion. For this purpose, LDOF incorporates descriptor matching techniques into the variational approach to emphasize matches with higher accuracy even in the presence of similar looking image areas. EpicFlow relies on a sparse-to-dense approach which detects and preserves edges. The *FlowFields* method presented by Bailer et al. (2015) builds up on the edge preserving interpolation of EpicFlow, but improves on its results by using a new hierarchical correspondence field search strategy based on either census or SIFTflow as data term.

The basic idea of *block matching*, on the other hand, is to divide an image into *macro blocks* of a given block size b and to find the best matching block in a reference image using error functions such as the sum of absolute differences (SAD).

Usually, only translational motion is taken into account. To avoid blocking artifacts at object boundaries, different techniques such as overlapping blocks, adaptive block size, multiscale approaches and filtering have been proposed (Choi et al., 2007). The search range can be limited to a maximum displacement range (*p-value*). A simple full search tests all possible block displacements within this range. More efficient search strategies can be applied, e.g., temporal motion prediction, which reduce the number of calculations at the cost of accuracy (Cuevas et al., 2013). Due to its high degree of parallelism, BM can be efficiently implemented on GPUs or FPGAs to achieve real-time processing.

Recently, motion estimation based on convolutional *deep neural networks* emerged (Dosovitskiy et al., 2015) and soon surpassed OF and BM methods in quality, as demonstrated e.g. by the KITTI benchmark (Menze and Geiger, 2015). As acquiring ground truth motion data for training purposes is challenging, unsupervised variants have been proposed (Ren et al., 2017), but they do not typically reach the same level of quality. In this paper, we evaluate the recent pre-trained networks FlowNet2 (Ilg et al., 2017) and LiteFlowNet (Hui et al., 2018) in the context of FS imagery.

Steiner et al. (2016) is the first work addressing the motion compensation problem for multispectral short wave infrared (SWIR) FS imagery. They estimate forwards and backwards motion fields using state of the art OF methods for each pair of related waveband channels, which are intensity consistent by nature. Due to the interpolation over large time spans, this *corresponding channel matching (CCM)* approach yields rather poor results on scenes involving non-constant motion.

In the context of Time-of-Flight (ToF) cameras, various motion estimation approaches have

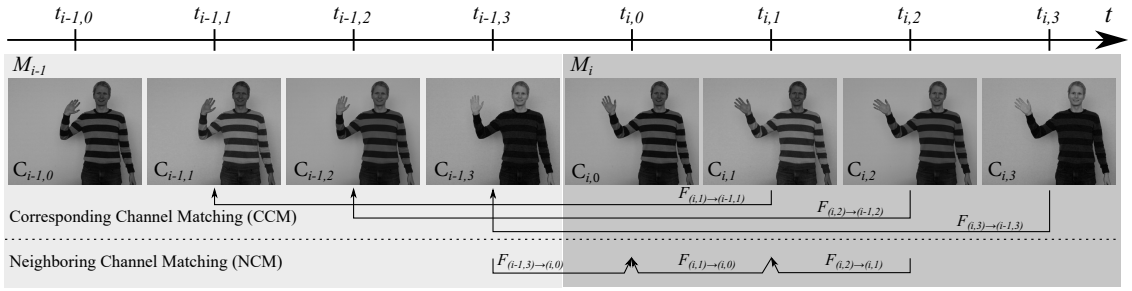


Figure 2: Flow calculations for two successive FS images with four channels using either corresponding channel matching (CCM) or neighboring channel matching (NCM).

been proposed. Most of them apply OF(Lindner and Kolb, 2009; Lefloch et al., 2013) or BM (Högg et al., 2013) on the set of phase images, thus implementing a *neighboring channel matching (NCM)* approach. To deal with the intensity inconsistency problem, they estimate motion on normalized phase intensity images. A different and computationally very efficient approach to motion compensation has been proposed by Schmidt and Jähne (2011). Their *pixelwise artifact correction (PAC)* method detects motion artifacts on pixel level by assuming temporally smooth intensity variation in the non-motion case: if the first channel contains no discontinuity but at least one of the following channels does, then the pixel is assumed to be affected by motion.

3. Field Sequential Motion Estimation Schemes

Consider a field sequential (FS) image stream consisting of images M_i, with $i \in \mathbb{N}$ being a sequential number, which themselves contain n channels $C_{i,w}$, which were acquired at sequential times $t_{i,w},\ w = 0, \ldots, n-1$, with w being the channel index. Furthermore, a discrete and equidistant acquisition time $\Delta t = t_{i,w} - t_{i,w-1}$ is assumed for each channel and a constant acquisition time $T = t_{i,0} - t_{i-1,0} = n\Delta t$ for the full image, as illustrated in Fig. 2. Formally, motion estimation for FS image streams has to compute the displacement vector fields $F_{(i,w)\to(i,0)}$ between any channels $C_{i,w},\ w > 0, \ldots, n-1$ and the first channel $C_{i,0}$, which serves as reference.

By applying a displacement vector field $F_{(i,w)\to(i,0)}$ to channel $C_{i,w}$, all pixel values $p(x, y)$ from $C_{i,w}$ are shifted according to the two-dimensional displacement vectors $\vec{d}(x, y) = F_{(i,w)\to(i,0)}(x, y)$. In the final "corrected" images $\tilde{C}_{i,w}$, the positions of moving objects will match those in the reference channel $C_{i,0}$, if the motion estimation has been accurate.

When optical flow is calculated directly between adjacent channels of the image sequence, i.e. between $C_{i,w}$ and $C_{i,w+1}$, purely intensity-based optical flow algorithms will produce invalid displacement vectors due to the violation of the intensity consistency assumption. In the following we describe two fundamental concepts to overcome this problem, i.e., *corresponding channel matching* (CCM; see Sec. 3.1) and *neighboring channel matching* (NCM; see Sec. 3.2), and discuss

approaches to modify existing motion estimation techniques in order to be applied with either concept.

3.1. Corresponding Channel Matching (CCM)

By using two consecutive FS images M_{i-1} and M_i and estimating motion only between pairs of corresponding channels, as shown in Fig. 2, the violation of the intensity inconsistency problem can be avoided. Despite the larger displacement between the compared images, state-of-the-art motion estimation techniques will most likely produce accurate displacement vectors based on this method. Assuming a constant and linear motion between corresponding channels $C_{i-1,w}$ and $C_{i,w}$, every flow vector $F_{(i-1,w)\to(i,w)}(x,y)$ is regarded as a linear combination of n identical partial vectors describing a pixels movement between $C_{i,w}$ and $C_{i,w-1}$,

$$F_{(i,w)\to(i,w+1)}(x,y) \equiv \frac{1}{n} F_{(i-1,w)\to(i,w)}(x,y). \tag{1}$$

Bidirectional, All Channel Optical Flow (CCM-B). The CCM method presented by Steiner et al. (2016) calculates a *forward flow* $F_{(i-1,w)\to(i,w)}$ and a *backward flow* $F_{(i,w)\to(i-1,w)}, w = 1,\ldots,n-1$ for each pair of channels $\langle C_{i-1,w},\ C_{i,w}\rangle$, $w > 0$. Both forward and backward flow are applied with weights $\frac{(n-w)}{n}$ and $\frac{w}{n}$ in order to interpolate a motion corrected channel $\tilde{C}_{i,w}, w = 1,\ldots,n-1$, for the reference time $t_{i,0}$:

$$\tilde{C}_{i,w} = \tfrac{(n-w)}{n} F_{(i-1,w)\to(i,w)}[C_{i-1,w}] \oplus \tfrac{w}{n} F_{(i,w)\to(i-1,w)}[C_{i,w}]. \tag{2}$$

The bidirectional interpolation function $\oplus$ calculates the intensity of every pixel in $\tilde{C}_{i,w}$ by averaging the corresponding pixel values in both $C_{i-1,w}$ and $C_{i,w}$. In conjunction with the detection of occlusions, this function provides high interpolation accuracy. The main disadvantage of this approach is its extremely high computational complexity, as it requires $2\cdot(n-1)$ OF calculations for each FS image.

In the following, we discuss alternative approaches that reduce computational complexity, but generally also reduce the accuracy of compensation; see Sec. 7.

