Composition of Similarity Metrics for Correspondence Matching in Depth Estimation

Hubert Żabiński
Poznań University of Technology
Institute of Multimedia Telecommunications
Poznań, Poland
ORCID: 0000-0003-3345-2110

Olgierd Stankiewicz
Poznań University of Technology
Institute of Multimedia Telecommunications
Poznań, Poland
ORCID: 0000-0001-9691-9094

***Abstract*— Correspondence matching is a prerequisite step in dense depth estimation techniques. In this paper we consider various similarity metrics for correspondence matching and we present an approach which can be used to optimize it. Experimental results show that by careful selection of similarity metric can have positive impact on depth estimation quality and that the differences between various metrics range up to 60 percent points of bad-pixel depth map quality ratio. It has also been shown that usage of proposed composite similarity can lead to improved depth map quality, expressed as lower bad-pixel ratio.**

Keywords—similarity metric, similarity measure, depth estimation

I. INTRODUCTION

The modeling of 3D scenes from sets of views is an important task in many modern applications, which include 3D television (3DTV), immersive 6-DoF, robot vision, and self-driving cars. In such applications, there is a need for depth information which represent distances to the objects in the scene. One of the passive 3D depth sensing methods is stereo matching, which has been subject of extensive computer vision research during recent years. In the simplest form, stereo matching techniques use image pair analysis and correspondence search between image fragments to determine the disparity value for each point [1]. The quality of the estimated depth depends largely on the similarity metric (also known as dissimilarity/matching metric) used to find correspondence between the analyzed image fragments.

Many different similarity metrics for image fragment matching are known in the literature [2]. The simplest and most common are sum of absolute difference (SAD) and sum of square difference (SSD). Typically, these metrics are computed for block sizes ranging from as small as 1×1 to 32×32 and larger. More complex matching metrics, such as RANK or CENSUS have also been proposed in paper [3], [4]. Additionally, these metrics can be computed on image samples in multiple color spaces: RGB, YUV, HSV. Matching metrics based on image gradients have also been proposed in literature [5] and [6].

Despite many works, it is not clear which image matching metric is the best one, especially in context of various imaging conditions, e.g. represented by some image-set. Currently there is lack of research considering selection of matching metric and its influence on performance of depth estimation. Analysis of this influence is in the focus of this work.

II. SIMLARITY METRIC CALCULATION ARCHITECTURE

Determining similarity of the fragments of an images can be divided into three main stages: features extraction for each pixel of analyzed regions, comparison of determined features between regions in two analyzed views, and aggregation of calculated metric over analyzed region (Fig. 1).

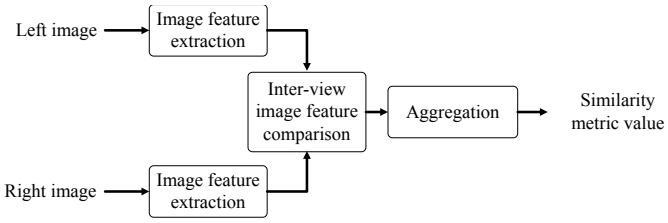

Fig. 1. General architecture of similarity/matching metric calculation

First step in the all metrics known from literature is feature extraction. This step is performed on basis of matched fragment unit, commonly per each pixel. This can be as simple as taking R, B, G component value, or calculating luminance value from R,G,B components. In more complex examples this step involves Rank transform or HoG cascade and gradients determination (e.g. as in SIFT). In the end, for each image fragment (pixel), considered vector of features is created, either based on the fragment itself or based on neighboring pixels. In order to use generalized model, we have divided this step into a following sub steps, from which one can be performed or can be omitted depending on the need.

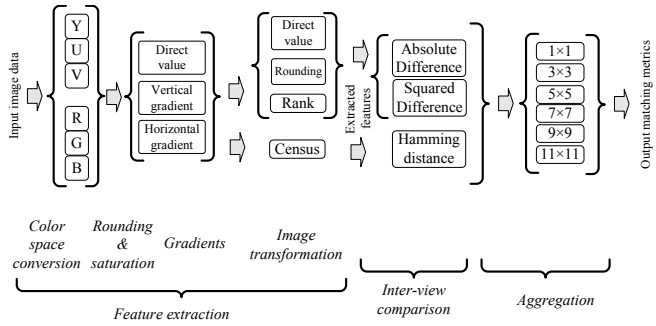

Fig. 2. Architecture of feature extraction.

