The Complete Rank Transform: A Tool for Accurate and Morphologically Invariant Matching of Structures

Oliver Demetz
demetz@mia.uni-saarland.de
David Hafner
hafner@mia.uni-saarland.de
Joachim Weickert
weickert@mia.uni-saarland.de

Mathematical Image Analysis Group
Saarland University
Saarbrücken, DE

Abstract

Most researchers agree that invariances are desirable in computer vision systems. However, one always has to keep in mind that this is at the expense of accuracy: By construction, all invariances inevitably discard information. The concept of morphological invariance is a good example for this trade-off and will be in the focus of this paper. Our goal is to develop a descriptor of local image structure that carries the *maximally possible* amount of local image information under this invariance. To fulfill this requirement, our descriptor has to encode the full ordering of the pixel intensities in the local neighbourhood. As a solution, we introduce the *complete rank transform*, which stores the intensity rank of every pixel in the local patch. As a proof of concept, we embed our novel descriptor in a prototypical $TV\text{-}L^1$-type energy functional for optical flow computation, which we minimise with a traditional coarse-to-fine warping scheme. In this straightforward framework, we demonstrate that our descriptor is preferable over related features that exhibit the same invariance. Finally, we show by means of public benchmark systems that our method produces - in spite of its simplicity - results of competitive quality.

1 Introduction

Incorporating invariances is a popular tool to make computer vision systems more robust under real-world conditions. Since being invariant means ignoring something, every invariance has to come at the price of a loss of information. Thus, there is an intrinsic trade-off between invariances and high accuracy. In our paper, we focus on severe illumination changes and descriptors that are invariant under such challenging conditions. In particular, we consider transforms that are morphologically invariant in the sense of being unaffected by monotonically increasing greyscale transformations [1]. This class comprises all descriptors that are based on the grey value order. A famous example is the median filter [22]. Zabih and Woodfill's *rank transform* [25] is another prominent representative of such illumination robust descriptors. It computes the rank of a pixel's intensity within the 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 [25], the *census transform* is proposed, which compares a pixel with all its neighbours and stores which one is larger. Besides encoding the rank, also some spatial information is stored in this descriptor. However, also here a lot of information is discarded. Thus, it would be desirable to develop a robust feature with strong invariance that discards as little information as possible.

1.1 Our Contribution

In this paper, we propose a novel descriptor that carries much more information than the rank, but exhibits the same strong invariance. Our general idea is to not restrict ourselves to the rank of the central pixel, but to compute the ranks of all pixels of the neighbourhood, i.e. to store the complete local intensity order. We claim that our novel complete rank descriptor can be used as a generally superior alternative to the census transform: It is equally computationally efficient, as parameter-free as the census transform, and leads 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 the focus of this paper and would examine very different aspects and properties of a descriptor.

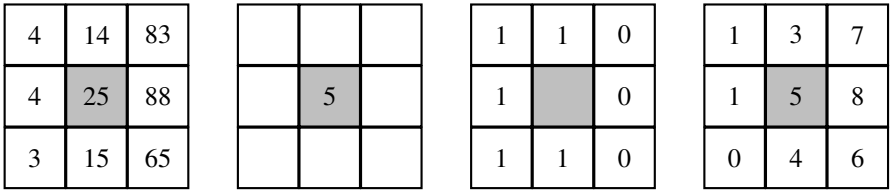
1.2 Related Work

Independently of Zabih and Woodfill's rank and census transforms [25], Pietikäinen et al. performed broad research on various kinds of *local binary patterns* (see [16] and references therein). Stein et al. [20] used the census transform as an efficient descriptor for structure matching in driver assistance systems and Fröba and Ernst [7] used the modified census transform for face recognition. More recently, several sparse interest point descriptors that are related to intensity order-based ideas have been proposed: With their chained circular neighbourhoods, Chan et al. [6] made a first step towards representing neighbourhood ordinal information. The LIOP descriptor of Wang et al. [23] describes the intensity order of a very large neighbourhood which is tailored for sparse interest point matching. A similar idea of matching order distributions is proposed by Tang et al. [21]. Mittal and Ramesh [14] 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, e.g. the structure-texture decomposition by Wedel et al. [24] or various invariant colour spaces in Mileva et al. [13]. There are several recent publications that incorporate the census transform in variational optical flow or stereo methods: Müller et al. propose a census-based data term for optical flow [15], and Ranftl et al. [18] as well as Mei et al. [12] present census-based stereo methods. The theoretical study of Hafner et al. [10] explains the reasons why census-based data terms for variational optical flow are successful.