Unidirectional, All Channel Optical Flow (CCM-U). This approach simplifies the interpolation method by using only one OF calculation for each pair of channels $C_{i-1,w}$ and $C_{i,w}$, $w = 1,\ldots,n-1$. It uses either the forwards or backwards flow depending on the current channel, to keep the length of the resulting motion vectors and, thus, the expected error as small as possible:

$$\tilde{C}_{i,w} = \begin{cases} \frac{w}{n} \cdot F_{(i,w)\to(i-1,w)}[C_{i,w}] & \text{if } w \le \frac{n}{2} \\ \frac{(n-w)}{n} \cdot F_{(i-1,w)\to(i,w)}[C_{i-1,w}] & \text{if } w > \frac{n}{2} \end{cases} \tag{3}$$

Unidirectional, Partial Channel Optical Flow (CCM-1 / CCM-2). The number of OF calculations can be further decreased by interpolating a given flow field to subsequent channels. Assume that the backwards flow $F_{(i,u)\to(i-1,u)}$ for channel u is known and the motion is constant during the acquisition time of both FS images. Then, the backward flow $F_{(i,v)\to(i-1,v)}$ for channel $v > u$ can be interpolated using $F_{(i,u)\to(i-1,u)}$:

$$F_{(i,v)\to(i-1,v)} = -\frac{v-u}{n} \cdot F_{(i,u)\to(i-1,u)}[F_{(i,u)\to(i-1,u)}]. \tag{4}$$

The motion corrected image $\tilde{C}_{i,v}$ is calculated according to Eq. 3 (case $w \leq \frac{n}{2}$). This way, the number of required optical flow calculations for each cube can be reduced down to one (CCM-1). However, the more flow fields are interpolated from a previous one, the higher the approximation error will be if the assumption of constant motion does not hold true. Neglecting this fact, one calculated flow field could even be extrapolated over several image cubes, further reducing the processing time at the cost of an even higher approximation error. In practice and depending on the amount and nature of expected motion in the scene, it seems to be a better choice to interpolate only a limited number of flow fields from others.

In the case of four (or more) channels per FS image, a more accurate interpolation can be achieved if a second flow field $F_{(i,w)\to(i-1,w)}$ of a subsequent channel w is used to bidirectionally interpolate $F_{(i,v)\to(i-1,v)}, u < v < w$ (CCM-2):

$$F_{(i,v)\to(i-1,v)} = -\frac{v-u}{n} \cdot F_{(i,u)\to(i-1,u)}[F_{(i,u)\to(i-1,u)}] \oplus \frac{w-v}{n} \cdot F_{(i,w)\to(i-1,w)}[F_{(i,w)\to(i-1,w)}]. \tag{5}$$

3.2. Neighboring Channel Matching (NCM)

The NCM approach commonly applied to ToF images (Lindner and Kolb, 2009; Lefloch et al., 2013; Högg et al., 2013) estimates motion fields $F_{(i,w)\to(i,w-1)}$ between adjacent channels $C_{i,w}$, $w > 0$ directly. To compensate motion in $C_{i,w}$, all partial flow fields $F_{(i,w)\to(i,w-1)}$ are applied to $C_{i,w}$ sequentially:

$$\tilde{C}_{i,w} = F_{(i,1)\to(i,0)}[\ldots[F_{(i,w)\to(i,w-1)}[C_{i,w}]]]. \tag{6}$$

NCM is a potentially more accurate alternative to CCM, as it keeps the object displacement minimal for each flow calculation and allows to compensate dynamic changes of motion speed and direction during the acquisition of the FS image. Obviously, though, it has to handle the intensity inconsistency problem between different spectral channels.

The expected total interpolation error can further be reduced by estimating motion between adjacent channels of two neighboring FS images forwards or backwards towards the closest reference channel; see Fig. 2. The compensated image $\tilde{C}_{i,w}$ can then be found by sequentially applying

the resulting flow vectors either forwards or backwards:

$$\tilde{C}_{i,w} = \begin{cases} F_{(i,1)\to(i,0)}[[F_{(i,w)\to(i,w-1)}[C_{i,w}]]] & \text{if } w \leq \frac{n}{2} \\ F_{(i-1,n-1)\to(i,0)}[[F_{(i-1,w)\to(i-1,w+1)}[C_{i-1,w}]]] & \text{if } w > \frac{n}{2} \end{cases} \tag{7}$$

4. Image Transformation and Correlation

In general, the realization of NCM methods requires handling the intensity differences between neighboring channels. Here, three different types of operations can be applied:

1. transformation of the image (channel) into another domain (e.g. gradients),
2. finding dense correlations between neighboring channels (using, e.g., cross-correlation), or
3. applying intensity correction (using, e.g., equalization)

Even though intensity transformation and correlation approaches can be applied sequentially, this is rarely done in literature. Therefore, we decided to apply only one of the methods and describe the related methods in this section. Intensity correction methods are summarized in Sec. 5.

The following image transformation approaches are evaluated in this respect, in Sec. 7.

Census Transform, proposed by Zabih and Woodfill (1994), describes the local spatial structure around a specific pixel of an image by calculating a binary vector $p_{x,y}$ for each pixel: if a neighboring pixel has a lower intensity than $p_{x,y}$, a 1 will be added to the vector, otherwise a 0. After the transformation, correspondence is calculated by finding the minimum Hamming distance.

Image Gradients describe the intensity variations in a pixel's local neighborhood and can be computed, e.g., using a Sobel filter (Jähne, 2005); see Fig. 3 f).

We investigated the following correlation based approaches that can be applied to two images in order to solve the intensity inconsistency problem:

Mutual Information is based on the entropy of an image pair and yields a high value if the information gain of a new image in addition to an existing image is low, i.e., if two images of the same scene are geometrically aligned. Mutual information is known to be robust against non-linear intensity relationships and has been proposed for both multispectral and multimodal image registration applications (Zitova and Flusser, 2003; Kern and Pattichis, 2007). It can be used as a cost function for block matching, but it cannot be linearized for the use in OF algorithms.

Cross-Spectral Feature Detection is frequently used in multispectral or multimodal image registration (Zitova and Flusser, 2003), where image transformation or warping parameters are estimated based on detected features. While these methods cannot be used to estimate dense motion fields between two images directly, they might be used for the registration of blocks in block matching algorithms.

Normalized Cross-Correlation **(NCC):** Cross-correlation is commonly used as cost function in order to find the position of specific features in an image (Jähne, 2005). NCC additionally normalizes the image which improves the robustness against illumination changes. NCC can be used as an inverse cost function for BM, as well as a linearized data term in OF (Steinbruecker et al., 2009; Werlberger, 2012).

Preliminary Method Selection. As we face the fundamental problem of combinatorial complexity when evaluating a large amount of approaches making up the final motion estimation method, we executed preliminary tests in order to exclude approaches for which we observe significant drawbacks in our context of motion estimation for FS imagery. Census transform, image gradients and cross-correlation deliver valuable results, so we use them in our exhaustive evaluation in Sec. 7. Applying mutual information was found to be computationally extremely expensive[2]. Thus, we excluded mutual information as its application to a larger set of test sequences and motion estimation methods would be impracticable. Cross-spectral feature detection delivered significantly inferior results when testing with some of the our multispectral data sets from Sec. 7. Thus, we also excluded this method from our full evaluation.

5. Intensity Correction Methods

There are several ways to address the intensity inconsistency problem in case of NCM motion estimation using intensity correction approaches. In the following, we assume a grayscale image I that is intensity corrected, resulting in $\tilde{I}$. Fig. 3 illustrates their effect on the channels of an FS image.

The following approaches to reduce the intensity inconsistency between the spectral channels are evaluated in Sec. 7.

Global Linear Normalization is a simple linear mapping of the used intensity range $[I_{min}, I_{max}]$ to a new range $[0, \tilde{I}_{max}]$ (Petrou and Petrou, 2010); see Fig. 3 b).

[2]We tested the fast approximative implementation from Shams and Barnes (2007) in combination with block matching. Here, the average execution time for a single image pair with a resolution of 640×480 pixels and a (small) search window of 11x11 pixels requires $\approx 300s$ on a typical PC.

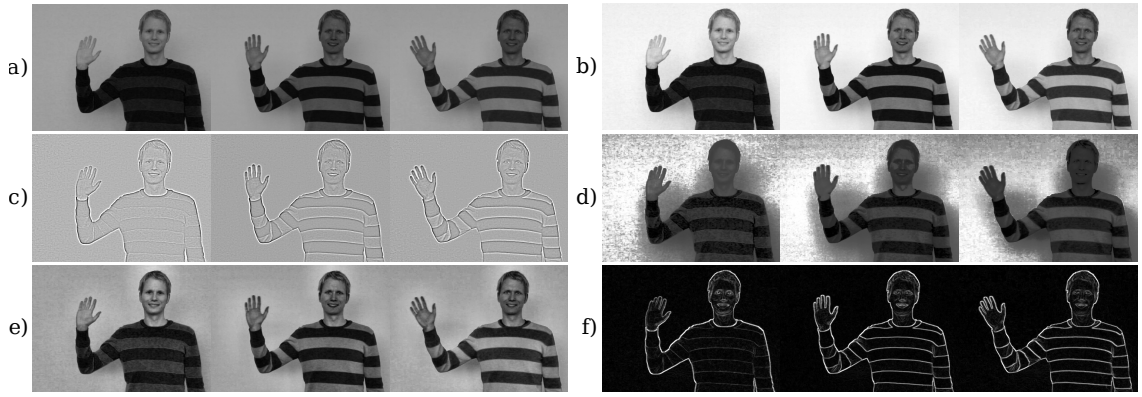


Figure 3: Examples of methods applied on an FS image with three channels: a) original image; b) global normalization; c) local normalization; d) histogram equalization; e) CLAHE; f) gradients.