Feature extraction steps, illustrated in Fig. 2, are as follows:

- Color space conversion - from input color space to color space for further processing, e.g. RGB, YC_bC_r (widely denoted YUV, also in this paper), HSV, L*a*b*, etc.
- Rounding - determining the number of bits per sample (skipping the least significant bits of representation) – especially important for noisy input, or for hardware optimization, e.g. in context of GPU or FPGA/ASIC implementations. In this paper we do not focus on hardware implementations and therefore we do not consider this step in set of analyzed metrics.
- Calculation of horizontal/vertical gradients. This step allows for finer localization and matching of edges of objects. In our work we employ simple pixel-difference.

2021 Signal Processing Symposium (SPSympo) | 978-1-6654-1274-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/SPSYMPO51155.2020.9593880

- Image transformations - Hadamard, Rank, Census, DCT, HoG. It is worth noting that the transform for a given pixel in the image can be determined basing on a larger support window and represent the features assigned to that particular pixel.

After the determination the features vectors (Fig. 2), the next stage is comparison of features of two fragments of images (from two views). Please note that this depends on particular form of given feature, e.g. if it is a number, it can be compared by calculating the absolute difference (AD) of values or squared difference of values (SD). If given feature is a bit vector (as in the case of Census transform), it can be compared by counting the number of different bits (Hamming distance).

Please note, that if the considered image fragment consists of many pixels, this is performed independently for each of them.

After the aforementioned step another saturation process can be employed. If the considered metric value (representing dissimilarity) is too big, it can be safely assumed that the given image fragments do not match. It is pointless to determine precisely how much they don't match, so values above given threshold can be saturated.

The last step of similarity metric calculation (Fig. 2) is aggregation of the obtained matching metric values over the entire fragment, e.g. inside the currently processed block or segment. A simple sum, or the average of the metric value obtained for the pixels belonging to the block or segment can be used. There are also suggestions in the literature for weighing the value of a metrics depending on, for example, the value of the estimated noise level at a given point, the difference in contrast to the center point, or the distance from the center of the block or segment. There are also proposals to use the matching metric value only for selected points, e.g. every second point, to reduce the number of required operations and thus reduce the implementation cost. The size of aggregation window can as small as small as 1×1 to 32×32 and larger.

III. Considered variants of similarity metrics

The amount of all possible similarity metrics, resulting from all possible combinations of operations mention above is horrendous. Therefore, for the sake of conclusiveness, we have limited number of considered metrics in various stages of our research. The overall similarity metrics are as follows:

- Color spaces: RGB and YUV color spaces – each component is processed independently of others, e.g. R, G, B, Y, U, V.
- Direct value of each component and horizontal/vertical gradient of each component.
- Transforms: Direct value, and Rank and Census transforms, in sizes from 3×3 to 11×11.
- Feature comparison: absolute difference (AD) per feature, squared difference (SD) per feature, Hamming distance between features represented as bit-vectors.
- Aggregation in windows of size from 1×1 to 11×11, while not exceeding maximum range of 11×11 input pixels.

For example, for an RGB image stored in a file, the determination of feature vectors is illustrated in Fig. 3. The input image in RGB color space consists of three color components. In the first step input images are converted to YUV color space. As a result for further processing, 6 components are passed: 3 input components (R, G, B) and 3 new components (designated Y, U, V). Next, for each component the vertical and horizontal gradients are being determined. So we obtain, 12 new components (6 gradients in vertical direction, and 6 gradients in horizontal direction)as follows:a component representing the vertical gradient of the red component $\partial_V R$ and the horizontal gradient of the luminance component $\partial_H Y$. In total, 18 components are further processed. The next step is to determine the Rank and Census transformations for all the components determined so far. Five sizes of both Rank and Census transforms are used: from 3×3 to 11×11. After these operations we get 198 components, containing different features of the input image.

