Universität des Saarlandes

Fachrichtung 6.1 – Mathematik

Preprint Nr. 348

Morphologically Invariant Matching of Structures with the Complete Rank Transform

Oliver Demetz, David Hafner and Joachim Weickert

Saarbrücken 2014

Fachrichtung 6.1 – Mathematik
Universität des Saarlandes

Preprint No. 348
submitted: May 28, 2014

Morphologically Invariant Matching of Structures with the Complete Rank Transform

Oliver Demetz
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science
Saarland University, Campus E1.7, 66041 Saarbrücken, Germany
demetz@mia.uni-saarland.de

David Hafner
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science
Saarland University, Campus E1.7, 66041 Saarbrücken, Germany
hafner@mia.uni-saarland.de

Joachim Weickert
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science
Saarland University, Campus E1.7, 66041 Saarbrücken, Germany
weickert@mia.uni-saarland.de

Edited by
FR 6.1 – Mathematik
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
Germany

Fax:	+ 49 681 302 4443
e-Mail:	preprint@math.uni-sb.de
WWW:	http://www.math.uni-sb.de/

Abstract

Invariances are one of the key concepts to render computer vision algorithms robust against severe illumination changes. However, there is no free lunch: With any invariance comes an unavoidable loss of information. The goal of our paper is to introduce two novel descriptors which minimise this loss: the complete rank transform and the complete census transform. They are invariant under monotonically increasing intensity rescalings, while containing a *maximally possible* amount of information.

To analyse our descriptors, we embed them as constancy assumptions into a variational framework for optic flow computation. As a suitable regularisation term, we choose the total generalised variation that favours piecewise affine solutions. Our experiments focus on the KITTI benchmark where robustness w.r.t. illumination changes is one of the main issues. The results demonstrate that our descriptors yield state-of-the-art accuracy.

1 Introduction

Especially in uncontrolled real-world scenarios, robustness is one of the most important features of computer vision algorithms. Because it is hardly possible to design a model that can handle all eventualities explicitly, incorporating invariances is an essential alternative to gain robustness. However, being invariant means ignoring something, thus every invariance leads to a loss of information. This article aims at analysing a class of descriptors that exhibit an extraordinarily strong invariance, the so-called morphological invariance. Descriptors of this class remain unchanged if the underlying signal undergoes any transform that is monotonically increasing (Alvarez et al, 1993); e.g. additive, multiplicative and even exponential rescalings.
Since a monotonically increasing rescaling does not alter the ordering of the intensity values, the considered descriptor class comprises all approaches that are based on this grey value order. For instance, a famous example is the median filter of Tukey (1971). Zabih and Woodfill's rank transform (Zabih and Woodfill, 1994) is another prominent representative of such illumination robust descriptors. It computes the rank of a pixel's intensity within a local neighbourhood. Their transform is invariant against any monotonically increasing intensity changes. However, it is clear that only storing the rank of the pixel also means to discard all other local information. In the same paper (Zabih and Woodfill, 1994), the census transform is proposed, which compares a pixel with all its neighbours and stores which one is larger. In this way (besides encoding the rank in a different form), also some spatial

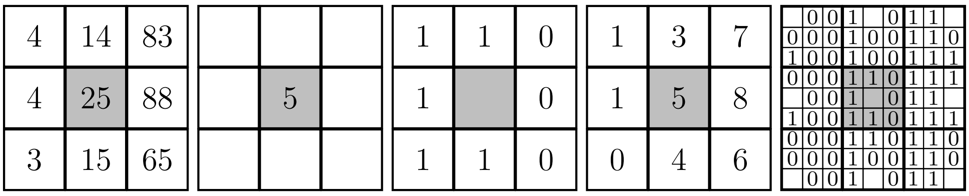
information is stored. However, also here a lot of information is discarded. Thus, it would be desirable to develop a robust feature that is morphologically invariant and discards as little information as possible.

Our Contributions. This article is an extended and revised version of our conference contribution (Demetz et al, 2013), where we introduced the complete rank transform as a tool for morphologically invariant matching of structures. We extend our conference paper in several aspects. First, we present further analysis and experiments w.r.t. the complete rank transform. Second, we derive an additional transform that formally contains the same amount of information. However, the novel so-called *complete census transform* is a binary signature that suggests a natural and very well-suited distance metric. By analysing the difference between this metric and the sum of squared distances that we propose for complete rank signatures, we can explain in which sense the latter is a good approximation of the former. Third, we improve our variational framework by extending the regularisation term to total generalised variation. This second order term favours piecewise affine solutions which appear frequently in realistic scenarios. We show that our proposed descriptors can be used as a generally superior alternative to the census transform: They are as parameter-free as the census transform, and lead to clearly improved results.
We want to stress that we discuss all these descriptors from the point of view of designing a data term for optical flow. Sparse interest point matching is not in our focus and would examine very different aspects and properties of a descriptor.

Related Work. There are many other transforms in literature that are related to our idea: independently of Zabih and Woodfill's rank and census transforms (Zabih and Woodfill, 1994), Pietikäinen et al. performed broad research on various kinds of *local binary patterns* (see the book on Local Binary Patterns of Pietikäinen et al (2011) and references therein). However the majority of these local binary patterns discards rather more than less available information (Chen et al, 2013), hence go in the opposite direction of our research. Stein (2004) use the census transform as an efficient descriptor for sparse structure matching in driver assistance systems, and Fröba and Ernst (2004) use the modified census transform for face recognition. The BRIEF descriptor of Calonder et al (2012) is a variation of the census transform which performs the comparisons on arbitrary pixel pairs in the neighbourhood. The first appearance of ordinal measures of full patches in the literature goes back to work on block matching based stereo correspondence of Bhat and Nayar (1998). Related to that, also more recently, several sparse interest point descriptors building on intensity order-based ideas have been proposed: With their chained circular neighbourhoods, Chan et al (2012) make a first step towards representing neighbourhood ordinal information. The LIOP descriptor of Wang et al (2011) describes the intensity order of a very large neighbourhood and is specifically tailored for sparse interest point matching. A similar idea of matching order distributions is proposed by Tang et al (2009). Mittal and Ramesh (2006) combine order and intensity information to increase the robustness against Gaussian noise.

A classical application domain where local descriptors are matched is optical flow. A large number of publications on this topic also consider the problem of illumination robustness. Most of these attempts are based on invariance, such as the gradient constancy assumption introduced by Uras et al (1988) that is invariant under global additive changes, as well as constancy assumptions on higher order derivatives by Papenberg et al (2006). A higher level of invariance can be achieved with the normalised cross correlation (Steinbrücker et al, 2009; Werlberger et al, 2010) which is also invariant under multiplicative changes. The work of Liu et al (2011) also falls in this class of invariance, where the SIFT descriptor (Lowe, 2004) is used for establishing correspondence. Other attempts to achieve invariance include the structure-texture decomposition by Wedel et al (2008) as well as the Histogram of Oriented Gradients-based method by Rashwan et al (2013). In presence of color imagery, invariance can also be achieved by exploiting the dichromatic reflection model (van de Weijer and Gevers, 2004) or by switching to other color spaces as performed in Mileva et al (2007). A remarkable exception from the invariance-based approaches to achieve robustness is the work of Gennert and Negahdaripour (1987) where deviations from the brightness constancy assumption are estimated explicitly. In Xu et al (2010) and Kim et al (2013) invariant data terms are incorporated in an adaptive way by switching locally between different constancy assumptions.

There are several recent publications that incorporate the census transform in variational optical flow or stereo methods: Müller et al (2011) propose a census-based data term for optical flow, and Ranftl et al (2012) as well as Mei et al (2011) present census-based stereo methods. Braux-Zin et al (2013) combine census and grey value constancy assumption in a data term for optic flow and additionally integrate sparse feature matches. The theoretical study of Hafner et al (2013) explains the reasons why census-based data terms for variational optic flow are successful.



(a) Intensities. (b) Rank. (c) Census. (d) Complete rank. (e) Complete census.

Figure 1: Illustration of the presented intensity order transforms (*(b)–(d)*) with a 3 × 3 neighbourhood patch (*(a)*), where the reference pixel is marked in grey.