Local Linear Normalization compensates for non-uni-form illumination within an image (Sage, 2011); see Fig. 3 c). Using the windowed mean $m_I(x, y)$ and variance $\sigma_I(x, y)$ for each pixel (x, y), the normalized intensity $\tilde{I}(x, y)$ computes as:

$$\tilde{I}(x, y) = \frac{I(x, y) - m_I(x, y)}{\sigma_I(x, y)}. \tag{8}$$

Histogram Equalization uniformly distributes the intensity values over the available intensity range (Petrou and Petrou, 2010); see Fig. 3 d). It normalizes the histogram $H(i)$ of an input image and calculates the cumulative distribution $H'(i)$, which is used to remap the intensity values:

$$\tilde{I}(x, y) = H'(I(x, y)) \text{ with } H'(i) = \sum_{0 \le j \le i} H(j). \tag{9}$$

Contrast Limited Adaptive Histogram Equalization (CLAHE) performs the histogram equalization in a local per-pixel window. As this operation tends to amplify noise in homogeneous areas, the CLAHE algorithm introduces a clipping limit for histogram redistribution (Pizer et al., 1987); see Fig. 3 e).

6. Evaluation Setup

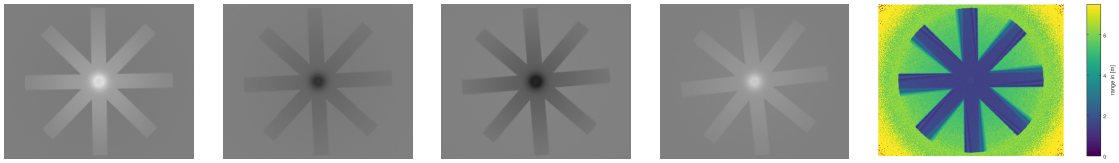
Datasets. For evaluation of the motion estimation methods for FS imagery we have prepared five different types of datasets. In all cases the number of channels is $n = 3$ or $n = 4$.

Middlebury: All **Middlebury** evaluation samples with at least 8 frames from the Middleburry benchmark[3] by Baker et al. (2011).

[3] see `http://vision.middlebury.edu/flow`

(a) Linear stage. (b) Rotating wheel. (c) Head tilting. (d) Walking.

Figure 4: Examples of the test scenarios included in the FS multimodal dataset



(a) Phase image 0. (b) Phase image 1. (c) Phase image 2. (d) Phase image 3. (e) Range image.

Figure 5: Sequence of four simulated ToF raw phase images (channels) and the resulting depth frame.

Please note that sequences with less than 8 frames, as well as other, frequently used datasets such as KITTI that do not provide RGB sequences of sufficient length cannot be incorporated in our evaluation.

From each of the sample sequences we derive an **RGB** FS image sequence with ground truth images by dropping two of the three color channels yielding a 3-channel FS image sequence. Furthermore, we generate **RGB-R** FS image sequences that contain a fourth channel generated by converting the RGB image into a gray scale image with reduced brightness. This channel resembles the so-called *dark reference frame* commonly acquired in *active multispectral camera systems* in order to subtract background illumination.

MPI Sintel: The **MPI Sintel** dataset[4] by Butler et al. (2012) (marked as "final") are used in the **RGB** and **RGB-R** dataset.

Own sRGB: This dataset is acquired using an active multispectral SWIR video sequence for three different scenarios (see Fig. 4): *Linear stage* (laterally moving test pattern), *rotating wheel*, and *human movement* showing a person's upper-body performing several movement patterns. We generated FS data **RGB, RGB-R** analogous to **Middlebury** and **MPI Sintel**.

Short Wave Infrared (SWIR): The human movement data in **Own sRGB** provides the SWIR test data. Ground Truth is available using the additional RGB video stream (see below for

[4]see `http://sintel.is.tue.mpg.de/`

details).

Time-of-Flight (ToF): ToF ground truth is extremely hard to access, as the channels, commonly called phase images in ToF imaging, are directly processed in the camera and there is no option to either trigger the exposure of phase images nor to access the explicit timing of the exposition. Therefore, we use simulated ToF imagery based on Lambers et al. (2015); Bulczak et al. (2018), for which ground truth motion fields can be explicitly extracted. We used two sample scenes, one with a lateral moving object and one with a rotating star-like shape. Both scenes have been acquired with two different velocities.

The simulator generates four phase images (channels) per depth frame and includes noise and motion artifacts (see Fig. 5). We deactivate the simulation of background intensity in order to be able to apply the same quality measures as for the multispectral data sets that do not include this intensity bias.

Note that all datasets except MPI Sintel contain indoor scenes as motion estimation in the context of both multispectral and ToF sensors typically targets such scenes.

Details for RGB and MS Dataset Acquisition. The multispectral datasets have been captured using an active FS-based SWIR camera system with three wavebands and a dark reference channel (-R), in combination with a high quality RGB camera with the same frame rate. The cameras were arranged in a staring imager configuration in a well-lit environment, which is common for FS NIR imaging systems. Although the subject is the same in all of the human movement sequences, different movement patterns have been captured to provide diversity.

Any negative effect caused by demosaicing of the RGB camera's Bayer pattern is accounted for by recording in high definition with 1920x1080 pixels and downsampling the images to the resolution of the SWIR camera's images, i.e. 636x508 pixels.

The ground truth for our MS dataset is created from the additional RGB sequences. As the SWIR camera system uses $n = 4$ channels, the RGB camera simultaneously acquires four virtual channels, i.e. R, G, B, and dark reference (-R). For evaluation, we use a cross-compensation approach that applies the optical flow calculated for the SWIR imagery to the RGB image sequence and compares the result to the corresponding RGB full frame.

To match the field of view of the RGB camera to the SWIR camera, the RGB imagery is shifted and cropped appropriately. However, the baseline between both cameras of $\approx 20cm$ induces a slightly different perspective and thus a mismatch in the motion fields. To estimate this mismatch, we recorded a second dataset where the SWIR camera was replaced by a second RGB camera. Applying the same cross-compensation procedure to this stereo-like setup, we found a *baseline error* for the comparison $\mathrm{IE}_{\mathrm{base}} \approx 2.7$. As $\mathrm{IE}_{\mathrm{base}}$ is by far lower than the error of the best FS

motion compensation method with IE ≈ 6.6, we find our cross-compensation approach to be valid within this range.

Quality Measures. The objective comparison of a compensated image with the ground truth image is performed using the following quality measures:

1. *Interpolation Error (IE)* (Baker et al., 2011) is defined as the root mean square of the L2 norm of the vector of spectral channel differences between the interpolated and ground truth images, analog to the Middlebury OF evaluation,
2. *Structural Similarity Index Metric (SSIM)*, which describes the similarity of images based on structural information and is inspired by the human visual perception (Wang et al., 2004), and
3. *Spectral Error (SE)*, which we define as the root mean square of all pixel's spectral angular distance (Petrou and Petrou, 2010).

7. Results and Discussion

With 379 combinations of methods and preprocessing options, the total amount of results is very extensive. Here, we present only a representative selection and summarize the findings. The complete results can be found in the digital supplemental material, which presents all details about influences of individual steps and changes of preprocessing methods or algorithms to the results.