1.3 Organisation of our Publication

Our paper is organised as follows: In Section 2, we review the rank and census transforms. After that, Section 3 introduces the complete rank transform. In Section 4, we embed our novel descriptor into a variational framework and demonstrate its benefits in the experimental Section 5. We conclude the paper with a summary and an outlook in Section 6.



(a) Intensity values. (b) Rank. (c) Census. (d) Complete rank.

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.

2 Rank and Census Transform

Before describing our novel patch-based descriptor, let us briefly revisit two closely related descriptors by Zabih and Woodfill [25] that form the basis of our proposal. Both of them belong to the class of patch-based intensity order descriptors and share the property of being *morphologically invariant*, i.e. invariant with respect to any monotonically increasing grey value rescalings. In particular, they describe a signature vector $\boldsymbol{s}$ that encodes how the grey value of a reference pixel compares to its neighbours.

The *rank transform* (RT) encodes for each pixel the position of its grey value in the ranking of all grey values in its neighbourhood. Practically, this rank is determined by counting the number of neighbours with a smaller grey value than the reference pixel. Typically, a neighbourhood consists of a square patch of $k := K \times K$ pixels. Then, the rank transform formally maps each pixel to its scalar rank signature $s_{\mathrm{RT}} \in \{0, \dots, k-1\}$. Regarding our sample intensity patch in Figure 1(a), we have

$$s_{\mathrm{RT}} = 5\,, \tag{1}$$

since the five neighbouring grey values 3, 4, 4, 14, and 15 are smaller than the intensity 25 of the central pixel.

The second descriptor is the *census transform* (CT), which has also been introduced by Zabih and Woodfill [25]. It 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. In practice, 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, otherwise it is 0. In the final binary signature, all bits are concatenated. While the order of this concatenation is in general arbitrary, it must be consistent in the whole image such that each bit can be uniquely associated with one neighbour. In mathematical terms, each pixel in the image is mapped to a binary signature $\boldsymbol{s}_{\mathrm{CT}} \in \{0,1\}^{k-1}$ of length $k-1$. Consequently, the census signature of our example reads

$$\boldsymbol{s}_{\mathrm{CT}} = (1,1,0,1,0,1,1,0)^{\top}\,, \tag{2}$$

where the binary digits are read out line by line, from top left to bottom right, cf. Figure 1(c). It is obvious that the number of ones in a census signature coincides with the rank of that pixel.

3 Complete Rank Transform

Although the two signatures by Zabih and Woodfill [25] exhibit the same morphological invariance, by construction the census transform obviously encodes more information than the pure rank.

The goal of this section is to propose a novel transform that extracts as much local image information as possible of the input data while preserving the same desired invariance against monotonically increasing illumination changes. To this end, we make use of the following observation: Not only the rank of the central intensity value, but even the *complete intensity order* of the considered patch is morphologically invariant.

Hence, we will now modify and extend the rank transform in such a way that it incorporates this novel idea of encoding the *complete* intensity order instead of only encoding the rank of the reference pixel.

3.1 Construction

The construction of our novel *complete rank transform* (CRT) is straightforward: First, we compute the rank of each pixel in the considered neighbourhood patch. Practically, for each element of the patch this comes down to determining the number of pixels with smaller intensity. Next, we concatenate these ranks as for the census transform to obtain the complete rank descriptor. Mathematically, for a patch size k, the complete rank transform is a mapping

$$\boldsymbol{s}_{\mathrm{CRT}} : \Omega \to \Pi_k \subset \{0,\dots,k-1\}^k, \tag{3}$$

where the co-domain Π_k is the set of all possible rankings of k elements including multiple occurrences of the same intensity value. The cardinality of this set is the k-th ordered Bell number (OBN(k)). The principle of our CRT is illustrated in Figure 1(d). The corresponding signature is given by

$$\boldsymbol{s}_{\mathrm{CRT}} = (1,3,7,1,5,8,0,4,6)^\top, \tag{4}$$

where the digits are, as before, read out line by line.

3.2 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.

As already mentioned, Zabih and Woodfill's census transform carries spatial information which is not included in the rank transform. Instead of storing all ranks, one could alternatively also apply the census transform to every pixel in the patch, i.e. compare each pixel with each other pixel. However, in this complete case, both transforms comprise this spatial information component by construction. Consequently, there is no difference between the two complete signatures, and there even exists a bijection: Each complete rank signature uniquely induces a complete census signature, and vice versa. Hence, exactly the same amount of information is contained in both (complete) signature types. The reason, however,

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

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

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 [10], 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 [17] 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.