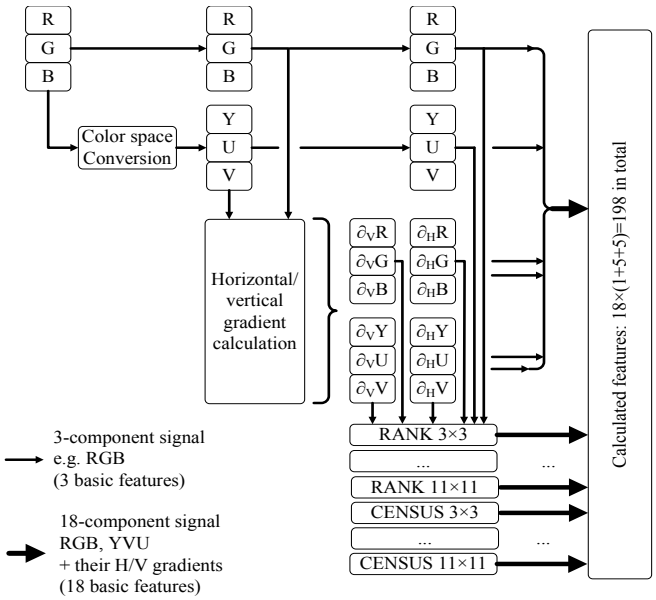

Fig. 3. Considered variants of feature calculation.

Final metrics considered in our research are generated basing on the 198 features presented in Fig 3 with the latter use of inter-view difference calculation (AD/SD/Hamming distance) and/or aggregation (Fig. 3). As mentioned, maximally 11×11 input pixels window size is used. Please note that the input range of pixels depend on both aggregation windows size and the size of transform.

- Example 1: For Rank transform of size 7×7, the following aggregation windows sizes are used 1×1, 3×3, 5×5 (because 5×5 aggregation extends 7×7 transform window to maximal input size of 11×11).
- Example 2: For Absolute Difference metric without any transform (support window of 1×1), all aggregation windows between 1×1 and 11×11 are used.

For all combinations of features and aggregation sizes, in total that would yield in $N = 1026$ similarity metrics. Such a big number is still not feasible for research, especially for statistical covariance analysis, which has $O(N^2)$ computational complexity. Therefore in our research, presented in Section VI, we have worked with selected subsets of these similarity metrics.

IV. COMPOSITE SIMILARITY METRICS

The state-of-the-art suggests that even better results can be achieved if the depth estimation is performed with the use of composite similarity metric, e.g. linear combination of some basic similarity metrics. There is however open research question on how exactly such basic metrics should be combined (and which of them) in order to achieve the best results. Therefore, in our research we also consider application of statistical analysis tools to achieve a composite metric.

We have considered the following analysis and optimization methods:

- Principal Component Analysis (PCA),
- Fisher's linear Discriminant Analysis (FDA) [7],
- Linear Discriminant Analysis (LDA),
- Steepest-descent optimization (StDe).

All of those methods allows to find a linear combination of input basic similarity metrics in order to attain optimized similarity metric.

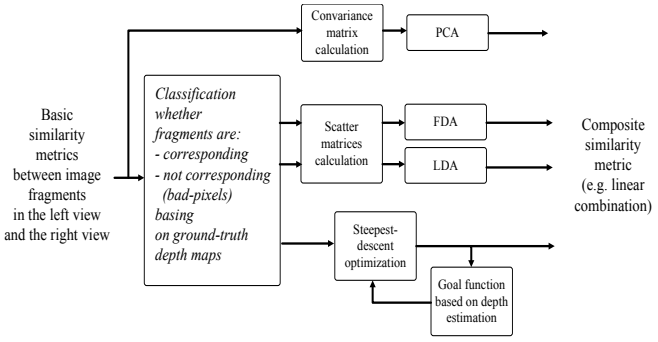

Fig. 4. Composite similarity metric obtaining scheme.