Table 1 states all methods and algorithms applied and explains the abbreviations used in the following evaluation. CCM and NCM are the **General Concepts** applicable to motion estimation for FS imagery. For CCM, we need to specify the **Motion Estimation Scheme** that defines how the full flow for an FS image frame is computed, e.g., using the uni- or bidirectional approach (see Sec. 3). Remember that CCM-2 only makes sense if we have four (or more) channels per FS image, thus CCM-2 can only be applied to **RGB-R** and **ToF** data sets. Either of the resulting concept-scheme combinations is applied to the original or a transformed version of the channels (**Image Transformation**; see Sec. 4) that has optionally been processed by an **Intensity Correction** method (see Sec. 5). The resulting combination can be implemented using any kind of BM or dense OF algorithm. In this work, GPU-accelerated implementations of Brox, TV-L1, Lucas-Kanade (LK), LDOF and Huber-L1-based optical flow, as well as full search and fast approximate BM algorithms from standard libraries (OpenCV 2.4.11 and FlowLib 3.0) are used. They are complemented by a (multithreading) CPU implementation of FlowFields, which is not currently available as a GPU-accelerated version. All Brox- and FlowLib-based algorithms have been applied twice, once with recommended (quality-oriented) parameters and once with parameters optimized

General concept	
CCM	corresp. channel matching
NCM	neighboring channel matching
CCM Motion Estimation Scheme	
-B	all channels bidirectional
-U	all channels unidirectional
-2	partial (2 channels)
-1	partial (1 channel)
Image Transformation	
-I	intensity (i.e. no transformation)
-TG	transformation to gradients
-TC	transformation to census
-C	correlation
Intensity Correction	
N	global normalization
L	local normalization
H	histogram equalization
C	contrast limited adaptive histogram equalization
Algorithms	
BM	Block Matching, sum of absolute differences
FBM	Fast Block Matching, BM with restricted set of candidates
LK	Lucas-Kanade OF
Br	Brox OF
TVL1	TV-L1
HL1	Huber-L1
FHL1	Fast Huber-L1, parameters optimized for speed (Werlberger et al., 2009)
HQS	Huber-L1 with quadratic fitting, sum of absolute differences (Werlberger, 2012)
H1C	Huber regularization term, L1 data term, compensation of brightness constancy violations
H2C	Huber regularization term, L2 data term, compensation of brightness constancy violations
TGVC	2nd order Total Generalized Variation w. Census transform (Bredies et al., 2010; Ranftl et al., 2014)
FF	FlowFields
LDOF	Large Displacement Optical Flow
PAC	Pixelwise Artifact Correction
FN	FlowNet2 (Ilg et al., 2017)
LFN	LiteFlowNet (Hui et al., 2018)

* = optimized for speed by authors of this paper, see Sec. 7

** = part of algorithm

Table 1: Abbreviations used in Tab. 2, Fig. 6 and Fig. 7, 8, 9, 10, 11.

for speed, which is denoted with a *. Optimal parameters have been found experimentally[5].

As some of the algorithms already include an image transformation, e.g. to gradients, we explicitly mark this with **. In our evaluation we also feed gradient images to these algorithms, leading to an overall image transformation of type TG+TG**. Some algorithms internally use intensity and gradient images for estimating motion, denoted as (I+TG)**. If fed with gradient images, these algorithms read TG+(I+TG)**.

Computational efficiency is measured on a standard desktop computer with a recent Intel CPU and nVidia graphics card using the FS multispectral **RGB-R** datset. Note, however, that there is some variation in system setup between methods, so the timing information in Tab. 2 and Fig. 6 should be interpreted as a rough estimate.

Analog to the Middlebury evaluation, all methods were ranked for each test sequence based on all described quality measures. Tab. 2 shows the average ranks of the top-20 combinations of algorithms and approaches with respect to multispectral dataset (including **Own sRGB**, **Middlebury** and **MPI Sintel**) in the upper part. The middle part of Tab. 2 shows the ranking with respect to the **ToF** data set. For comparison, results of the original algorithms without optimization for FS data sequences and different CCM optimizations have been added in the lower part. In addition, Fig. 6 allows to easily compare the motion compensation performance to the computational efficiency of the different methods. A higher resolution plot can also be found in the supplemental material. All abbreviations are explained in Tab. 1.

To illustrate the performance of different approaches, example images from each dataset and a selection of motion compensation results are shown in Fig. 7, 8, 9, 10, 11.

Comparing CCM and NCM. The multispectral data sets (including **Middlebury** and **MPI Sintel**) are handled best with NCM methods. This is mainly due to the fact, that multispectral data sets exhibit less intensity inconsistencies than ToF data. Furthermore, non-linear motion can be better captured using neighboring channel matching (NCM) due to shorter interpolation intervals. CCM-methods, on the other hand, produce primarily superior results on **ToF** data sets. However, several NCM methods work well on **ToF** data sets. Namely NCM-TG+TG** with the HL1 algorithm, which is among the top-20 for both, multispectral and ToF.

Varying the number of OF in CCM. Tab. 2 includes CCM results using all channels bi- (CCM-B) and unidirectional (CCM-U), first and last channel (CCM-2), as well as first channel only (CCM-1) based on the Brox algorithm that performs best for CCM. For all algorithms, a reduction of the number of OF calculations decreases the processing time almost proportionally, while

[5]BM: search field parameter $p = 20$, block size $bs = 25$; Brox: 3 instead of 10 inner and solver iterations each; FlowLib: 3 instead of 10 iterations and warps each.

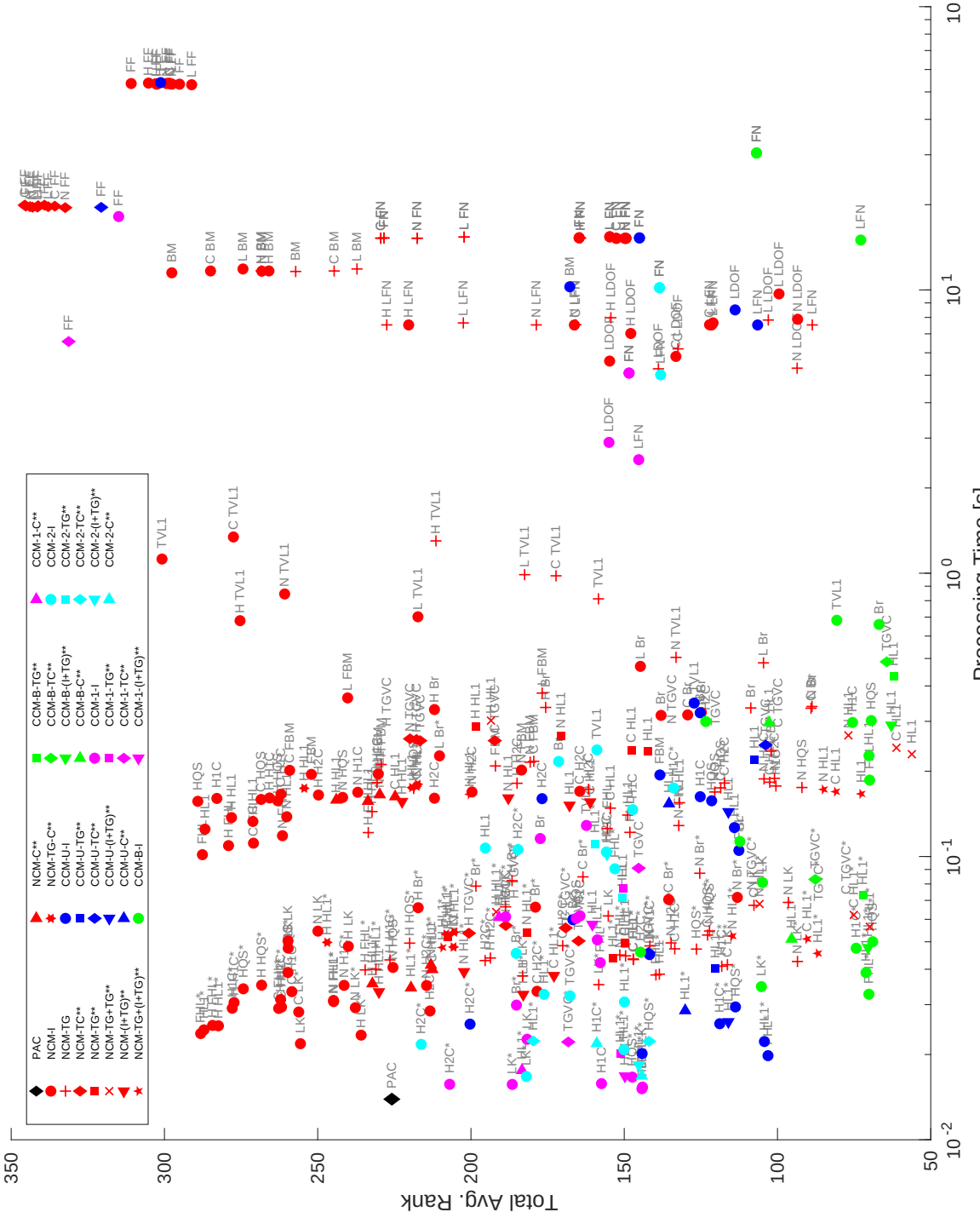


Figure 6: Scatter plot showing accuracy and processing time of the described approaches. An SVG version of this plot is available in the supplementary material.