4 Variational Optical Flow Model

In this section, our goal is to embed the considered signatures into a variational energy functional for optical flow computation [11]. To this end, we assume that the input images have been mapped by one of the introduced transforms to a vector-valued function $\boldsymbol{c}:\Omega\times[0,\infty)\to D^{\kappa}$. For colour images, typically the signature length is tripled since we concatenate the signatures of each channel.

4.1 Energy Formulation

We propose a generic variational model which assumes the signatures of corresponding pixels in the first frame and in the second frame to coincide. This assumption, together with a prior knowledge about the spatial regularity of the flow field, is expressed in the functional

$$E(u,v)=\int_{\Omega}\Big(\,\Psi\big(\tfrac{1}{\kappa}|\boldsymbol{c}(\boldsymbol{x}+\boldsymbol{w})-\boldsymbol{c}(\boldsymbol{x})|^2\big)+\alpha\cdot\Psi\big(|\nabla u|^2+|\nabla v|^2\big)\,\Big)\,\mathrm{d}\boldsymbol{x}, \tag{5}$$

whose minimiser is the sought flow field $(u,v)^{\top}:\Omega\to\mathbb{R}^2$ and the rectangular image domain is denoted by $\Omega\subset\mathbb{R}^2$. The first term (*data term*) 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. The second term (*smoothness term*) penalises the magnitude of the gradient $\nabla:=(\partial_x,\partial_y)^{\top}$ of the flow field. Further, the two terms are balanced by the regularisation parameter $\alpha>0$.

We would like to point out that we intentionally choose a functional with such a transparent and generic structure: In the data term, we do not assume the constancy of any derivatives or higher order expressions of the signature and apply a joint robustification (also in the case of colour images). The same holds for the smoothness term, where we do not incorporate more advanced concepts. The penaliser function [6, 19]

$$\Psi(s^2) = 2\lambda\sqrt{s^2+\lambda^2} - 2\lambda^2 \tag{6}$$

is also the same with equal numerical parameter $\lambda = 10^{-2}$ for both terms. As a consequence, the influence of the various descriptors can be clearly examined. Any further model assumptions or more sophisticated penalisation or regularisation strategies could just obscure the influence and effects of our robust data descriptors. This combination can be seen as a continuous and differentiable approximation to the widely-used $TV-L^1$ model [26].

4.2 Minimisation

According to the calculus of variations, the minimiser of our energy functional has to fulfil the Euler-Lagrange equations. Before stating them explicitly, let us introduce the following abbreviations for the sake of readability:

$$\boldsymbol{c}_x := \partial_x \boldsymbol{c}(\boldsymbol{x}+\boldsymbol{w}), \qquad \boldsymbol{c}_y := \partial_y \boldsymbol{c}(\boldsymbol{x}+\boldsymbol{w}), \qquad \boldsymbol{c}_t := \boldsymbol{c}(\boldsymbol{x}+\boldsymbol{w}) - \boldsymbol{c}(\boldsymbol{x}), \tag{7}$$

$$\Psi'_d := \tfrac{1}{\kappa}\Psi'\left(\tfrac{1}{\kappa}|\boldsymbol{c}_t|^2\right), \qquad \Psi'_s := \Psi'\left(|\nabla u|^2 + |\nabla v|^2\right). \tag{8}$$

Here, the spatial derivatives of vector-valued quantities are computed componentwise. With these abbreviations, the minimality conditions can be concisely formulated by

$$\Psi'_d \cdot \boldsymbol{c}_t^\top \boldsymbol{c}_x - \alpha \cdot \operatorname{div}\left(\Psi'_s \cdot \nabla u\right) = 0, \tag{9}$$

$$\Psi'_d \cdot \boldsymbol{c}_t^\top \boldsymbol{c}_y - \alpha \cdot \operatorname{div}\left(\Psi'_s \cdot \nabla v\right) = 0, \tag{10}$$

with $\boldsymbol{n}^\top \nabla u = 0$ and $\boldsymbol{n}^\top \nabla v = 0$ as boundary conditions.

4.3 Numerical Algorithm and Implementation

Our solution of the PDEs (9) and (10) essentially resembles the warping scheme by Brox *et al.* [3]. However, when implementing this warping scheme it is important not to apply the coarse-to-fine strategy to the raw input images, because any averaging in the downsampling strategy or interpolation destroys the morphological invariance of the descriptor. To preserve this invariance of the CRT signatures, the images *must not* be resampled or smoothed. The remedy is to first compute the signatures of the original input images and subsequently apply the coarse-to-fine warping pyramid to these computed signatures.