The first choice for our research was Fisher's linear Discriminant Analysis (FDA) and its generalized version: Linear Discriminant Analysis (LDA). Both of these methods rely on classification of pairs of image fragments (in the left view and in the right view) into two classes: corresponding ones, and not corresponding ones (a.k.a. bad-pixels) and thus whether they should be matched in correspondence search, or not (Fig. 4). Therefore these methods require information about true depth values in the dataset and thus ground truth depth maps. This may constitute a disadvantage in a case, where one tries to find optimal metric for data set for which ground-truth depth maps are not known. Both of these methods incorporates calculation of scatter matrices (within-class and between-classes covariance matrices) for all considered metrics. Size of these matrices is $N \times N$ and thus FDA/LDA analysis not feasible for big N values.

Principal Component Analysis (PCA) on the other hand does not require classification of data for corresponding/not corresponding features. Therefore it is applicable for datasets for which ground-truth depth data is not available. It however also incorporates calculation of covariance matrix and for all considered metrics, and moreover, later calculation of its eigen vectors. Therefore, this method is also not feasible for big number N of analyzed similarity metrics.

The last method that we have considered is classical Steepest-descent optimization. This method does not incorporate calculation of covariance/scatter matrices and thus is more suitable for big N values. It however employs iterative calculation of the goal function which in the case considered in the research is related to performing depth estimation with some particular algorithm. Therefore results of this method are dependent on the selection of depth estimation algorithm employed for the optimization.

The dataset for analysis was constructed on the foundation of Middlebury image database [8], widely used in literature related to research on depth estimation:

- The color images for two views have been used to generate all possible correspondence pairs of image fragments (in the left and in the right view).
- Between those fragments, considered basic similarity metrics have been calculated.
- Classification was done basing on reference depth maps in Middlebury image set.

V. QUALITY ASSESSMENT

In order to assess the quality of the considered similarity metrics in a wide context (e.g. GPU/hardware implementation, where complexity of estimation algorithms has to be compromised), a very simple, greedy depth estimation algorithm was used: Winner Takes All (WTA). In WTA algorithm, the disparity is estimated locally, solely basing on values of similarity metric. Therefore, there are no other factors (e.g. regularization/optimization algorithm) that could distort the result of the experiments.

To assess the quality of the determined depth map, a measurement of the number of points for which the disparity values were incorrectly determined according to the "Bad pixels" measure [1]. “Bad pixels ±1” variant has been used: given disparity can differ maximally by 1 in order to be classified as correct pixel. Other pixels are classified as bad.

In the Middlebury test image database, each stereoscopic pair is accompanied by a reference depth map, which is a benchmark data for the performance of depth estimation algorithm. The percentage of points with incorrect disparity value in the estimated depth map is evaluated in comparison with the reference depth map, excluding points for which it is impossible to determine the correct value of the disparity by definition, e.g. image's extreme points or occluded areas in the scene.

VI. EXPERIMENTAL RESULTS

We have performed four experiments with different selection of input similarity metrics in order to increase conclusiveness.

A. Color space

In this experiment we have considered which color space is more beneficial: RGB or YUV. Therefore we have not considered transforms (RANK/CENSUS) or gradients, and only considered sum of absolute differences (AD) with various aggregation window sizes. The results are presented in Table I. For the sake of brevity, we omit rows which were insignificant, e.g. most rows related to red (R) and green (G) components.

TABLE I. RESULTS OF EXPERIMENT WITH COLORSPACES: BAD-PIXEL RATIO RESULTS FOR SIMILARITY METRICS, AVERAGED OVER 27 MIDDLEBIRY TEST IMAGES AND OPTIMIZED WEIGHTS FOR STDE OPTIMIZED SIMILARITY METRIC.