Organisation. Our paper is organised as follows: In Section 2, we discuss the rank and census transforms, as well as their complete counterparts. After that, Section 3 discusses appropriate measures of patch dissimilarity for each of the transforms. In Section 4, we embed our novel descriptors into a variational framework and demonstrate their benefits in the experimental Section 5. We conclude the paper with a summary and an outlook in Section 6.

2 Morphologically Invariant Descriptors

Let us now give an overview over the class of morphologically invariant transformations (cf. Figure 1), i.e. transforms that are invariant under any global monotonically increasing rescaling of the input signal.
Formally, each transform maps a local image patch to a m-dimensional signature vector $\boldsymbol{s} : \mathbb{R}^k \to \mathbb{R}^m$. In this paper, we define the image patch as the k closest neighbouring pixels w.r.t. the spatial Euclidean distance. For didactic reasons, we represent the patch intensity values by a k-dimensional vector $\boldsymbol{f}$, where the values are ordered by increasing spatial distance from the centre. Consequently, the intensity of the central pixel is assigned to the first entry f_1.

2.1 Rank

The *rank transform* (RT) was proposed by Zabih and Woodfill (1994) and encodes for each pixel the position of its grey value in the ranking of all grey values in the neighbourhood. In other words, it is the number of neighbours with a smaller grey value than the central one. Formally, the rank transform maps each pixel to its scalar rank signature $s_{\mathrm{RT}} \in \{0, \ldots, k-1\}$, and can be computed as

$$s_{\mathrm{RT}}(\boldsymbol{f}) := \sum_{i=2}^{k} \mathbb{1}_{(f_i < f_1)} \,, \tag{1}$$

where $\mathbb{1}_{(x)}$ denotes the indicator function

$$\mathbb{1}_{(x)} := \begin{cases} 1 & \text{if } x \text{ is true,} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

2.2 Census

In the same paper, Zabih and Woodfill (1994) also introduced another descriptor, the so-called *census transform* (CT). It has attracted a lot of attention in recent years and can be seen as an extension of the rank transform: Besides encoding the rank, it adds a spatial component by expressing the relationship between the central pixel and each of its neighbours explicitly. Specifically, one bit of information is stored for each pixel of the neighbourhood: If the neighbour is smaller than the reference pixel the bit is 1, and 0 otherwise. In the final binary signature, all bits are concatenated. While the order of this concatenation is in general arbitrary, it has to be consistent such that each bit can be uniquely associated with one neighbour. In mathematical terms, each image patch of size k is mapped to a binary signature $\boldsymbol{s}_{\mathrm{CT}} \in \{0,1\}^{k-1}$ of length k–1. We choose the following formal representation to compute a census signature:

$$\boldsymbol{s}_{\mathrm{CT}}(\boldsymbol{f}) := \left(\mathbb{1}_{(f_2 < f_1)}, \ldots, \mathbb{1}_{(f_k < f_1)}\right)^{\top} . \tag{3}$$

Hence, every neighbouring pixel is compared to the central one. Furthermore, the sum of the digits of a census signature coincides with the rank s_{RT} of that pixel.

2.3 Complete Rank

Although the two signatures by Zabih and Woodfill (1994) exhibit the same morphological invariance, the census transform obviously encodes by construction more information than the pure rank.

However, there is still some more information that can be used without losing the desired invariance. To this end, let us now introduce an extension of Zabih and Woodfill's basic transform: the *complete rank transform* (CRT). We will see that the resulting signature carries much more information than its predecessors.

Given the census signature of an image patch, we know which pixels in the patch are smaller than the *central* one. However, the relationships among *all* neighbours cannot be determined by the pure census information. For instance, if two neighbouring pixels are both smaller than the central one, it is still unclear which of the two neighbours is smallest.

To also encode this information, we propose the complete rank transform. We compute the rank of each pixel of the patch and store this information in a k-dimensional signature $\boldsymbol{s}_{\mathrm{CRT}} \in \{0,\dots,k-1\}^k$:

$$s_{\mathrm{CRT}}(\boldsymbol{f}) := (s^1_{\mathrm{RT}},\dots,s^k_{\mathrm{RT}})^\top , \tag{4}$$

where

$$s^j_{\mathrm{RT}} := \sum_{\substack{i=1\\ i\neq j}}^{k} \mathbb{1}_{(f_i<f_j)} \,. \tag{5}$$

With this CRT signature, the whole intensity order is represented. From the viewpoint of morphological invariance, this is the maximal amount of information that can be extracted without leaving this class of invariance.

The computation rule for CRT signatures as shown in Equation 4 is demonstrative and intuitively understandable, but also inefficient (quadratic complexity in k). However, essentially what has to be done is to sort the intensities. Thus, we propose to use an efficient sorting algorithm such as *Quicksort* for this task ($\mathcal{O}(k\log k)$); see e.g. Press et al (2007).

2.4 Complete Census

After motivating the complete rank transform via the missing relationship information between all pixels in the patch, another transform comes naturally into mind, namely an analogue extension of the census transform: the *complete census transform* (CCT).

Instead of storing all k ranks, it stores for each pixel of the patch whether it is smaller or larger than *any other* pixel in the patch. Thus, we obtain a signature $\boldsymbol{s}_{\mathrm{CCT}} \in \{0,1\}^{k\cdot(k-1)}$ which contains all census signatures with each of the pixels as reference:

$$\boldsymbol{s}_{\mathrm{CCT}}(\boldsymbol{f}) := (\boldsymbol{s}^1_{\mathrm{CT}},\dots,\boldsymbol{s}^k_{\mathrm{CT}})^\top , \tag{6}$$

with

$$\boldsymbol{s}^j_{\mathrm{CT}} := \left(\mathbb{1}_{(f_1<f_j)},\dots,\mathbb{1}_{(f_{j-1}<f_j)},\mathbb{1}_{(f_{j+1}<f_j)},\dots,\mathbb{1}_{(f_k<f_j)}\right) . \tag{7}$$

Evidently, the original census signature from Equation 3 coincides with ${\boldsymbol{s}^1_{\mathrm{CT}}}^\top$. The information contained in complete rank and complete census is equivalent. This can be seen from the bijection between them: It makes no difference if we compute the CCT signature directly from the intensity values or from the CRT signature of a patch. In the opposite direction, the complete rank digits are just the sums of corresponding CCT bits:

$$(s_{\mathrm{CRT}}(\boldsymbol{s}_{\mathrm{CCT}}))_j = \sum_{i=1}^{k-1} (\boldsymbol{s}^j_{\mathrm{CT}})_i \,. \tag{8}$$

Both complete rank and complete census signatures, do also represent tied ranks, i.e. if pixels in the patch have the same intensity. Thus, the number of possible signatures for a patch with k pixels is the k-th *ordered Bell number* (Sloane and Plouffe, 1995) OBN(k) (also called k-th Fubini number), which is defined by

$$\mathrm{OBN}(k) \;=\; \sum_{i=0}^{k} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} j^k \,. \tag{9}$$

It expresses the maximally possible number of weak orderings of a set of k elements.

2.5 Discussion

In each pixel, our complete rank signature contains the full local image intensity order. Obviously this is much more information than the rank or census signatures carry. In particular, it is impossible to encode more local image information without leaving the class of morphologically invariant descriptors. The reason for this is that the only property that cannot be changed by a monotonic function is monotonicity, i.e. whether one pixel is larger than the other or not. However, the reason to prefer our proposed complete rank signature is its much more compact representation and lower dimensionality, compared to the complete census signature.

Nevertheless, this alternative census-inspired perspective offers an unexpected insight: As pointed out in (Hafner et al, 2013), each binary digit of a census signature can be regarded as the sign of the corresponding directional derivative (in a finite difference sense). Thus, from this point of view, one can conclude that the complete rank transform inherently contains rich local differential information. In this regard, dealing with derivatives of such signatures as in (Puxbaum and Ambrosch, 2010) actually corresponds to second order image derivative information. This fact is not obvious from just considering the rank representation and should be kept in mind.

For the sake of clarity, we summarise the discussed transforms and compare their essential properties in Table 1.