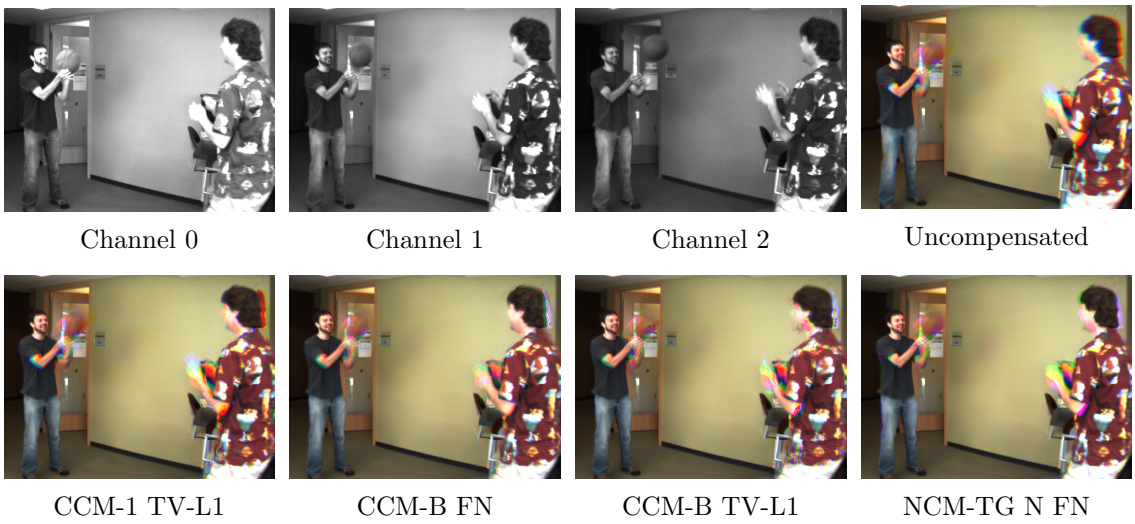


Figure 7: Examples for an image before and after motion compensation (dataset Middlebury)

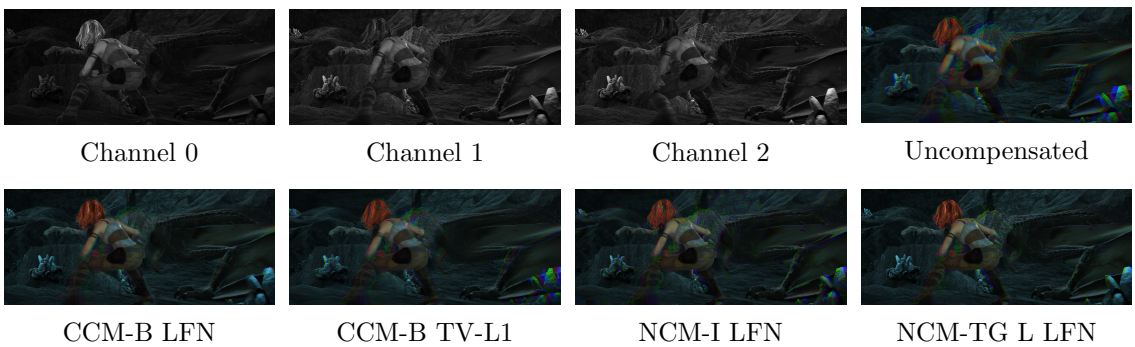


Figure 8: Examples for an image before and after motion compensation (dataset MPI Sintel)

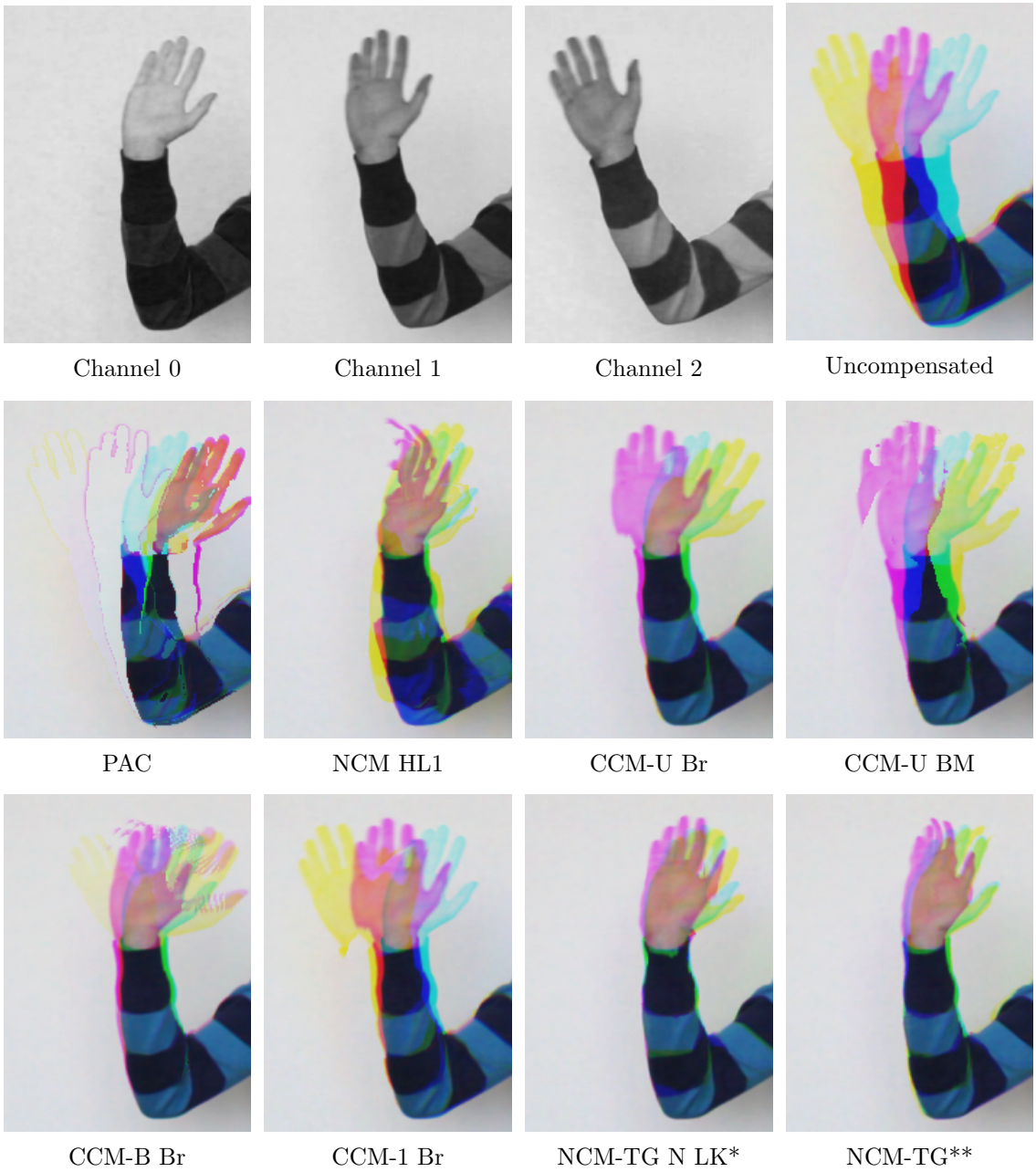


Figure 9: Examples for an image before and after motion compensation (dataset Own sRGB)