Similarity metric		**Bad pixels (±1 threshold)**	**Optimized Weight (StDe)**
R	AD	86.81%	0.01
	SAD 11×11	42.21%	0.01
G	AD	85.56%	0.03
	SAD 11×11	39.19%	0.00
B	AD	85.20%	0.01
	SAD 3×3	52.50%	0.06
	SAD 5×5	42.97%	0.03
	SAD 7×7	39.19%	0.00
	SAD 9×9	37.29%	0.00
	SAD 11×11	**36.28%**	**0.01**
Y	AD	85.72%	0.00
	SAD 3×3	51.88%	0.00
	SAD 5×5	43.04%	0.00
	SAD 7×7	39.28%	0.00
	SAD 9×9	37.44%	0.00
	SAD 11×11	36.48%	0.14
U	AD	88.95%	0.02
	SAD 3×3	61.72%	0.01
	SAD 5×5	51.30%	0.03
	SAD 7×7	46.38%	0.04
	SAD 9×9	43.47%	0.01
	SAD 11×11	41.59%	0.17
V	AD	89.11%	0.00
	SAD 3×3	62.16%	0.01
	SAD 5×5	52.23%	0.07
	SAD 7×7	46.89%	0.04
	SAD 9×9	43.76%	0.03
	SAD 11×11	41.84%	0.22
	Optimized (PCA)	**36.91%**	-
	Optimized (FDA)	**37.09%**	-
	Optimized (LDA)	**36.01%**	-
	Optimized (StDe)	**34.03%**	-

As it can be seen, application of stand-alone similarity input metrics gives the best results for Sum of Absolute Differences (SAD) aggregated in windows of size 11×11 for blue (B) component, which is slightly better than for Y component (36.48%). This results of course depends on the particular used Middlebury dataset and for different dataset can be different. Therefore, as for this dataset ground-truth data is available, we could just use blue component and attain quite good results. The big windows size is not surprising as typically moderately big windows aggregation sized yield with better quality depth maps.

More interestingly, what we can see is that usage of PCA (without knowledge about ground-truth data) allows us to constitute a composite optimized metric which performs only slightly worse (36.91% of bad pixels) than the best performing metric.

Usage of Fisher's discriminant Analysis (FDA), performed with the usage of ground-truth data, in this case yields even worse results, but the usage of Linear Discriminant Analysis (LDA) provides interesting improvement to about 36.01% of bad pixels. Usage of Steepest-descent optimization algorithm gives even better results: 34.03% of bad-pixels. As compared to the best performing stand-alone input similarity metric (SAD 11×11 Blue, 36.28%), the gain is of about 2 percent points.

In the right column of Table I it can be observed that in fact the optimized StDe composite metric consists mostly of three components: Y, U and V, all using 11×11 aggregation window. This means that it is optimal to use YUV color space, even though, in the case of challenge of stand-alone components, similarity metric based on blue (B) was the best performing.

Therefore in the further research, we have focused on YUV color space.

B. Sum of Differences: Absolute or Squared?

In second experiment we have considered which inter-view feature comparison method performs better: sum of absolute differences (AD) or sum of squared differences (SD. The results are presented in Table II.

As it can be seen, results attained for Absolute Differences (AD) are slightly better than those attained with the use of Squared Differences (SD). Also, the best performing stand-alone similarity metric is based on AD with 11×11 aggregation window: 36.48%.

Moreover it can be seen that also in the considered experiment, usage of proposed optimization techniques leads to improvement of results. Application of PCA (which does not requires ground-truth depth maps) allows for improvement to 35.27% of bad pixels. Noticeably, this is better results than usage of FDA (which requires ground-truth data). Even better results (35.17% of bad pixels) can be attained when LDA is used. The best results again has been attained with the use of Steepest-descent optimization algorithm (StDe) which yielded 34.49% of bad pixels. It can be noted that the result here is slightly worse than in the previous experiment, presumably because other components (R,G,B) were not available to optimization algorithms.

C. Gradients.

In the third experiment we have included gradients in our considerations. For the sake of brevity we present only the most significant rows in Table III. It can be seen that usage of even one stand-alone metric based on gradients (specifically, vertical gradient of luminance component: Y ∂_V, aggregated in 11×11 window) provides the best performing results so far: 30.43 of bad pixels. Noticeably, worse results are attained when most of analysis and optimization methods are used: PCA, FDA and LDA, correspondingly 32.93%, 36.07%, and 32.54% of bad pixels. It can be however observed, that those of PCA and LDA are in fact better that those attained in previous experiments.