Table 1: Comparison of the proposed intensity order transforms. The number of pixels in the considered neighbourhood is given by k.

transform	range $\mathcal{D}$ of one digit	signature length m	spatial information	size of descriptor space
rank (RT)	$\{0, \ldots, k-1\}$	1	–	k
census (CT)	$\{0, 1\}$	$k-1$	✓	2^{k-1}
complete rank (CRT)	$\{0, \ldots, k-1\}$	k	✓	OBN(k)
complete census (CCT)	$\{0, 1\}$	$k(k-1)$	✓	OBN(k)

3 Signature Distance Metrics

Besides the question which signature to chose, an equally important decision to take is the metric in which to compare the chosen signatures.
For the classical rank and census transform the answer is clear: For ranks, the absolute value of their differences is an appropriate metric because smaller rank difference means more similar patches. Let, similar to Section 2, $\boldsymbol{f}$ and $\boldsymbol{g}$ denote two patches to compare. Then, the corresponding metric for rank reads

$$d(s_{\mathrm{RT}}(\boldsymbol{f}), s_{\mathrm{RT}}(\boldsymbol{g})) = |s_{\mathrm{RT}}(\boldsymbol{f}) - s_{\mathrm{RT}}(\boldsymbol{g})| \,. \tag{10}$$

In case of census signatures, their Hamming distance is a natural choice since it reflects the number of pixel comparisons that are in agreement:

$$d(\boldsymbol{s}_{\mathrm{CT}}(\boldsymbol{f}), \boldsymbol{s}_{\mathrm{CT}}(\boldsymbol{g})) = \sum_{i=1}^{k-1} \mathbb{1}_{((s_{\mathrm{CT}}(\boldsymbol{f}))_i = (s_{\mathrm{CT}}(\boldsymbol{g}))_i)} \,. \tag{11}$$

In the context of ternary census signatures, Vogel et al (2013) propose the *Centralised Sum of Absolute Distances* (CSAD) as a convex approximation. However, this approximation looses many invariances, in fact even the invariance under multiplicative rescalings is lost. Thus, for us this is no option.

The straightforward generalisation of the absolute rank difference to its complete counterpart would be the Euclidean norm of the difference vector ($p = 2$) or the sum of absolute component differences ($p = 1$):

$$d(\boldsymbol{s}_{\mathrm{CRT}}(\boldsymbol{f}), \boldsymbol{s}_{\mathrm{CRT}}(\boldsymbol{g})) = \left(\sum_{j=1}^{k} |(s_{\mathrm{CRT}}(\boldsymbol{f}))_j - (s_{\mathrm{CRT}}(\boldsymbol{g}))_j|^p \right)^{1/p} \,. \tag{12}$$

However, one is actually interested in the number of pixel comparisons in the patch not being in agreement. In this regard, the desired dissimilarity measure can be obtained by applying the Hamming distance to the complete census signatures:

$$d(\boldsymbol{s}_{\mathrm{CCT}}(\boldsymbol{f}), \boldsymbol{s}_{\mathrm{CCT}}(\boldsymbol{g})) = \sum_{i=1}^{k(k-1)} \mathbb{1}_{((s_{\mathrm{CCT}}(\boldsymbol{f}))_i = (s_{\mathrm{CCT}}(\boldsymbol{g}))_i)} \, . \tag{13}$$

Interestingly, the two latter metrics exhibit a close relationship that becomes clear by plugging Equation 8 into the former one:

$$d(\boldsymbol{s}_{\mathrm{CRT}}(\boldsymbol{f}), \boldsymbol{s}_{\mathrm{CRT}}(\boldsymbol{g})) = \left(\sum_{j=1}^{k} \left| \sum_{i=1}^{k-1} (\boldsymbol{s}^{j}_{\mathrm{CT}}(\boldsymbol{f}))_i - \sum_{i=1}^{k-1} (\boldsymbol{s}^{j}_{\mathrm{CT}}(\boldsymbol{g}))_i \right|^{p} \right)^{1/p} . \tag{14}$$

Comparing Equation 13 and 14, one can see that instead of counting each individual disagreeing pixel comparison, the disagreements are accumulated for each of the k CRT components. This accumulation performs a best case estimate, i.e. as many census digits as possible are assumed to coincide. In other words, the best possible case for each CRT component is assumed.

In the case of the sum of absolute differences ($p = 1$), the metric exactly represents the lowest possible bound on the Hamming CCT distance. For the Euclidean distance ($p = 2$), the individual rank differences are amplified by the square function. Thus, more disagreeing pixel comparisons than the least possible are presumed.

To conclude, CCT in combination with the Hamming distance might be the most intuitive measure. However, CCT introduces a large computational overhead compared to CRT; signature length $k(k-1)$ versus k. Nevertheless, we have seen that the CRT metrics approximate this CCT metric in a meaningful way. It is up to our experiments to show the difference of both signature-metric combinations in terms of accuracy.

4 Variational Optical Flow Model

In this section, we demonstrate the suitability our morphologically invariant features for optic flow computation. To this end, we embed the features into a variational framework that is based on the seminal work of Horn and Schunck (1981). It allows a transparent and flexible modelling while being able to provide accurate state-of-the-art results as demonstrated in various optic flow benchmarks.

4.1 Energy Formulation

Generally, we assume that the input images have been mapped by one of the introduced transforms to a vector-valued function $\boldsymbol{s} : \Omega \times [0, \infty) \to \mathcal{D}^m$.

Here, $\Omega \subset \mathbb{R}^2$ denotes the 2D rectangular image domain. For colour images, typically the signature length is tripled since we concatenate the signatures of each channel.

Our generic variational approach models the constancy assumption that signatures of corresponding pixels in the first frame and in the second frame coincide. Accordingly, we propose to compute the sought optic flow field $(u, v)^\top \colon \Omega \to \mathbb{R}^2$ as the minimiser of the following energy functional:

$$E = \int_\Omega \big(M + \alpha \cdot R_1 + \beta \cdot R_2\big)\,\mathrm{d}x\,\mathrm{d}y\,, \tag{15}$$

where the data term

$$M = \Psi\big(\tfrac{1}{m}|\boldsymbol{s}(\boldsymbol{x}+\boldsymbol{w}) - \boldsymbol{s}(\boldsymbol{x})|^2\big) \tag{16}$$

implements the discussed signature constancy assumption. It penalises differences between the signature at position $\boldsymbol{x} := (x, y, t)^\top$ in the first frame and its corresponding one at $\boldsymbol{x}+\boldsymbol{w} := (x+u, y+v, t+1)^\top$ in the second frame. We apply the penalisation function (Cohen, 1993; Schnörr, 1994)

$$\Psi(z^2) = 2\lambda\sqrt{z^2+\lambda^2} - 2\lambda^2 \tag{17}$$

with the small positive parameter λ to handle outliers and occlusions.

The regularisation terms

$$R_1 = \Psi\big(|\boldsymbol{\nabla} u - \boldsymbol{a}|^2 + |\boldsymbol{\nabla} v - \boldsymbol{b}|^2\big) \tag{18}$$

and

$$R_2 = \Psi\big(|\boldsymbol{\mathcal{J}}\boldsymbol{a}|^2 + |\boldsymbol{\mathcal{J}}\boldsymbol{b}|^2\big) \tag{19}$$

model a prior knowledge about the spatial smoothness of the flow field. Here, $\boldsymbol{\nabla} := (\partial_x, \partial_y)^\top$ denotes the spatial gradient operator and $\boldsymbol{\mathcal{J}}\boldsymbol{a}$ and $\boldsymbol{\mathcal{J}}\boldsymbol{b}$ the Jacobian matrices of the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. Further, all terms are balanced by the regularisation parameters $\alpha > 0$ and $\beta > 0$.

Let us discuss the applied regularisation terms in more detail. Setting $\boldsymbol{a}$ and $\boldsymbol{b}$ to $\boldsymbol{0}$ everywhere yields a first order smoothness penalty, that approximates the penalisation of the total variation (TV) of the flow field (Brox et al, 2004). However, jointly optimising for $\boldsymbol{a}$ and $\boldsymbol{b}$ in combination with the optic flow approximates a second order smoothness assumption. Here, R_1 plays the role of a coupling term: It enforces the flow derivatives $\boldsymbol{\nabla} u$ and $\boldsymbol{\nabla} v$ to equal $\boldsymbol{a}$ and $\boldsymbol{b}$. Thus, the first order smoothness term R_2 on $\boldsymbol{a}$ and $\boldsymbol{b}$ features a second order smoothness assumption on u and v. The resulting regularisation can be seen as a continuous and differentiable version of the second order total generalised variation (TGV^2) smoothness term of Bredies et al (2010). In a linear contest, such an approximation has been considered by Hewer et al (2013). In contrast to the first order smoothness term that leads to piecewise constant flow fields, it favours piecewise affine solutions.