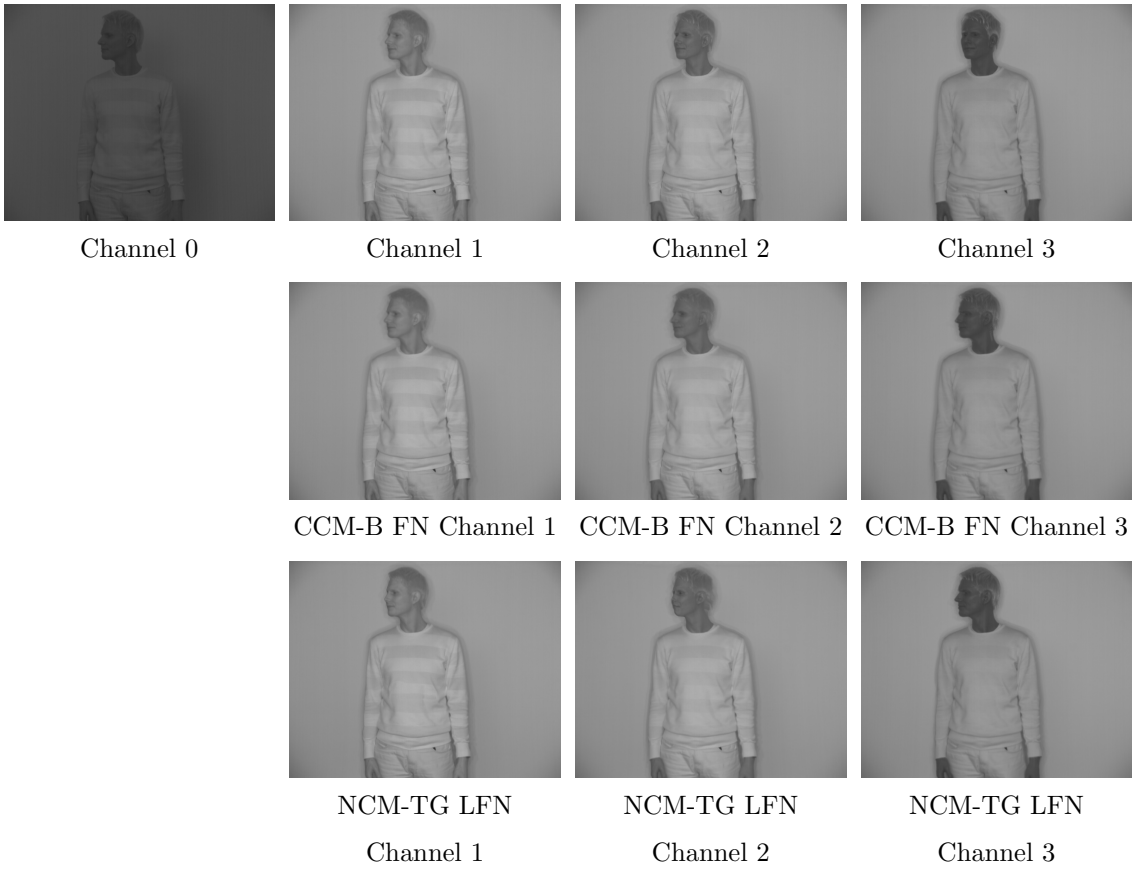


Figure 10: Examples for an image before and after motion compensation (dataset SWIR)

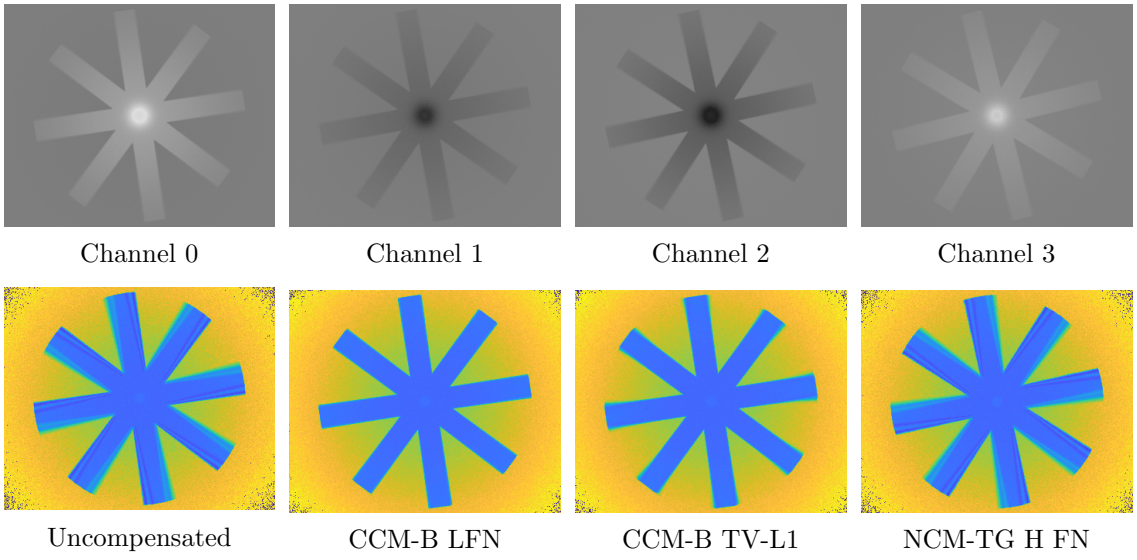


Figure 11: Examples for an image before and after motion compensation (dataset ToF)

simultaneously increasing the error in a very predictable way; see Fig. 6.

Handling of Inconsistent Intensities with NCM. Without image transformation and intensity correction, only the Brox algorithm (NCM-I-Br) is capable of handling the inconsistent intensities with NCM-I methods to some degree. Applying image transformation only (NCM-∗-<none>), transformation to gradient, potentially applied twice, i.e. during preprocessing and, again, within the algorithm itself, clearly yields the best results. The census transformation and correlation-based methods (NCM-C) cannot compete. Applying intensity correction only (NCM-I-∗), Brox and LDOF perform well on multispectral data if global or local normalization is applied.

Influence of the OF Algorithm. Ignoring image transformation and intensity correlation-based, there is no clear tendency in terms of OF algorithms, neither for the multispectral nor for the ToF data sets. While LDOF, as one of the more modern approaches is quite successful on multispectral data sets, more classical approaches like Brox and Huber-based algorithms yield comparable results; on ToF data sets they even dominate. Surprisingly, the most modern algorithm in the evaluation, FlowFields, performed worst. A possible explanation for this finding could be that it's SIFTflow matching approach is optimized for color rather than grayscale images.

When taking processing time into account, the normal and speed-optimized Huber-L1 (HL1, FHL1), as well as Lucas-Kanade optical flow (LK) deliver outstanding results.

Pixelwise Artifact Correction (PAC). This approach from Schmidt and Jähne (2011) is the only one specifically developed to correct ToF raw data. It performs comparably bad regarding quality, but the approach is computationally very effective and fast, although it does not rely on GPU acceleration.

Deep Neural Networks. The deep neural network based methods FlowNet2 (FN) and LiteFlowNet (LFN) both perform well. Note that we use publically available pre-trained implementations of both. These have been trained on RGB data, whereas here we apply them to individual channels of our FS imagery which are interpreted as grayscale images. Training either method specifically for a given FS type (multispectral or ToF) will likely result in improved quality.

8. Conclusions

This paper presents and evaluates approaches to apply existing motion estimation methods to field-sequential (FS) imagery, originating from multispectral dynamic scene captures or Time-of-Flight cameras. The major challenge here is the assumption of consistent intensities for corresponding pixels made by most motion estimation approaches, which is in general not fulfilled for adjacent channels of FS imagery.

While corresponding channel matching (CCM) methods estimate motion fields between corresponding channels of successive FS images to avoid intensity inconsistencies, neighboring channel matching (NCM) estimates motion fields between neighboring channels within a single FS image, which requires a successful handling of inconsistent intensities between the channels but (potentially) benefits from interpolation for shorter time intervals and displacement vectors.

We combine existing motion estimation schemes with known image transformation and/or intensity correction methods, leading to an overall set of 379 combinations of FS motion compensation approaches, implemented using state of the art algorithms.

We present the new *FS database* containing datasets with ground truth acquired using RGB, multispectral SWIR, and ToF camera simulators, which will be available to the scientific public in order to promote further research in this field. Our evaluation also involves data from the **Middlebury** and the **MPI Sintel** datasets.

Due to the variety in the FS database that includes also strongly intensity inconsistent ToF phase images as well as moderate inconsistent multispectral imagery, there is not "the best" method superior to others. There is, however, a clear tendency, that NCM methods are more successful for moderate intensity inconsistency. For strong intensity inconsistency, CCM methods perform best, while NCM in combination with gradient transformation (potentially applied twice) still give good results.

Acknowledgments.. This research was partially funded by German Research Foundation (DFG) within the Research Training Group GRK 1564 "Imaging New Modalities" and the DFG project grant KO-2960/12-1.

Data sets.. The full data used in this paper is available at `https://www.cg.informatik.uni-siegen.de/data/fsmotion2019` (this location will change for a final release).

References

S. Daly, X. Feng, Method and system for field sequential color image capture using color filter array, 2004. US Patent 6,690,422.

A. Gowen, C. O'Donnell, P. Cullen, G. Downey, J. Frias, Hyperspectral imaging – an emerging process analytical tool for food quality and safety control, Trends in Food Science & Technology 18 (2007) 590–598. doi:doi:10.1016/j.tifs.2007.06.001.

J. Brauers, T. Aach, S. Helling, Multispectral image acquisition with flash light sources, Journal of Imaging Science and Technology 53 (2009) 31103–1–31103–10. doi:doi:doi:10.2352/J.ImagingSci.Technol.2009.53.3.031103.

S. Helling, E. Seidel, W. Biehlig, Algorithms for spectral color stimulus reconstruction with a seven-channel multispectral camera, Conf. on Colour in Graphics, Imaging, and Vision 2004 (2004) 254–258.

T. Bourlai, N. Narang, B. Cukic, L. Hornak, On designing a swir multi-wavelength facial-based acquisition system, in: Infrared Technology and Applications XXXVIII, volume 8353, 2012, p. 83530R. doi:doi:10.1117/12.919392.