The most important observation is however that the best results, again, can be attained with the use of Steepest-descent (StDe) optimization algorithm, which yields with only 27.36% of bad pixels. The mixture of input similarity metrics found by StDe algorithm mostly consists of the mentioned vertical gradient of luminance component Y ∂_V aggregated in 11×11 window, but also contains YVU components without gradients, and small shares of gradients of chrominance components (UV).

TABLE II. RESULTS OF EXPERIMENT WITH COMPARISON METHOD: BAD-PIXEL RATIO RESULTS FOR SIMILARITY METRICS, AVERAGED OVER 27 MIDDLEBIRY TEST IMAGES AND OPTIMIZED WEIGHTS FOR STDE OPTIMIZED SIMILARITY METRIC.

Similarity metric			Bad pixels (±1 threshold)	Optimized Weight (StDe)
Y	SD	1×1	85.72%	0.00
		3×3	53.91%	0.00
		5×5	45.67%	0.00
		7×7	42.12%	0.00
		9×9	40.21%	0.00
		11×11	39.13%	0.00
Y	AD	1×1	85.72%	0.00
		3×3	51.88%	0.00
		5×5	43.04%	0.02
		7×7	39.28%	0.00
		9×9	37.44%	0.00
		11×11	**36.48%**	**0.11**
U	SD	1×1	88.95%	0.00
		3×3	61.72%	0.01
		5×5	51.30%	0.00
		7×7	46.38%	0.00
		9×9	43.47%	0.00
		11×11	41.59%	0.01
U	AD	1×1	88.95%	0.00
		3×3	60.59%	0.00
		5×5	50.11%	0.00
		7×7	45.20%	0.01
		9×9	42.32%	0.03
		11×11	40.54%	0.26
V	SD	1×1	89.11%	0.00
		3×3	62.16%	0.00
		5×5	52.23%	0.00
		7×7	46.89%	0.00
		9×9	43.76%	0.01
		11×11	41.84%	0.01
V	AD	1×1	89.11%	0.02
		3×3	61.34%	0.00
		5×5	51.17%	0.02
		7×7	45.80%	0.04
		9×9	42.77%	0.10
		11×11	41.02%	0.35
		Optimized (PCA)	**35.27%**	-
		Optimized (FDA)	**36.15%**	-
		Optimized (LDA)	**35.17%**	-
		Optimized (StDe)	**34.49%**	-

D. Rank and Census transforms with Gradients

In the final third experiment we have employed Rank and Census Transforms, in sizes ranging from 3×3 to 11×11, and with aggregation in windows of size from 1×1 to 11×11, while not exceeding maximum range of 11×11 input pixels. For the sake of brevity in Table IV we present only the most significant rows: mostly related to Census and Rank transforms for luminance (Y) component.

There are few interesting thinks to be notices. Firstly, the results attained by higher-order Census transforms (windows size with aggregation between 9×9 and 11×11 between are in range of about 26-41% and mostly below 30% of bad pixels. The best stand-alone results are attained for census transform of size 5×5 with aggregation window 7×7, resulting in 11×11 total input window: 26.19%. This metric in fact outperforms any other metric considered in the abovementioned research – even those related to optimized metrics attained with Steepest-descent (StDe) algorithm.

TABLE III. RESULTS OF EXPERIMENT WITH GRADIENTS: BAD-PIXEL RATIO RESULTS FOR SIMILARITY METRICS, AVERAGED OVER 27 MIDDLEBIRY TEST IMAGES AND OPTIMIZED WEIGHTS FOR STDE OPTIMIZED SIMILARITY METRIC.