4.2 Multiresolution Technique

Due to the data term (16), the energy in (15) is non-convex w.r.t. the optic flow variables u and v. As a remedy we consider, similar to Brox et al (2004), a fixed point iteration that is embedded it into a multiresolution pyramid-based approach. This leads to a linearised and convex version of the data term that allows to handle large displacements. On each pyramid level ℓ, we solely compute small flow increments du^ℓ and dv^ℓ. The corresponding linearised version of data term (16) reads

$$M^\ell = \Psi\left(\tfrac{1}{m}|\boldsymbol{s}_x \cdot du^\ell + \boldsymbol{s}_y \cdot dv^\ell + \boldsymbol{s}_t|^2\right), \tag{20}$$

where the derivatives of the vector-valued signature $\boldsymbol{s}$, i.e.

$$\begin{aligned} \boldsymbol{s}_x &:= \partial_x\, \boldsymbol{s}(\boldsymbol{x} + \boldsymbol{w}^\ell), && (21)\\ \boldsymbol{s}_y &:= \partial_y\, \boldsymbol{s}(\boldsymbol{x} + \boldsymbol{w}^\ell), && (22)\\ \boldsymbol{s}_t &:= \boldsymbol{s}(\boldsymbol{x} + \boldsymbol{w}^\ell) - \boldsymbol{s}(\boldsymbol{x}) && (23) \end{aligned}$$

are calculated componentwise.
The computed flow increments du^ℓ and dv^ℓ are then successively used to update the overall flow:

$$u^{\ell+1} = u^\ell + du^\ell \qquad \text{and} \qquad v^{\ell+1} = v^\ell + dv^\ell\,. \tag{24}$$

Please note that the coupling term R_1 couples the derivative of the *complete* flow to the auxiliary variables in each level, i.e.

$$R_1^\ell = \Psi\left(|\boldsymbol{\nabla}(u^\ell + du^\ell) - \boldsymbol{a}|^2 + |\boldsymbol{\nabla}(v^\ell + dv^\ell) - \boldsymbol{b}|^2\right). \tag{25}$$

The regularisation term R_2 stays unaffected by this multiresolution strategy. In summary, the following incremental energy is to be minimised w.r.t. du^ℓ and dv^ℓ (in the case of first order smoothness), and additionally w.r.t. $\boldsymbol{a}$ and $\boldsymbol{b}$ (second order smoothness):

$$E^\ell = \int_\Omega \left(M^\ell + \alpha \cdot R_1^\ell + \beta \cdot R_2\right) \mathrm{d}x\, \mathrm{d}y\,. \tag{26}$$

4.3 Minimisation

Following the calculus of variations (Gelfand and Fomin, 2000), the minimiser of the energy (26) has to fulfil the Euler-Lagrange equations. For the sake of readability, let us first introduce the following abbreviations:

$$\Psi'_M := \tfrac{1}{m} \cdot \Psi'\left(\tfrac{1}{m}|\boldsymbol{s}_x \cdot du^\ell + \boldsymbol{s}_y \cdot dv^\ell + \boldsymbol{s}_t|^2\right) , \tag{27}$$

$$\Psi'_{R_1} := \Psi'\big(|\boldsymbol{\nabla}(u^\ell + du^\ell) - \boldsymbol{a}|^2 + |\boldsymbol{\nabla}(v^\ell + dv^\ell) - \boldsymbol{b}|^2\big) , \tag{28}$$

$$\Psi'_{R_2} := \Psi'\big(|\boldsymbol{\mathcal{J}}\boldsymbol{a}|^2 + |\boldsymbol{\mathcal{J}}\boldsymbol{b}|^2\big) . \tag{29}$$

With these abbreviations, the Euler-Lagrange equations for the optic flow increments du^ℓ and dv^ℓ can be formulated as

$$\begin{aligned}\Psi'_M \cdot \boldsymbol{s}_x^\top (\boldsymbol{s}_x \cdot du^\ell + \boldsymbol{s}_y \cdot dv^\ell + \boldsymbol{s}_t)& \\ - \alpha \cdot \operatorname{div}\big(\Psi'_{R_1} \cdot \big(\boldsymbol{\nabla}(u^\ell + du^\ell) - \boldsymbol{a}\big)\big) &= 0 ,\end{aligned} \tag{30}$$

$$\begin{aligned}\Psi'_M \cdot \boldsymbol{s}_y^\top (\boldsymbol{s}_x \cdot du^\ell + \boldsymbol{s}_y \cdot dv^\ell + \boldsymbol{s}_t)& \\ - \alpha \cdot \operatorname{div}\big(\Psi'_{R_1} \cdot \big(\boldsymbol{\nabla}(v^\ell + dv^\ell) - \boldsymbol{b}\big)\big) &= 0 ,\end{aligned} \tag{31}$$

with the natural boundary conditions

$$(\boldsymbol{\nabla}(u^\ell + du^\ell) - \boldsymbol{a})^\top \boldsymbol{n} = 0 , \tag{32}$$

$$(\boldsymbol{\nabla}(v^\ell + dv^\ell) - \boldsymbol{b})^\top \boldsymbol{n} = 0 , \tag{33}$$

where $\boldsymbol{n}$ is the outer normal vector to the boundary of Ω.

Applying a second order smoothness assumption, one additionally has to solve for the coupling variables $\boldsymbol{a}$ and $\boldsymbol{b}$:

$$\alpha \cdot \Psi'_{R_1} \cdot (\boldsymbol{\nabla}(u^\ell + du^\ell) - \boldsymbol{a}) - \beta \cdot \boldsymbol{\nabla}^\top \left(\Psi'_{R_2} \cdot (\boldsymbol{\mathcal{J}}\boldsymbol{a})^\top\right) = \boldsymbol{0} , \tag{34}$$

$$\alpha \cdot \Psi'_{R_1} \cdot (\boldsymbol{\nabla}(v^\ell + dv^\ell) - \boldsymbol{b}) - \beta \cdot \boldsymbol{\nabla}^\top \left(\Psi'_{R_2} \cdot (\boldsymbol{\mathcal{J}}\boldsymbol{b})^\top\right) = \boldsymbol{0} , \tag{35}$$

with $(\boldsymbol{\mathcal{J}}\boldsymbol{a})\boldsymbol{n} = \boldsymbol{0}$ and $(\boldsymbol{\mathcal{J}}\boldsymbol{b})\boldsymbol{n} = \boldsymbol{0}$ as boundary conditions.

4.4 Numerical Algorithm and Implementation

We assume the images to be sampled on a regular grid with horizontal and vertical grid size h_1 and h_2. All occurring spatial derivatives of the signatures $\boldsymbol{s}$ in (30) and (31) are computed by means of the 4th-order stencil $(1, -8, 0, 8, -1)/(24h_d)$, $d = 1, 2$, while the temporal derivative $\boldsymbol{s}_t$ is determined by a simple forward difference.