N. Gat, Imaging spectroscopy using tunable filters: a review, in: Proc. SPIE Wavelet Applications VII, volume 4056, 2000, pp. 50–64. doi:doi:10.1117/12.381686.

H. Steiner, S. Sporrer, A. Kolb, N. Jung, Design of an active multispectral swir camera system for skin detection and face verification, Journal of Sensors 2016 (2016). doi:doi:10.1155/2016/9682453, article ID 9682453.

R. Lange, P. Seitz, Solid-state time-of-flight range camera, IEEE Journal of quantum electronics 37 (2001) 390–397.

A. Kolb, E. Barth, R. Koch, R. Larsen, Time-of-Flight cameras in computer graphics, in: Computer Graphics Forum (Eurographics STAR), volume 29, 2010, pp. 141–159.

M. Lambers, S. Hoberg, A. Kolb, Simulation of Time-of-Flight sensors for evaluation of chip layout variants, IEEE Sensors 15 (2015) 4019–4026.

S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, R. Szeliski, A database and evaluation methodology for optical flow, Int. J. of Computer Vision 92 (2011) 1–31. doi:doi:10.1007/s11263-010-0390-2.

D. Fortun, P. Bouthemy, C. Kervrann, Optical flow modeling and computation: A survey, Computer Vision and Image Understanding 134 (2015) 1 – 21. doi:doi:10.1016/j.cviu.2015.02.008.

M. Jakubowski, G. Pastuszak, Block-based motion estimation algorithms — a survey, Opto-Electronics Review 21 (2012) 86–102. doi:doi:10.2478/s11772-013-0071-0.

D. J. Butler, J. Wulff, G. B. Stanley, M. J. Black, A naturalistic open source movie for optical flow evaluation, in: A. F. et al. (Eds.) (Ed.), European Conf. on Computer Vision (ECCV), Part IV, LNCS 7577, Springer-Verlag, 2012, pp. 611–625.

B. K. Horn, B. G. Schunck, Determining optical flow, in: Proc. SPIE, volume 0281, 1980, pp. 319–331. doi:doi:10.1117/12.965761.

B. D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: Proc. Int. Conf. Artificial Intelligence (IJCAI), IJCAI'81, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1981, pp. 674–679.

T. Brox, A. Bruhn, N. Papenberg, J. Weickert, High accuracy optical flow estimation based on a theory for warping, in: T. Pajdla, J. Matas (Eds.), Computer Vision (ECCV), volume 3024 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2004, pp. 25–36. doi:doi:10.1007/978-3-540-24673-2_3.

C. Zach, T. Pock, H. Bischof, A duality based approach for realtime tv-l1 optical flow, in: Proc. DAGM Conf. Pattern Recognition, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 214–223.

M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, H. Bischof, Anisotropic huber-l1 optical flow, in: Proc. British Conf. Machine Vision (BMVC), London, UK, 2009.

M. Werlberger, Convex Approaches for High Performance Video Processing, Ph.D. thesis, Institute for Computer Graphics and Vision, Graz University of Technology, Graz, Austria, 2012.

T. Brox, J. Malik, Large displacement optical flow: Descriptor matching in variational motion estimation, IEEE Trans. on Pattern Analysis and Machine Intelligence 33 (2011) 500–513. doi:doi:10.1109/TPAMI.2010.143.

J. Revaud, P. Weinzaepfel, Z. Harchaoui, C. Schmid, Epicflow: Edge-preserving interpolation of correspondences for optical flow, in: Computer Vision and Pattern Recognition, 2015.

C. Bailer, B. Taetz, D. Stricker, Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation, in: Proc. IEEE Int. Conf. on Computer Vision, 2015.

B.-D. Choi, J.-W. Han, C.-S. Kim, S.-J. Ko, Motion-compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation, IEEE Trans. Circuits and Systems for Video Technology 17 (2007) 407–416. doi:doi:10.1109/TCSVT.2007.893835.

E. Cuevas, D. Zaldivar, M. A. Perez-Cisneros, D. Oliva, Block matching algorithm based on differential evolution for motion estimation, Engineering Applications of Artificial Intelligence 26 (2013) 488–498.

A. Dosovitskiy, P. Fischer, E. Ilg, P. Häusser, C. Hazırbaş, V. Golkov, P. v.d. Smagt, D. Cremers, T. Brox, FlowNet: Learning optical flow with convolutional networks, in: IEEE Int. Conf. Computer Vision (ICCV), 2015. doi:doi:10.1109/ICCV.2015.316.

M. Menze, A. Geiger, Object scene flow for autonomous vehicles, in: IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015. doi:doi:10.1109/CVPR.2015.7298925.

Z. Ren, J. Yan, B. Ni, B. Liu, X. Yang, H. Zha, Unsupervised deep learning for optical flow estimation, in: Proc. AAAI Conference on Artificial Intelligence, 2017. URL: `https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/viewPaper/14388`.

E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, T. Brox, Flownet 2.0: Evolution of optical flow estimation with deep networks, in: IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017. doi:doi:10.1109/CVPR.2017.179.

T. Hui, X. Tang, C. C. Loy, Liteflownet: A lightweight convolutional neural network for optical flow estimation, in: IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018. doi:doi:10.1109/CVPR.2018.00936.

M. Lindner, A. Kolb, Compensation of motion artifacts for Time-of-Flight cameras, in: Proc. Dynamic 3D Imaging, volume 5742 of *LNCS*, Springer, 2009, pp. 16–27.

D. Lefloch, T. Hoegg, A. Kolb, Real-time motion artifacts compensation of tof sensors data on GPU, in: Proc. SPIE Defense, Security - Three-Dimensional Imaging, Visualization, and Display, 2013, pp. 87380U–87380U–7.

T. Högg, D. Lefloch, A. Kolb, Real-time motion artifact compensation for PMD-ToF images, in: Proc. Workshop Imaging New Modalities, German Conference of Pattern Recognition (GCPR), volume 8200 of *LNCS*, Springer, 2013, pp. 273–288.

M. Schmidt, B. Jähne, Efficient and robust reduction of motion artifacts for 3d time-of-flight cameras, in: Proc. Int. Conf. 3D Imaging (IC3D), 2011, pp. 1–8. doi:doi:10.1109/IC3D.2011.6584391.

R. Zabih, J. Woodfill, Non-parametric local transforms for computing visual correspondence, in: Computer Vision (ECCV), volume 801 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 1994, pp. 151–158. doi:doi:10.1007/BFb0028345.

B. Jähne, Digital Image Processing, Springer Berlin Heidelberg, 2005.

B. Zitova, J. Flusser, Image registration methods: a survey, Image and Vision Computing 21 (2003) 977 – 1000. doi:doi:10.1016/S0262-8856(03)00137-9.

J. Kern, M. Pattichis, Robust multispectral image registration using mutual-information models, IEEE Trans. Geoscience and Remote Sensing 45 (2007) 1494–1505. doi:doi:10.1109/TGRS.2007.892599.

F. Steinbruecker, T. Pock, D. Cremers, Advanced data terms for variational optic flow estimation, in: Proc. Vision, Modeling and Visualization Workshop, 2009, pp. 155–164.

R. Shams, N. Barnes, Speeding up mutual information computation using nvidia cuda hardware, in: Proc. Conf. Digital Image Computing Techniques and Applications, 2007, pp. 555–560. doi:doi:10.1109/DICTA.2007.4426846.

M. Petrou, C. Petrou, Image Processing: The Fundamentals, 2 ed., John Wiley & Sons, 2010.

D. Sage, Local normalization, Biomedical Image Group, EPFL, 2011. URL: `http://bigwww.epfl.ch/sage/soft/localnormalization/`.

S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, K. Zuiderveld, Adaptive histogram equalization and its variations, Computer Vision, Graphics, and Image Processing 39 (1987) 355 – 368. doi:doi:10.1016/S0734-189X(87)80186-X.

D. Bulczak, M. Lambers, A. Kolb, Quantified, interactive simulation of AMCW ToF camera including multipath effects, Sensors 18 (2018) 13.

Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Processing 13 (2004) 600–612. doi:doi:10.1109/TIP.2003.819861.

K. Bredies, K. Kunisch, T. Pock, Total generalized variation, SIAM Journal on Imaging Sciences 3 (2010) 492–526. doi:doi:10.1137/090769521.

R. Ranftl, K. Bredies, T. Pock, Non-local total generalized variation for optical flow estimation, in: Computer Vision (ECCV), 2014, pp. 439–454.