All components of our method and especially the *Fast Jacobi* solver [9] are perfectly suited for an implementation on parallel hardware architectures such as modern GPUs. 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.

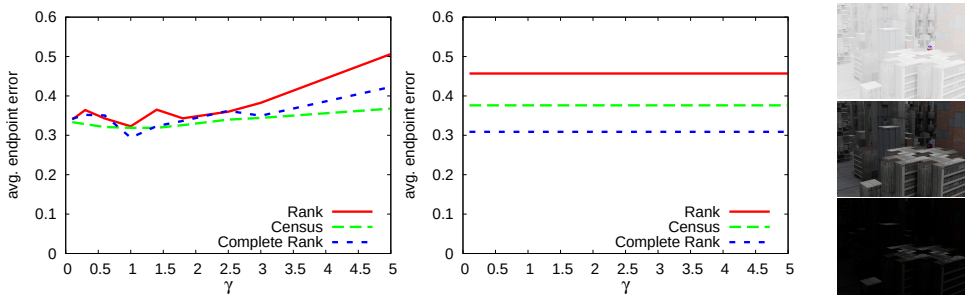


Figure 2: Behaviour under γ changes. The plots show the results of our method under γ variations of the second frame of the *Urban2* sequence. *Left plot:* Behaviour when warping the raw images. *Right plot:* Behaviour when warping the signature data. *Right, from top to bottom:* γ-corrected second frames with $\gamma = \{0.1, 1, 3\}$.

5 Experiments

Choice of parameters. Due to its simplicity, only very few parameters have to be chosen. The downscaling factor of the coarse-to-fine warping scheme is fixed to 0.95 and we repeat the incremental flow computations on each level 4 times. Furthermore, the numerical parameter for the robust function Ψ is kept constant at $\lambda = 10^{-2}$. Practically, this means that the only free parameter of our optical flow method is the regularisation weight α, which we optimise for every experiment. However, throughout our experiments, the optimal α always was in the interval $[0.05, 0.2]$ for the CRT-based data term.

Invariance to γ changes. Our first experiment examines the behaviour of the proposed method under monotonically increasing intensity changes. To this end, we consider a typical image sequence $f : \Omega \to [0, 255]$ and apply a γ-correction to each channel of the second frame:

$$f_\gamma(\boldsymbol{x}) := 255 \cdot \left(\tfrac{1}{255} f(\boldsymbol{x})\right)^\gamma \tag{11}$$

In Figure 2, we illustrate the necessity of the modification of the warping scheme. As expected, if the raw images are subject to sampling and smoothing in the coarse-to-fine approach, the invariance of the signatures is lost. In contrast, by first evaluating our invariant descriptor and then applying the coarse-to-fine scheme, the desired unconditional invariance is achieved in practice. For this experiment we computed flow fields under varying γ-corrections using the *Urban2* training sequence of the Middlebury [7] benchmark. To ensure a fair comparison, the regularisation parameter α has been optimised for each graph.

Accuracy Evaluation. In our second experiment, we compare the accuracy of the various descriptors. To this end, we first compute flow fields for four real-world test sequences of the KITTI benchmark [8], 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. The numbers for the method of Zimmer et al. [23] and Bruhn and Weickert [4] are taken from this website. The method [4] is particularly interesting to compare, since our regularisation and minimisation strategy follows the ideas

[1] http://dagm.de/index.php?Itemid=149

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. [28]	37.3	32.3	23.2	62.9	38.9
Bruhn/Weickert [4]	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	**29.8**	**22.8**	**22.6**	**61.5**	**34.2**

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	dimetr.	grove2	grove3	hydr.	urban2	urban3	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**

in this paper. As one can see, the complete rank transform consistently outperforms the competing methods.

Next we assess the error rates on the Middlebury training images, cf. Table 3. Note that these sequences are 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.

Public benchmark systems. We also assess the accuracy of our method on the novel KITTI Vision Benchmark Suite [8], which offers a large amount of training and testing image sequences of real-world outdoor driving situations. Our method ranks 9th at the time of submission. However, the four top-ranking methods on this benchmark (parenthesised in Table 4) additionally rely on stereo information, which we do not need.