Moreover, in contrast to standard pyramid-based approaches, we *do not* downsample the input images themselves, but their transformed versions. Otherwise, the desired morphological invariance would be lost when averaging or interpolating raw intensity values in the downsampling strategy. To compute the downsampled transformed images, we presmooth them with a

Gaussian whose standard deviation is proportional to the current grid size $\eta^\ell \cdot h_d$, where $\eta = 0.95$ is the downsampling factor.
On each pyramid level, we have to solve a sparse nonlinear system of equations, where the nonlinearities are caused by the Ψ' terms (27), (28), and (29). To solve these systems, we apply the lagged nonlinearity method (Vogel and Oman, 1996) which basically consists of two nested loops: In the inner loops, we keep the nonlinearity terms fixed and thus, only have to solve a sparse linear system of equations. The nonlinearity terms are then subsequently updated in the outer loop. As a fast and easy implementable linear system solver, we use the *Fast Jacobi* method of Grewenig et al (2013). It is perfectly suited for an implementation on parallel hardware architectures such as modern GPUs. Basically, it is based on a standard Jacobi solver. However, varying cyclic under- and over-relaxations ω where even half of them may violate the stability limit allow an enormous speed-up. More precisely, one iteration step at the pyramid level ℓ with pixel index i and iteration index k is for the flow increments du^ℓ and dv^ℓ given by

$$\begin{aligned} du_i^{\ell,k+1} = (1-\omega)\cdot du_i^{\ell,k} \;+\; \omega \;\cdot\; \Bigg(& -\Psi'_{Mi} \cdot \boldsymbol{s}_{xi}^\top(\boldsymbol{s}_{yi}\cdot dv_i^{\ell,k} + \boldsymbol{s}_{ti}) \\ & + \sum_{d=1}^{2} \sum_{j\in\mathcal{N}_d(i)} \alpha \cdot \frac{\Psi'_{R_1 i} + \Psi'_{R_1 j}}{2h_d} \cdot \Big(\frac{u_j^{\ell,k} - u_i^{\ell,k} + du_j^{\ell,k}}{h_d} + a_{di}^{\ell,k} - a_{dj}^{\ell,k} \Big) \Bigg) \\ & \Bigg/ \Bigg(\Psi'_{Mi} \cdot \boldsymbol{s}_{xi}^\top \boldsymbol{s}_{xi} + \sum_{d=1}^{2} \sum_{j\in\mathcal{N}_d(i)} \alpha \cdot \frac{\Psi'_{R_1 i} + \Psi'_{R_1 j}}{2h_d^2} \Bigg), \end{aligned} \tag{36}$$

$$\begin{aligned} dv_i^{\ell,k+1} = (1-\omega)\cdot dv_i^{\ell,k} \;+\; \omega \;\cdot\; \Bigg(& -\Psi'_{Mi} \cdot \boldsymbol{s}_{yi}^\top(\boldsymbol{s}_{xi}\cdot du_i^{\ell,k} + \boldsymbol{s}_{ti}) \\ & + \sum_{d=1}^{2} \sum_{j\in\mathcal{N}_d(i)} \alpha \cdot \frac{\Psi'_{R_1 i} + \Psi'_{R_1 j}}{2h_d} \cdot \Big(\frac{v_j^{\ell,k} - v_i^{\ell,k} + dv_j^{\ell,k}}{h_d} + b_{di}^{\ell,k} - b_{dj}^{\ell,k} \Big) \Bigg) \\ & \Bigg/ \Bigg(\Psi'_{Mi} \cdot \boldsymbol{s}_{yi}^\top \boldsymbol{s}_{yi} + \sum_{d=1}^{2} \sum_{j\in\mathcal{N}_d(i)} \alpha \cdot \frac{\Psi'_{R_1 i} + \Psi'_{R_1 j}}{2h_d^2} \Bigg), \end{aligned} \tag{37}$$

where $\mathcal{N}_1$ and $\mathcal{N}_2$ describe the neighbouring pixels in horizontal and vertical direction, respectively. In an analogous way, the iteration step for

$\boldsymbol{a}=(a_1,a_2)^\top$ and $\boldsymbol{b}=(b_1,b_2)^\top$ reads for $p=1,2$

$$\begin{aligned} a_{pi}^{k+1} = (1-\omega)\cdot a_{pi}^{k} + \omega\ \cdot\ & \left(\frac{\alpha \cdot {\Psi'}_{R_1 i}}{2h_p} \cdot \left(u_{n_p^+}^{\ell,k} - u_{n_p^-}^{\ell,k} + du_{n_p^+}^{\ell,k} - du_{n_p^-}^{\ell,k}\right) \right. \\ & \left. + \sum_{d=1}^{2} \sum_{j\in\mathcal{N}_d(i)} \beta \cdot \frac{{\Psi'}_{R_2 i} + {\Psi'}_{R_2 j}}{2h_d^2} \cdot a_{pj}^{k} \right) \\ & \Bigg/ \left(\alpha \cdot {\Psi'}_{R_1 i} + \sum_{d=1}^{2} \sum_{j\in\mathcal{N}_d(i)} \beta \cdot \frac{{\Psi'}_{R_2 i} + {\Psi'}_{R_2 j}}{2h_d^2} \right), \end{aligned} \tag{38}$$

$$\begin{aligned} b_{pi}^{k+1} = (1-\omega)\cdot b_{pi}^{k} + \omega\ \cdot\ & \left(\frac{\alpha \cdot {\Psi'}_{R_1 i}}{2h_p} \cdot \left(v_{n_p^+}^{\ell,k} - v_{n_p^-}^{\ell,k} + dv_{n_p^+}^{\ell,k} - dv_{n_p^-}^{\ell,k}\right) \right. \\ & \left. + \sum_{d=1}^{2} \sum_{j\in\mathcal{N}_d(i)} \beta \cdot \frac{{\Psi'}_{R_2 i} + {\Psi'}_{R_2 j}}{2h_d^2} \cdot b_{pj}^{k} \right) \\ & \Bigg/ \left(\alpha \cdot {\Psi'}_{R_1 i} + \sum_{d=1}^{2} \sum_{j\in\mathcal{N}_d(i)} \beta \cdot \frac{{\Psi'}_{R_2 i} + {\Psi'}_{R_2 j}}{2h_d^2} \right), \end{aligned} \tag{39}$$

where n_1^- and n_1^+ describe the left and right neighbouring pixels in horizontal direction. In a similar way, the vertical neighbours are denoted by n_2^- and n_2^+.
Our reference implementation runs on an NVidia Geforce GTX 460 graphics card and is written in CUDA. On this platform, the typical computation time of our method on image sequences of size 640 × 480 is 13 seconds per flow field for the CRT descriptor with first order smoothness.

5 Experiments

The experimental evaluation of our method is structured as follows: First, we demonstrate that the invariance against any monotonically increasing rescaling is indeed fulfilled by simulating such remappings artificially. Similarly, we analyse the behaviour under noise with additive white Gaussian noise. In the second part of our experiments, we focus on the KITTI vision benchmark suite (Geiger et al, 2012), and assess the performance of our method with this real-world data set.

Choice of Parameters. Due to its simplicity, only very few parameters have to be chosen. The main free parameters of our optical flow method is the weight of the coupling term α as well as the smoothness weight β for the regularisation term on the auxiliary variables $\boldsymbol{a}$ and $\boldsymbol{b}$. As it turns out, the coupling weight has the largest influence on the results and was chosen in the range $[10^{-1}, 10^{-3}]$. The smoothness weight β, however, only has an indirect influence on the flow. We have chosen it fixed to $\beta = 3$ throughout our experiments. The last free parameters of our framework are the contrast parameters of the subquadratic functions. Also here, no adaptation per image sequence is necessary, and $\lambda = 10^{-2}$ for the data and smoothness term as well as $\lambda = 0.5$ for the coupling term work well.

We repeat the incremental flow computations on each level four times. Our Fast Jacobi-based numerical scheme performs 5 outer and 20 inner interactions per incremental computation, and the upsampling of the flow as well as the coupling variables is performed with bilinear interpolation. The back-registration at each level is performed with bicubic interpolation.

5.1 Behaviour under Synthetic Perturbations

Concerning synthetic perturbations, we consider the eight training image sequences of the Middlebury benchmark (Baker et al, 2011) because no severe illumination changes are present and reliable ground truth flow fields are available. To assess the accuracy of an estimated flow field, we evaluate the *average endpoint error* (Otte and Nagel, 1994):

$$\mathrm{EPE}(\boldsymbol{w}, \boldsymbol{w}_{\mathrm{gt}}) = \frac{1}{\Omega} \int_{\Omega} |\boldsymbol{w}(\boldsymbol{x}) - \boldsymbol{w}_{\mathrm{gt}}(\boldsymbol{x})| \, \mathrm{d}\boldsymbol{x}\,, \tag{40}$$

where $\boldsymbol{w} = (u, v)^{\top}$ is the estimated and $\boldsymbol{w}_{\mathrm{gt}} = (u_{\mathrm{gt}}, v_{\mathrm{gt}})^{\top}$ is the known ground truth displacement field.