Table 2: Results of the top-20 ranking approaches with respect to the multispectral dataset (**Own sRGB**, **Middlebury** and **MPI Sintel**) (top section), the top-20 ranking approaches with respect to the **ToF** data sets (middle section), and selected additional approaches (lower section) on our datasets; values are given as averaged ranks unless otherwise noted.

Concept, Scheme	**Image Transform.**	**Int. Corr.**	**Algo.**	**Tot. MS Avg.**	**Tot. ToF Avg.**	**Own sRGB** RGB	**Own sRGB** RGB-R	**SWIR**	**Middlebury** RGB	**Middlebury** RGB-R	**MPI Sintel** RGB	**MPI Sintel** RGB-R	**Time** [s]
Uncomp			-	296.59	161.08	261.97	285.52		251.78	283.2	303.64	318.19	-
NCM	-TG**		HL1	45.11	239.42	25.18	28.55	43.41	106.97	29.75	40.56	41.39	0.23
NCM	-I	N	Br	46.54	203.17	45.39	20.21	94.74	86.63	24.53	36.25	18.00	0.32
NCM	-TG**	C	HL1	46.64	248.67	31.12	25.82	36.81	129.00	38.95	39.75	25.06	0.24
NCM	-I	L	LDOF	47.67	151.33	44.76	23.21	45.19	67.83	23.00	64.81	64.92	9.68
NCM	-TG	L	LDOF	50.58	155.42	58.85	30.70	50.00	77.82	35.58	50.17	50.94	7.83
NCM	-TG	N	LDOF	50.84	136.33	95.73	36.18	24.30	71.17	24.87	61.92	41.72	5.29
NCM	-TG**	N	HL1	55.61	285.58	36.88	21.12	136.19	110.17	41.18	32.00	11.72	0.27
NCM	-I	N	LDOF	57.18	129.67	156.15	25.52	73.63	62.18	27.70	39.75	15.36	7.88
NCM	-(I+TG)**	C	HL1	57.78	264.42	78.21	33.61	34.67	142.03	37.23	46.78	31.92	0.16
NCM	-TG	N	LK	58.15	134.75	55.21	20.82	47.96	105.17	44.18	90.61	43.08	0.07
NCM	-TG	N	LK*	59.82	127.17	67.76	14.27	47.41	91.95	46.87	100.06	50.42	0.04
NCM	-TG+TG**		HL1	65.51	46.83	37.30	29.64	41.48	139.57	59.55	74.58	76.44	0.23
NCM	-I	L	LFN	66.80	175.25	109.67	88.39	25.59	50.82	98.53	32.56	62.03	7.65
NCM	-TG	N	HQS	69.23	114.83	44.67	51.85	65.74	123.03	44.73	109.28	45.28	0.18
NCM	-I	C	Br	69.68	188.92	20.55	73.73	58.81	94.00	142.82	18.39	79.47	0.32
NCM	-I	H	LDOF	70.21	225.50	124.42	40.76	176.56	73.87	27.78	34.83	13.28	7.02
NCM	-TG	H	LDOF	74.73	234.00	95.00	35.97	149.89	97.67	45.35	56.58	42.64	7.98
NCM	-TG	L	Br	74.97	134.17	74.48	37.33	59.07	111.23	83.02	87.25	72.39	0.48
NCM	-TG**	C	HL1*	75.27	224.25	54.61	44.55	83.07	148.95	78.47	65.42	51.83	0.05
NCM	-TG+TG**	N	HL1	75.50	78.25	52.61	45.48	80.37	141.53	90.15	69.94	48.39	0.27
CCM-B	-I		FHL*	118.11	22.25	100.39	139.45	52.00	69.28	109.83	173.03	182.78	0.03
CCM-B	-I		HL1*	118.98	23.25	100.88	140.03	54.19	70.28	111.20	173.06	183.25	0.04
CCM-B	-TG**		HL1	97.82	26.42	88.88	128.85	48.59	51.52	95.45	127.42	144.06	0.43
CCM-B	-TC**		TGVC	102.18	26.50	88.82	123.30	46.96	60.13	96.10	145.92	154.06	0.49
CCM-B	-I		H1C*	119.19	29.50	102.42	139.61	53.22	71.22	110.62	173.94	183.33	0.05
CCM-B	-I		HQS*	106.55	31.33	85.03	122.06	50.59	69.55	112.53	147.69	158.42	0.05
CCM-B	-TG**		HL1*	111.00	33.17	105.33	143.03	48.30	66.50	103.77	151.94	158.14	0.07
NCM	-TG+TG**	C	HL1	87.49	34.92	67.42	46.70	52.30	179.73	113.70	84.14	68.42	0.24
CCM-B	-(I+TG)**		HL1*	105.81	35.08	95.48	132.24	46.89	72.57	108.85	139.56	145.06	0.05
CCM-B	-TC**		TGVC*	139.29	36.00	140.79	168.48	67.63	97.97	126.10	190.50	183.56	0.08
CCM-B	-I		FHL	103.49	36.42	88.45	121.70	54.30	55.17	94.28	144.06	166.44	0.19
CCM-B	-I		HL1	103.49	36.83	88.85	123.52	49.89	55.95	94.15	145.47	166.64	0.23
NCM	-TG+TC**	C	TGVC*	118.22	37.00	77.94	49.48	119.59	155.17	154.13	139.42	131.78	0.07
CCM-B	-(I+TG)**		HL1	87.33	38.42	80.76	112.45	46.56	44.68	87.57	116.28	123.03	0.29
CCM-B	-I		TVL1	122.72	38.58	112.52	143.94	65.70	61.52	107.15	174.67	193.53	0.68
NCM	-TG+TG**	C	HL1*	110.30	39.92	80.82	56.97	71.70	195.67	145.85	122.19	98.89	0.06
NCM	-TG		H1C*	194.64	41.75	114.00	164.76	119.48	226.85	298.55	192.53	246.33	0.04
CCM-B	-I		H1C	104.54	46.50	88.82	122.55	50.22	58.07	97.87	148.28	166.00	0.30
NCM	-TG+TG**		HL1	65.51	46.83	37.30	29.64	41.48	139.57	59.55	74.58	76.44	0.23
NCM	-TG+TG**		HL1*	92.62	47.17	59.21	44.85	59.26	166.10	94.98	115.06	108.86	0.06
NCM	-I		BM	314.19	281.00	283.85	351.30	322.52	279.05	342.77	287.03	332.83	11.50
NCM	-TG		BM	244.12	270.25	218.12	233.18	259.44	217.22	272.50	246.06	262.31	11.62
CCM-B	-I		Br	84.74	49.00	70.15	97.24	56.52	49.38	77.60	111.83	130.44	0.66
CCM-U	-I		Br	131.95	118.50	114.97	152.61	85.96	77.17	127.20	170.50	195.25	0.32
NCM	-TG		Br	138.52	78.92	60.06	167.42	127.93	112.13	199.38	102.00	200.72	0.33
NCM	-I		TVL1	316.43	285.17	265.79	354.12	317.85	285.47	366.20	267.78	357.78	1.12
NCM	-I	H	FBM	183.24	277.25	189.91	156.36	255.37	209.50	124.17	193.50	153.86	0.20
NCM	-I		HL1	317.31	256.25	271.61	363.64	311.44	293.20	360.43	272.28	348.56	0.12
NCM	-TC**		TGVC	129.21	303.92	142.33	109.45	189.85	163.15	121.43	80.97	97.28	0.26
NCM	-I	N	LK	304.08	215.17	284.70	354.48	292.48	231.80	361.27	250.17	353.69	0.05
NCM	-C**		HL1	154.97	312.25	196.64	123.27	271.15	191.98	122.22	106.36	73.17	0.16
NCM	-I	H	HL1	210.25	346.00	237.79	137.36	316.67	275.88	192.97	210.06	101.00	0.14
CCM-U	-I		LDOF	114.82	112.83	135.30	161.21	48.22	55.87	104.55	131.33	167.25	8.50
NCM	-I		LDOF	174.20	135.33	118.91	282.45	142.85	61.90	290.15	51.47	271.67	5.61
NCM	-I		FF	346.42	346.42	307.55	326.67	273.44	184.35	261.63	266.42	307.03	53.54
NCM	-I	L	FF	291.13	332.00	289.82	280.85	314.78	182.32	168.47	264.39	251.25	53.01
CCM-B	-I		FN	106.84	121.00	120.42	123.85	39.70	50.12	79.85	110.44	124.33	30.46
NCM	-I	L	FN	154.72	191.00	178.09	149.94	56.37	111.50	139.95	87.69	105.56	15.40
NCM	-I	H	LFN	106.75	333.92	162.88	131.09	164.70	63.97	109.43	38.72	76.44	7.53
CCM-1	-I		LFN	200.87	89.67	222.55	262.21	100.30	145.48	209.13	213.42	253.00	2.52