For the sake of completeness, we also evaluated our method on the Middlebury benchmark [2]. 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: At the time of submission, our method ranks 43.5th whereas the method of Brox et al. [3] ranks 49.5th. Moreover, the much more advanced method by Zimmer et al. [27] only ranks slightly higher on rank 38.7. These results are remarkable, since they prove that the invariant data term of our simple model includes hardly less information than the combined grey value and gradient information of [3, 27].

6 Conclusions

With the complete rank transform (CRT), we have introduced a novel descriptor of local image structure that is invariant under all monotonically increasing greyscale rescalings (morphological invariance). Although this descriptor is not the first one exhibiting this invariance, it preserves that maximal amount of local image information within the class of morpholog-

Table 4: Top 20 KITTI benchmark results. Our proposed method *CRTflow* is written in boldface. Methods in brackets exploit stereo information.

Rank	Method	Out-Noc	Out-All	Avg-Noc	Avg-All
(1)	(PR-Sf+E)	(4.08 %)	(7.79 %)	(0.9 px)	(1.7 px)
(2)	(PCBP-Flow)	(4.08 %)	(8.70 %)	(0.9 px)	(2.2 px)
(3)	(MotionSLIC)	(4.36 %)	(10.91 %)	(1.0 px)	(2.7 px)
(4)	(PR-Sceneflow)	(4.48 %)	(8.98 %)	(1.3 px)	(3.3 px)
5	TGV2ADCSIFT	6.55 %	15.35 %	1.6 px	4.5 px
6	Data-Flow	8.22 %	15.78 %	2.3 px	5.7 px
7	TVL1-HOG	8.31 %	19.21 %	2.0 px	6.1 px
8	MLDP-OF	8.91 %	18.95 %	2.5 px	6.7 px
9	**CRTflow**	**9.71 %**	**18.88 %**	**2.7 px**	**6.5 px**
10	fSGM	11.03 %	22.90 %	3.2 px	12.2 px
11	TGV2CENSUS	11.14 %	18.42 %	2.9 px	6.6 px
12	IQFlow	18.93 %	28.33 %	3.6 px	8.8 px
(13)	(GC-BM-Bino)	(18.93 %)	(29.37 %)	(5.0 px)	(12.0 px)
14	C+NL-M	19.17 %	26.35 %	7.4 px	14.5 px
15	eFolki	19.34 %	28.79 %	5.2 px	10.8 px
(16)	(GC-BM-Mono)	(19.49 %)	(29.88 %)	(5.0 px)	(12.1 px)
17	HS	19.92 %	28.86 %	5.8 px	11.7 px
18	RSRS-Flow	20.74 %	29.68 %	6.2 px	12.1 px
19	ALD	21.35 %	30.65 %	10.9 px	16.0 px
20	LDOF	21.86 %	31.31 %	5.5 px	12.4 px

ically invariant descriptors. This makes it particularly attractive for pattern matching applications where also high accuracy is desired such as optical flow estimation. Our embedding of the CRT into a $TV-L^1$-energy for computing the optical flow confirms this. The experiments give two insights: First, in contrast to the plain census-based descriptor, our complete rank does give good results even in the case that no challenging illumination changes occur. Second, the novel signature is consistently superior to the census transform, without introducing additional computational costs. Hence, it should serve as a generally preferable alternative to it. Currently, we are integrating our novel descriptor in more advanced models and assess its sparse feature matching capabilities. We are also interested in adaptive strategies to decide locally between our robust descriptor and other constancy assumptions.

References

[1] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel. Axioms and fundamental equations in image processing. *Archive for Rational Mechanics and Analysis*, 123:199–257, 1993.

[2] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011.

[3] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estima-

tion based on a theory for warping. In T. Pajdla and J. Matas, editors, *Computer Vision – ECCV 2004*, volume 3024 of *LNCS*, pages 25–36. Springer, 2004.

[4] A. Bruhn and J. Weickert. Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In *Proc. Tenth International Conference on Computer Vision*, volume 1, pages 749–755, Beijing, China, October 2005. IEEE Computer Society Press.

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

[6] I. Cohen. Nonlinear variational method for optical flow computation. In *Proc. Eighth Scandinavian Conference on Image Analysis*, volume 1, pages 523–530, Tromsø, Norway, May 1993.

[7] Bernhard Fröba and Andreas Ernst. Face detection with the modified census transform. In *Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR 2004)*, pages 91–96. IEEE Computer Society, 2004.

[8] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, USA, June 2012.