Similarity metric			Bad pixels (±1 threshold)	Optimized Weight (StDe)
Y	-	1×1	85.72	0.07
		3×3	51.88	0.00
		5×5	43.04	0.00
		7×7	39.28	0.00
		9×9	37.44	0.00
		11×11	36.48	0.00
	∂_H	1×1	94.33	0.00
		3×3	55.96	0.02
		5×5	42.31	0.02
		7×7	37.22	0.02
		9×9	35.11	0.07
		11×11	34.19	0.07
	∂_V	1×1	92.99	0.03
		3×3	48.91	0.11
		5×5	37.42	0.07
		9×9	31.39	0.02
		11×11	**30.43**	**0.22**
U	-	11×11	41.59	0.02
	∂_H	1×1	97.27	0.01
		3×3	97.27	0.01
	∂_V	3×3	73.25	0.02
		5×5	61.42	0.02
V	-	3×3	61.34	0.04
	-	5×5	51.17	0.02
	∂_H	3×3	77.75	0.01
		Optimized (PCA)	**32.93%**	-
		Optimized (FDA)	**36.07%**	-
		Optimized (LDA)	**32.52%**	-
		Optimized (StDe)	**27.36%**	-

Slightly worse (for about 2-3 percent points) results are attained with Rank transform, which also peaks at 30.53% and pixels in the case of transform size 5×5 and with aggregation window of size 7×7.

Another important observation is that in this experiment, statistical-based optimization methods like PCA, FDA and LDA were unable to find better optimized similarity metric. Their bad pixels results are correspondingly 31.67%, 30.74%, and 30.56%, which is about 4 percent point worse than results attained by census transform used stand-alone.

The last observation is that just like in previous cases, a significant improvement can be attained with the use of Steepest-descent (StDe) optimization algorithm. It allows to improve bad-pixel-radio for about 1.5 percent points as related to the best-performing census transform. Noticeably, the result attained with similarity metric optimized with the use of StDe (25.04%) is coming mostly from usage of the best performing census transform (weight 0.4). The remaining share of the optimized transform belongs to census transform calculated over vertical gradients (weight about 0.04) and classical sum of absolute differences (SAD) with window size 7×7 (weight also about 0.04). It can be noticed that this window size is considerably smaller that window sizes selected by optimization in previous experiments (typically 11×11).

TABLE IV. BAD-PIXEL RATIO RESULTS FOR SIMILARITY METRICS, AND OPTIMIZED WEIGHTS FOR STDE OPTIMIZED SIMILARITY METRIC

Similarity metric				Bad pixels (±1 threshold)	Optimized Weight (StDe)
Gradient	**Transform**	**Aggregation**	**Input window**		
-	none (SAD)	1×1	1×1	85.72%	0.00
		3×3	3×3	51.88%	0.00
		5×5	5×5	43.04%	0.00
		7×7	7×7	39.28%	0.04
		9×9	9×9	37.44%	0.00
		11×11	11×11	36.48%	0.00
	rank 3×3	1×1	3×3	97.51%	0.00
		3×3	5×5	58.36%	0.00
		5×5	7×7	41.51%	0.00
		7×7	9×9	35.01%	0.00
		9×9	11×11	32.02%	0.00
	rank 5×5	1×1	5×5	93.52%	0.00
		3×3	7×7	49.97%	0.00
		5×5	9×9	35.52%	0.00
		7×7	11×11	30.53%	0.00
	rank 7×7	1×1	7×7	91.28%	0.00
		3×3	9×9	47.04%	0.00
		5×5	11×11	33.82%	0.00
	rank 9×9	1×1	9×9	89.99%	0.00
		3×3	11×11	45.55%	0.00
	rank 11×11	1×1	11×11	89.14%	0.00
	census 3×3	1×1	3×3	80.90%	0.00
		3×3	5×5	43.06%	0.00
		5×5	7×7	32.19%	0.00
		7×7	9×9	28.30%	0.00
		9×9	11×11	26.67%	0.00
	census 5×5	1×1	5×5	59.00%	0.00
		3×3	7×7	34.92%	0.00
		5×5	9×9	28.69%	0.00
		7×7	**11×11**	**26.19%**	**0.43**
	census 7×7	1×1	7×7	47.56%	0.00
		3×3	9×9	31.38%	0.00
		5×5	11×11	27.26%	0.00
	census 9×9	1×1	9×9	41.69%	0.00
		3×3	11×11	29.63%	0.13
	census 11×11	1×1	11×11	38.38%	0.00
∂_V	census 3×3	1×1	3×3	84.48%	0.00
		3×3	5×5	50.80%	0.00
		5×5	7×7	39.88%	0.00
		7×7	9×9	35.15%	0.00
		9×9	11×11	32.52%	0.04
	census 5×5	1×1	5×5	63.84%	0.00
		3×3	7×7	42.24%	0.00
		5×5	9×9	35.71%	0.00
		7×7	11×11	32.26%	0.04
	census 7×7	1×1	7×7	52.93%	0.00
		3×3	9×9	38.60%	0.00
		5×5	11×11	33.95%	0.00
	census 9×9	1×1	9×9	47.37%	0.00
		3×3	11×11	36.84%	0.00
	census 11×11	1×1	11×11	44.22%	0.00
	Optimized (PCA)			**31.67%**	-
	Optimized (FDA)			**30.74%**	-
	Optimized (LDA)			**30.56%**	-
	Optimized (StDe)			**25.04%**	-