Invariance to γ Changes. Our first experiment examines the behaviour of the proposed method under monotonically increasing intensity changes. To this end, we consider the eight Middlebury training image sequences and apply a γ-correction to the second frame:

$$f_{\gamma}(\boldsymbol{x}) := 255 \cdot \left(\tfrac{1}{255} f(\boldsymbol{x})\right)^{\gamma} . \tag{41}$$

The results of this experiment are depicted in Figure 2. In practice, such a gamma correction is performed with floating point accuracy. However, to simulate the image acquisition process in a digital camera more realistically, the subsequent quantisation step must be taken into account. The problem with this nonlinear post-processing step is that it can alter the intensity order. We simulate the quantisation at two different bit depths: Most often,

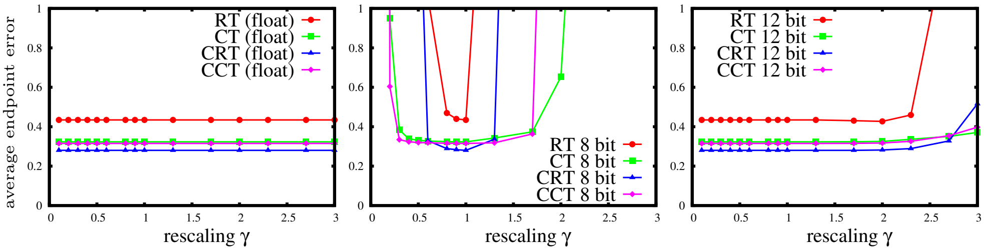


Figure 2: Behaviour under γ changes. The plots show the results of our method under γ variations of the second frames. *Left plot:* Behaviour without quantisation. *Centre plot:* Behaviour with re-quantisation with 8 bit after the γ rescaling. The invariance is destroyed. *Right plot:* Behaviour with 12 bit quantisation. For reasonable values for γ, the invariance is not affected.

digital images are quantised with 8 bit. As can be seen in Figure 2, the theoretically unconditional invariance does not hold for any transform in this case. However, many cameras that offer RAW sensor data quantise with 12 bit. Also many CMOS sensors and high-quality webcams offer a capture mode with such an increased dynamic range. Thus, we have also requantised the adjusted images with 12 bit and analysed those results. From Figure 2, one can see that these 4 bit more tonal resolution are in practice enough to restore the invariance almost completely. To ensure a fair comparison, the regularisation parameter α has been optimised for each graph.

Sensitivity to Noise. In this experiment we perturb the input image sequences with zero-mean Gaussian noise of varying standard deviations and measure the resulting accuracy. The outcome of this experiment is depicted graphically in Figure 3. Compared to the census transform, the complete rank as well as the complete census transform loose a bit less accuracy if the contamination with noise increases. Moreover, while the complete rank signature performs slightly better than complete census at low noise levels, this relation is reversed for higher levels where the complete census seems to be less vulnerable.

5.2 Real-world experiments

Since our method is tailored towards challenging illumination conditions, we also focus our evaluation on image material where such conditions are present. In that respect, the KITTI Vision Suite (Geiger et al, 2012) offers a good testbed for our needs. It provides a huge amount of image sequences captured from a driving car, along with corresponding ground truth flow

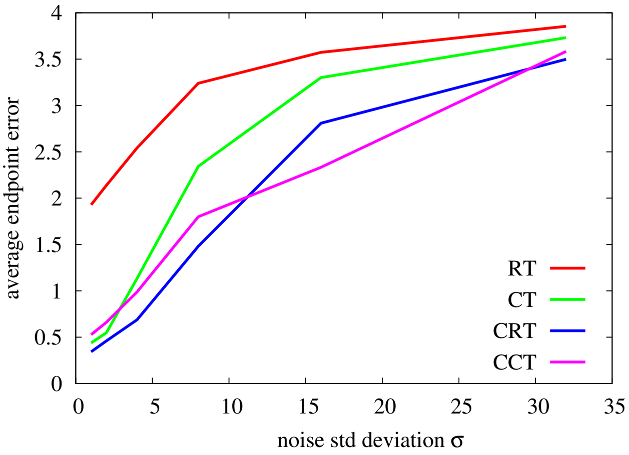


Figure 3: Behaviour of the average endpoint error under additive Gaussian noise of varying standard deviation. Depicted is the average endpoint error over the eight image sequences of the Middlebury training data set.

fields that are acquired with a laser scanning technique. Due to the inherent small-scale imprecisions of the ground truth data acquisition process of the KITTI benchmark, the usual error measures such as the average endpoint error (AEE) (Otte and Nagel, 1994) are not well suited for a quantitative evaluation. Thus, the common measure for the KITTI benchmark is the *bad pixel* (BP) error (Geiger et al, 2012):

$$\mathrm{BP}K(\boldsymbol{w}, \boldsymbol{w}_{\mathrm{gt}}) = \frac{1}{\Omega} \int_{\Omega} \mathbb{1}_{(|\boldsymbol{w}(\boldsymbol{x}) - \boldsymbol{w}_{\mathrm{gt}}(\boldsymbol{x})| < K)} \, \mathrm{d}\boldsymbol{x} \,. \tag{42}$$

For instance, the BP3 error, which we will always consider, expresses the percentage of estimated flow vectors that differ by more than 3 pixels form the measured ground truth solution, i.e. the percentage of pixels with an endpoint error above 3 pixels.

Neighbourhood Size. In our experiments, the neighbourhood size has shown to have a large impact on the results. In Figure 4, we depict the results of our experiment on this parameter. One can see that the complete rank and census transform outperform their incomplete predecessors. For the CRT descriptor, a minimum is attained at $k = 13$.

Regularisation Term. Next, we compare the TV-model from our conference paper to our improved TGV-model. To this end, we first compute flow fields for four real-world test sequences of the KITTI benchmark (Geiger

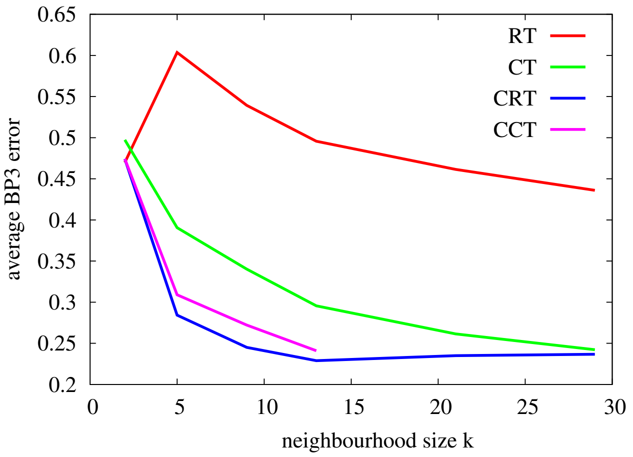


Figure 4: Behaviour of the average BP3 error when varying the size of the neighbourhood. Due to its high dimensionality, we did not test larger neighbourhoods for the complete census transform.

Table 2: Behaviour in real-world scenarios. Errors are given in terms of the *BP3* measure, i.e. the percentage of pixels having a Euclidean error larger than 3.

KITTI image sequence:	#11	#15	#44	#74	average
Zimmer et al (2011)	37.3	32.3	23.2	62.9	38.9
Bruhn and Weickert (2005)	33.9	47.7	32.4	71.4	46.7
Census Transform	36.5	28.6	28.5	63.8	39.4
Complete Rank Transform (TV)	29.8	22.8	22.6	61.5	34.2
Complete Rank Transform (TGV)	**22.9**	**13.5**	**15.2**	**56.3**	**27.0**

et al, 2012), which exhibit severe illumination changes. We have chosen the same set of images as selected for the *GCPR 2013 - Special Session on Robust Optical Flow*[1]. Table 2 summarises the obtained results. As reference, the numbers for the method of Zimmer et al (2011) and Bruhn and Weickert (2005) are taken from the website of this special session. The method of Bruhn and Weickert (2005) is particularly interesting to compare, since our former regularisation strategy is similar to the ideas in this paper. As one can see, the complete rank transform consistently outperforms the competing methods.