[9] S. Grewenig, J. Weickert, C. Schroers, and A. Bruhn. Cyclic schemes for PDE-based image analysis. Technical Report 327, Department of Mathematics and Computer Science, Saarland University, Saarbrücken, March 2013.

[10] D. Hafner, O. Demetz, and J. Weickert. Why is the census transform good for robust optic flow computation? In A. Kuijper, T. Pock, K. Bredies, and H. Bischof, editors, *Scale-Space and Variational Methods in Computer Vision*. Springer, 2013.

[11] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[12] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang. On building an accurate stereo matching system on graphics hardware. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 467–474. IEEE, 2011.

[13] Y. Mileva, A. Bruhn, and J. Weickert. Illumination-robust variational optical flow with photometric invariants. In F. A. Hamprecht, C. Schnörr, and B. Jähne, editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 152–162. Springer, 2007.

[14] A. Mittal and V. Ramesh. An intensity-augmented ordinal measure for visual correspondence. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 849–856, 2006.

[15] T. Müller, C. Rabe, J. Rannacher, U Franke, and R. Mester. Illumination robust dense optical flow using census signatures. In R. Mester and M. Felsberg, editors, *Pattern Recognition*, volume 6835 of *Lecture Notes in Computer Science*. Springer, September 2011.

[16] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen. *Computer Vision Using Local Binary Patterns*. Springer, London, 2011.

[17] Philipp Puxbaum and Kristian Ambrosch. Gradient-based modified census transform for optical flow. In George Bebis, Richard D. Boyle, Bahram Parvin, Darko Koracin, Ronald Chung, Riad I. Hammoud, Muhammad Hussain, Kar-Han Tan, Roger Crawfis, Daniel Thalmann, David Kao, and Lisa Avila, editors, *Proc. International Symposium on Advances in Visual Computing (ISVC),, Part I*, volume 6453 of *Lecture Notes in Computer Science*, pages 437–448. Springer, 2010.

[18] R. Ranftl, S. Gehrig, T. Pock, and H. Bischof. Pushing the limits of stereo using variational stereo estimation. In *IEEE Intelligent Vehicles Symposium*, pages 401–407, 2012.

[19] C. Schnörr. Segmentation of visual motion by minimizing convex non-quadratic functionals. In *Proc. Twelfth International Conference on Pattern Recognition*, volume A, pages 661–663, Jerusalem, Israel, October 1994. IEEE Computer Society Press.

[20] Fridtjof Stein. Efficient computation of optical flow using the census transform. In C. E. Rasmussen, H. H. Bülthoff, B. Schölkopf, and M. A. Giese, editors, *Pattern Recognition*, volume 3175 of *Lecture Notes in Computer Science*, pages 79–86. Springer, 2004.

[21] F. Tang, S. H. Lim, N. L. Chang, and H. Tao. A novel feature descriptor invariant to complex brightness changes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2631–2638, 2009.

[22] J. W. Tukey. *Exploratory Data Analysis*. Addison–Wesley, Menlo Park, 1971.

[23] Z. Wang, B. Fan, and F. Wu. Local intensity order pattern for feature description. In *IEEE International Conference on Computer Vision (ICCV)*, pages 603–610, November 2011.

[24] A. Wedel, T. Pock, C. Zach, D. Cremers, and H. Bischof. An improved algorithm for TV-L1 optical flow. In D. Cremers, B. Rosenhahn, A.L. Yuille, and F.R. Schmidt, editors, *Statistical and Geometrical Approaches to Visual Motion Analysis*, Lecture Notes in Computer Science. Springer, September 2008.

[25] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In Jan-Olof Eklundh, editor, *Computer Vision – ECCV '94*, volume 801 of *Lecture Notes in Computer Science*, pages 151–158. Springer, Berlin, 1994.

[26] C. Zach, Pock T., and Bischof H. A duality based approach for realtime $tv-l^1$ optical flow. In F. Hamprecht, C. SchnÃűrr, and B. JÃd'hne, editors, *Pattern Recognition, Proceedings of the 29nd DAGM*, volume 4713 of *Lecture Notes in Computer Science*, pages 214–223. Springer, Berlin, 2007.

[27] H. Zimmer, A. Bruhn, J. Weickert, L. Valgaerts, A. Salgado, B. Rosenhahn, and H.-P. Seidel. Complementary optic flow. In D. Cremers, Y. Boykov, A. Blake, and F. R. Schmidt, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 5681 of *Lecture Notes in Computer Science*, pages 207–220. Springer, Berlin, 2009.

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