VII. CONCLUSSIONS.

In the paper we have presented an organized taxonomy of calculation of similarity metrics. We have evaluated them with the use of Middlebury test image dataset. The attained results show the importance of selection of similarity metric, The differences between various metrics range up to 60 percent points of bad-pixel depth map quality ratio. Also, the considered basic similarity metrics have been also fed as inputs for statistical analysis tools and optimization algorithm in order to attain composite, optimized similarity metric.

It has been also shown that usage of attained optimized similarity metrics can lead to improvement of quality of depth estimation, expressed by drop of bad-pixel ratio. I.e. usage of composite similarity metric resulting from Steepest-descent optimization allows to improvement of bad-pixel ratio for about 1 to 4 percent points as related to the best performing stand-alone similarity metric. This however requires and employs the knowledge about the target depth estimation algorithm and ground truth depth maps.

When target depth estimation algorithm is not known a priori the optimization of similarity metric, it has been shown, that statistical analysis like FDA or LDA can be used. Those techniques in some of the experiments allowed to improve bad-pixel ratio by about 1.5 percent points. It must be however noted that in the experiment employing Census and Rank transforms, those techniques failed to provide improved compound similarity metric.

It has been also noticed that although usage of Principal Component Analysis (PCA) does not lead to any significant improvements regarding bad-pixel ratio, its advantage is that it does not require a priori knowledge about target depth estimation algorithm nor ground truth depth maps. Therefore this technique can be used to find semi-optimal composite similarity metric in cases of datasets for which ground-truth depth maps are not available.

Finally, the overall conclusions about the source of attained improvements mainly lays in usage of Census Transform, which should be performed in windows size of about 7×7 to 9×9 and then additionally aggregated in window 3×3 to 5×5. This particular results is novel and very interesting as, typically, no aggregation over Census/Rank transforms is suggested in the literature.

ACKNOWLEDGMENT

Project funded by The National Centre for Research and Development in the LIDER Programme (LIDER/34/0177/L8/16/NCBR/2017).

REFERENCES

[1] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," 2006 IEEE Computer Society Conf. on Comp. Vision and Pattern Recogn. (CVPR'06), New York, NY, USA, 2006, pp. 519-528.

[2] O. Johannsen et al., "A Taxonomy and Evaluation of Dense Light Field Depth Estimation Algorithms," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 2017, pp. 1795-1812 226.

[3] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in European Conference on Computer Vision (ECCV). Springer, 1994.

[4] C. Stentoumis, E. Karkalou and G. Karras, "A review and evaluation of penalty functions for Semi-Global Matching," 2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2015, pp. 167-172.

[5] K. Wegner, O. Stankiewicz, "Similarity measures for depth estimation." 2009 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video. IEEE, 2009.

[6] I. Jung, J. Sim, C. Kim and S. Lee, "Robust stereo matching under radiometric variations based on cumulative distributions of gradients," 2013 IEEE International Conference on Image Processing, Melbourne, VIC, Australia, 2013, pp. 2082-2085.

[7] S. V. Sidorov,N. Y. Zolotykh, "Linear and Fisher Separability of Random Points in the d-dimensional Spherical Layer," 2020 Int. Joint Conf. on Neural Networks (IJCNN), Glasgow, UK, 2020, pp. 1-6.

[8] "Middlebury Webpage", http://vision.middlebury.edu/stereo/.