Public Benchmark Systems. First, we assess the error rates on the Middlebury training images, cf. Table 3. As also noted by Vogel et al (2013), the

[1] `http://www.dagm.de/symposien/special-sessions/`

Table 3: Quantitative comparison of the rank (RT), census (CT) and complete rank transform (CRT) on the Middlebury training images. Numbers are average endpoint errors $\times 10^{-1}$.

	rw	dim	gr2	gr3	hydr	urb2	urb3	yos	**avg**
RT	1.11	0.92	1.91	7.64	1.91	4.57	10.3	2.11	3.81
CT	1.02	0.90	1.69	6.46	**1.47**	3.78	8.19	1.69	3.16
CRT	**1.00**	**0.76**	**1.54**	**5.85**	1.58	**3.24**	**5.29**	**1.50**	**2.60**

Table 4: Detailed results of our method on the KITTI benchmark.

Error	**Out-Noc**	**Out-All**	**Avg-Noc**	**Avg-All**
2 pixels	8.84 %	15.38 %	2.0 px	3.9 px
3 pixels	6.71 %	12.09 %	2.0 px	3.9 px
4 pixels	5.68 %	10.23 %	2.0 px	3.9 px
5 pixels	5.01 %	8.97 %	2.0 px	3.9 px

image sequences of that benchmark exhibit mainly fronto-parallel motion, so we use our first order regularisation term here since it leads to better results. Furthermore, note that the Middlebury sequences are also less demanding with respect to illumination changes. Hence, the goal of this experiment is to show that also under normal lighting conditions reasonable flow fields can be obtained with our CRT-based data term. Furthermore, we prove with this experiment that our CRT is also in this setting generally preferable over the rank and census transform. Again, for each signature type, the regularisation parameter α has been optimised and then kept constant over all images.

The most interesting experiment here is the performance of our method in realistic scenarios, as provided by the KITTI Vision Benchmark Suite (Geiger et al, 2012). We have computed flow fields for all 195 test image sequences with a neighbourhood size of 13. Detailed results are shown in Table 4 where the bad pixel error measure is depicted for various thresholds and for ground truth information in *all* and *non-occluded* regions.

Additionally, we present in Table 5 a comparison of our method to the other participants of the benchmark. In our table, we only include published competing methods that consider the pure two-frame optic flow setup without stereoscopic assumptions. Methods that exploit such additional assumptions loose general applicability, because they are likely to fail e.g. in the presence of independently moving objects.

As one can see, our method clearly belongs to the top-ranking ones on this benchmark. Particularly when considering the ground truth information in all image regions, our method outperforms all others. One reason for this is

Table 5: Top KITTI benchmark results. We omitted all methods that use stereo information and all anonymous submissions.

Method	**Out-Noc**		**Out-All**		**Avg-Noc**		**Avg-All**	
TGV2ADCSIFT (Braux-Zin et al, 2013)	**6.20 %**	1	15.15 %	3	**1.5 px**	1	4.5 px	2
CRT-TGV (ours)	6.71 %	2	**12.09 %**	1	2.0 px	4	**3.9 px**	1
Data-Flow (Vogel et al, 2013)	7.11 %	3	14.57 %	2	1.9 px	3	5.5 px	3
DeepFlow (Weinzaepfel et al, 2013)	7.22 %	4	17.79 %	4	**1.5 px**	1	5.8 px	4
TVL1-HOG (Rashwan et al, 2013)	7.91 %	5	18.90 %	7	2.0 px	4	6.1 px	5
CRTflow (Demetz et al, 2013)	9.43 %	6	18.72 %	6	2.7 px	7	6.5 px	6
C++ (Sun et al, 2014)	10.04 %	7	20.26 %	8	2.6 px	6	7.1 px	8
C+NL (Sun et al, 2014)	10.49 %	8	20.64 %	9	2.8 px	8	7.2 px	9
fSGM (Hermann and Klette, 2013)	10.74 %	9	22.66 %	10	3.2 px	10	12.2 px	10
TGV2CENSUS (Ranftl et al, 2012)	11.03 %	10	18.37 %	5	2.9 px	9	6.6 px	7

the second order regulariser that is well suited for the typical divergent motion patterns of the KITTI benchmark. However, regarding the performance of the method of Ranftl et al (2012), which also incorporates a TGV-based regulariser, the benefits of our descriptor become apparent.

For the sake of completeness, we also evaluated our method with TV regularisation on the Middlebury benchmark (Baker et al, 2011). Since the test sequences of this benchmark exhibit almost no illumination changes or other scenarios that our highly invariant descriptor is designed for, we cannot expect top-ranking results on this benchmark. Nevertheless, it turns out that our prototypical variational model can in fact keep up with its nearest competitors: Our method ranks between the method of Brox et al (2004) and the much more advanced method by Zimmer et al (2009). These results are remarkable in the sense that they prove our invariant data term to include hardly less information than the combined grey value and gradient information of (Brox et al, 2004; Zimmer et al, 2009).

6 Conclusions

In our paper, we have driven the class of morphologically invariant local descriptors to the extreme: With the complete rank transform (CRT) and the complete census transform (CCT), we have introduced two descriptors that carry as much local image information as possible. Our transforms are well suited for pattern matching applications where highest accuracy is desired, such as optic flow estimation. We have demonstrated this within a variational framework, where we achieve state-of-the-art results for the KITTI benchmark. Since the CCT involves a natural distance metric, the comparison between two signatures is theoretically well justified. With the CRT, we present a lightweight and qualitatively good approximation to the CCT that offers much higher efficiency. We recommend it as the method of choice whenever robustness under uncontrolled lighting is essential.
In our ongoing work we are assessing the sparse feature matching capabilities of our signatures. First steps in this direction show promising results.

Acknowledgements. Our research is partly funded by the *Deutsche Forschungsgemeinschaft* through a *Gottfried Wilhelm Leibniz Prize* for Joachim Weickert. This is gratefully acknowledged.

References

Alvarez L, Guichard F, Lions PL, Morel JM (1993) Axioms and fundamental equations in image processing. Archive for Rational Mechanics and Analysis 123:199–257

Baker S, Scharstein D, Lewis JP, Roth S, Black MJ, Szeliski R (2011) A database and evaluation methodology for optical flow. International Journal of Computer Vision 92(1):1–31

Bhat DN, Nayar SK (1998) Ordinal measures for image correspondence. IEEE Trans Pattern Analysis and Machine Intelligence 20(4):415–423

Braux-Zin J, Dupont R, Bartoli A (2013) A general dense image matching framework combining direct and feature-based costs. In: Proc. IEEE International Conference on Computer Vision (ICCV), pp 185–192

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

Brox T, Bruhn A, Papenberg N, Weickert J (2004) High accuracy optical flow estimation based on a theory for warping. In: Pajdla T, Matas J (eds) Computer Vision – ECCV 2004, Part IV, Lecture Notes in Computer Science, vol 3024, Springer, Berlin, pp 25–36

Bruhn A, Weickert J (2005) Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In: Proc. IEEE International Conference on Computer Vision (ICCV), Beijing, China, vol 1, pp 749–755

Calonder M, Lepetit V, Ozuysal M, Trzcinski T, Strecha C, Fua P (2012) BRIEF: Computing a Local Binary Descriptor Very Fast. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(7):1281–1298

Chan CH, Goswami B, Kittler J, Christmas W (2012) Local ordinal contrast pattern histograms for spatiotemporal, lip-based speaker authentication. IEEE Transactions on Information Forensics and Security 7(2):602–612

Chen J, Kellokumpu VP, Zhao G, Pietikinen M (2013) RLBP: Robust local binary pattern. In: Burghardt T, Damen D, Mayol-Cuevas W, Mirmehdi M (eds) Proc. British Machine Vision Conference, BMVA Press, Bristol, UK

Cohen I (1993) Nonlinear variational method for optical flow computation. In: Proc. 8th Scandinavian Conference on Image Analysis, Tromsø, Norway, vol 1, pp 523–530

Demetz O, Hafner D, Weickert J (2013) The complete rank transform: A tool for accurate and morphologically invariant matching of structures. In: Burghardt T, Damen D, Mayol-Cuevas W, Mirmehdi M (eds) Proc. British Machine Vision Conference, BMVA Press, Bristol, UK

Fröba B, Ernst A (2004) Face detection with the modified census transform. In: Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR), Seoul, Korea, pp 91–96

Geiger A, Lenz P, Urtasun R (2012) Are we ready for autonomous driving? The KITTI vision benchmark suite. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, pp 3354–3361

Gelfand IM, Fomin SV (2000) Calculus of Variations. Dover, New York

Gennert MA, Negahdaripour S (1987) Relaxing the brightness constancy assumption in computing optical flow. Tech. Rep. 975, Artificial Intelligence Laboratory, Massachusetts Institiute of Technology

Grewenig S, Weickert J, Schroers C, Bruhn A (2013) Cyclic schemes for PDE-based image analysis. Tech. Rep. 327, Department of Mathematics, Saarland University, Saarbrücken, Germany

Hafner D, Demetz O, Weickert J (2013) Why is the census transform good for robust optic flow computation? In: Kuijper A, Pock T, Bredies K, Bischof H (eds) Scale-Space and Variational Methods in Computer Vision, Lecture Notes in Computer Science, vol 7893, Springer, Berlin, pp 210–221

Hermann S, Klette R (2013) Hierarchical scan-line dynamic programming for optical flow using semi-global matching. In: Park JI, Kim J (eds) Computer Vision - ACCV 2012 Workshops, Lecture Notes in Computer Science, vol 7729, Springer, Berlin, pp 556–567

Hewer A, Weickert J, Scheffer T, Seibert H, Diebels S (2013) Lagrangian strain tensor computation with higher order variational models. In: Burghardt T, Damen D, Mayol-Cuevas W, Mirmehdi M (eds) Proc. British Machine Vision Conference, BMVA Press, Bristol, UK

Horn B, Schunck B (1981) Determining optical flow. Artificial Intelligence 17:185–203

Kim TH, Lee HS, Lee KM (2013) Optical flow via locally adaptive fusion of complementary data costs. In: Proc. IEEE International Conference on Computer Vision (ICCV), IEEE Press, pp 3344–3351

Liu C, Yuen J, Torralba A (2011) SIFT flow: Dense correspondence across scenes and its applications. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(5):978–994

Lowe DL (2004) Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2):91–110

Mei X, Sun X, Zhou M, Jiao S, Wang H, Zhang X (2011) On building an accurate stereo matching system on graphics hardware. In: Proc. IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, pp 467–474

Mileva Y, Bruhn A, Weickert J (2007) Illumination-robust variational optical flow with photometric invariants. In: Hamprecht FA, Schnörr C, Jähne B (eds) Pattern Recognition, Lecture Notes in Computer Science, vol 4713, Springer, pp 152–162

Mittal A, Ramesh V (2006) An intensity-augmented ordinal measure for visual correspondence. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), New York, NY, vol 1, pp 849–856

Müller T, Rabe C, Rannacher J, Franke U, Mester R (2011) Illumination robust dense optical flow using census signatures. In: Mester R, Felsberg M (eds) Pattern Recognition, Lecture Notes in Computer Science, vol 6835, Springer, pp 236–245

Otte M, Nagel HH (1994) Optical flow estimation: Advances and comparisons. In: Eklundh JO (ed) Computer Vision – ECCV '94, Lecture Notes in Computer Science, vol 800, Springer, Berlin, pp 49–60

Papenberg N, Bruhn A, Brox T, Didas S, Weickert J (2006) Highly accurate optic flow computation with theoretically justified warping. International Journal of Computer Vision 67(2):141–158

Pietikäinen M, Hadid A, Zhao G, Ahonen T (2011) Computer Vision Using Local Binary Patterns. Springer, London

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical Recipes: The Art of Scientific Computing, 3rd edn. Cambridge University Press

Puxbaum P, Ambrosch K (2010) Gradient-based modified census transform for optical flow. In: Bebis G, Boyle RD, Parvin B, Koracin D, Chung R, Hammoud RI, Hussain M, Tan KH, Crawfis R, Thalmann D, Kao D, Avila L (eds) Advances in Visual Computing, Part I, Lecture Notes in Computer Science, vol 6453, Springer, pp 437–448

Ranftl R, Gehrig S, Pock T, Bischof H (2012) Pushing the limits of stereo using variational stereo estimation. In: Proc. IEEE Intelligent Vehicles Symposium, Alcala de Henares, Spain, pp 401–407

Rashwan H, Mohamed M, Garcia M, Mertsching B, Puig D (2013) Illumination robust optical flow model based on histogram of oriented gradients. In: Weickert J, Hein M, Schiele B (eds) Pattern Recognition, Springer, Berlin, Lecture Notes in Computer Science, vol 8142, pp 354–363

Schnörr C (1994) Segmentation of visual motion by minimizing convex non-quadratic functionals. In: Proc. IEEE International Conference on Pattern Recognition (ICPR), Jerusalem, Israel, vol A, pp 661–663

Sloane NJA, Plouffe S (1995) The Encyclopedia of Integer Sequences. Academic Press, San Diego

Stein F (2004) Efficient computation of optical flow using the census transform. In: Rasmussen CE, Bülthoff HH, Schölkopf B, Giese MA (eds) Pattern Recognition, Lecture Notes in Computer Science, vol 3175, Springer, Berlin, pp 79–86

Steinbrücker F, Pock T, Cremers D (2009) Advanced data terms for variational optic flow estimation. In: Magnor MA, Rosenhahn B, Theisel H (eds) Proceedings of the Vision, Modeling, and Visualization Workshop (VMV), DNB, pp 155–164

Sun D, Roth S, Black M (2014) A quantitative analysis of current practices in optical flow estimation and the principles behind them. International Journal of Computer Vision 106(2):115–137

Tang F, Lim SH, Chang NL, Tao H (2009) A novel feature descriptor invariant to complex brightness changes. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, pp 2631–2638

Tukey JW (1971) Exploratory Data Analysis. Addison-Wesley, Menlo Park

Uras S, Girosi F, Verri A, Torre V (1988) A computational approach to motion perception. Biological Cybernetics 60:79–87

van de Weijer J, Gevers T (2004) Robust optical flow from photometric invariants. In: Proc. IEEE International Conference on Image Processing (ICIP), pp 1835–1838

Vogel C, Roth S, Schindler K (2013) An evaluation of data costs for optical flow. In: Weickert J, Hein M, Schiele B (eds) Pattern Recognition, Lecture Notes in Computer Science, vol 8142, Springer, Berlin, pp 343–353

Vogel CR, Oman ME (1996) Iterative methods for total variation denoising. SIAM Journal on Scientific Computing 17(1):227–238

Wang Z, Fan B, Wu F (2011) Local intensity order pattern for feature description. In: Proc. IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, pp 603–610

Wedel A, Pock T, Zach C, Cremers D, Bischof H (2008) An improved algorithm for TV-L1 optical flow. In: Cremers D, Rosenhahn B, Yuille AL, Schmidt FR (eds) Statistical and Geometrical Approaches to Visual Motion Analysis, Lecture Notes in Computer Science, vol 5604, Springer, Berlin

Weinzaepfel P, Revaud J, Harchaoui Z, Schmid C (2013) Deepflow: Large displacement optical flow with deep matching. In: Proc. IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, pp 1385–1392

Werlberger M, Pock T, Bischof H (2010) Motion estimation with non-local total variation regularization. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2464–2471

Xu L, Jia J, Matsushita Y (2010) Motion detail preserving optical flow estimation. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society Press, pp 1293–1300

Zabih R, Woodfill J (1994) Non-parametric local transforms for computing visual correspondence. In: Eklundh JO (ed) Computer Vision – ECCV '94, Part II, Lecture Notes in Computer Science, vol 801, Springer, Berlin, pp 151–158

Zimmer H, Bruhn A, Weickert J, Valgaerts L, Salgado A, Rosenhahn B, Seidel HP (2009) Complementary optic flow. In: Cremers D, Boykov Y, Blake A, Schmidt FR (eds) Energy Minimization Methods in Computer Vision and Pattern Recognition, Lecture Notes in Computer Science, vol 5681, Springer, Berlin, pp 207–220

Zimmer H, Bruhn A, Weickert J (2011) Optic flow in harmony. International Journal of Computer Vision 3(93):368–388