Investigation on the separability of slums by multi-aspect TerraSAR-X dual-co-polarized high resolution spotlight images based on the multi-scale evaluation of local distributions

Andreas Schmitt, Tobias Sieg, Michael Wurm, Hannes Taubenböck

Abstract— **Following recent advances in distinguishing settlements vs. non-settlement areas from latest SAR data, the question arises whether a further automatic intra-urban delineation and characterization of different structural types is possible. This paper studies the appearance of the structural type "slums" in high resolution SAR images. Geocoded Kennaugh elements are used as backscatter information and Schmittlet indices as descriptor of local texture. Three cities with a significant share of slums (Cape Town, Manila, Mumbai) are chosen as test sites. These are imaged by TerraSAR-X in the dual-co-polarized high resolution spotlight mode in any available aspect angle. Representative distributions are estimated and fused by a robust approach. Our observations identify a high similarity of slums throughout all three test sites. The derived similarity maps are validated with reference data sets from visual interpretation and ground truth. The final validation strategy is based on completeness and correctness versus other classes in relation to the similarity. High accuracies (up to 87%) in identifying morphologic slums are reached for Cape Town. For Manila (up to 60%) and Mumbai (up to 54%), the distinction is more difficult due to their complex structural configuration. Concluding, high resolution SAR data can be suitable to automatically trace potential locations of slums. Polarimetric information and the incidence angle seem to have a negligible impact on the results whereas the intensity patterns and the passing direction of the satellite are playing a key role. Hence, the combination of intensity images (brightness) acquired from ascending and descending orbits together with Schmittlet indices (spatial pattern) promises best results. The transfer from the automatically recognized physical similarity to the semantic interpretation remains challenging.**

Key words — **Radar applications; Radar polarimetry; Radar remote sensing; Image analysis; Image classification; Image fusion; Image texture analysis; Pattern recognition; Urban areas.**

I. INTRODUCTION

A. Thematic introduction

About 827 million people across the globe live in conditions that UN-HABITAT [1] classifies as slums. Future projections suggest that by 2050 about two billion people will additionally live in self-built neighborhoods resulting from informal occupation and construction of precarious environments [1], [2]. The magnitude of these numbers reveals that informal city building is not a fringe phenomenon. Still, conventional data sources such as the census, maps or geodata on such locations are often outdated, incomplete, not reliable or missing completely. In consequence, spatiotemporal information on slums is scarce at the city scale. A global or at least large area inventory of informal settlements locations, extents or their morphology is inexistent [3].

Consequently, the numbers on slum population presented above are of high uncertainty [4].

Remote sensing is able to provide the only consistent and area-wide data set that allows realizing local, regional or global knowledge on slum locations and structural characteristics. Unfortunately, the only indicator from UN-Habitats slum definition allowing the usage of earth observation data is the "durability of housing conditions". Hence, recent studies try to relate this indicator to physical information present in EO-data in order to develop appropriate target classification approaches [3]. The most frequently used characteristics are high building density, small building sizes, and irregular, organic patterns [5], [6], [7]. A survey of current literature on classification techniques reveals that predominantly very high resolution optical satellite images have been exploited in slum mapping. Spectral as well as textural features within object-based image analysis in knowledge-based frameworks are dominating the developments [8], [9], [10], [11], [12], [13], [14], [15]. Studies using synthetic aperture radar (SAR) techniques for the detection of morphologic slums are rare. SAR data are of high relevance in this domain because of their capability to provide a stable data set independent from atmospheric conditions or day-time. This is an important advantage compared to data sets acquired by optical sensors especially with regard to tropical regions in which many of cities with a large share of slum structures are located. Furthermore, the detection of urban settlements by means of polarimetric SAR acquisitions is shown to be possible [16], [17]. However, the discrimination of different settlement types by SAR data sets is still challenging. The texture of the acquired images is a key feature for the separation between different settlement types due to the complexity of urban environments. The fundamental problem in the description of texture is the great variety of different texture measures from simple local statistics [18] to grey-level co-occurrence matrices [19] and their interpretative derivations [20]. Former studies search for the optimal descriptor, i.e. the (derived) SAR feature that shows the highest correlation with building density [21]. While former classification algorithms were designed for the use of a low number of distinct layers, recent studies try to include the whole entity of multifaceted descriptors by using self-learning approaches like the random forest method for classification of slums [22]. Though the classification from hundreds of feature layers is not a problem any longer, the calculation of those layers still is time and memory consuming. Hence, simple approaches are preferred for the classification of large areas, e.g. the global urban footprint [23]. Therefore, this study investigates the characteristics of slums in high resolution SAR images by their radiometric and spatial signature to distinguish between different settlement types. The data sets for this study are acquired by TerraSAR-X in the dual-co-polarized high resolution spotlight mode. Upcoming SAR satellite missions like the "Radarsat Constellation" [24] or the "High Resolution Wide Swath" mission [25] will probably be able to provide suitable image data with comparable radiometric resolution, but a much larger coverage.

B. Methodological introduction

TerraSAR-X images show a very high spatial resolution up to 1 m though imaged from space. A large set of measurements hence is available per point of interest. Hence the number of pixels within an area of interest is much higher compared to former SAR sensors with a lower spatial resolution of e.g. 30 meters. The mapping unit of one hectare formerly covered by roughly eleven pixels – which are necessary for the initial multi-looking

prior to the polarimetric decomposition – contains about 10,000 pixels when acquired by TerraSAR-X. Consequently, the traditional pixel-based estimation and classification of the backscattering mechanisms by very sophisticated polarimetric decompositions [26] is not appropriate for these data sets. Thus, it is reasonable to investigate agglomerations of measurements in the range of the mapping unit instead of dealing with isolated pixels.

Segment-based classification [27] represents one possible approach for this investigation introducing geometric features, as e.g. shape, size and context parameters of the segments. This requires a high quality of the initial segmentation which is challenging with respect to SAR data [28]. The segmentation algorithms applied to optical data assume a similarity of pixels which belong to the same segment and therewith, representing one object. In contrast, an object imaged by SAR generally appears as a collection of different backscattering effects, e.g. a residential house is characterized by the overlay of façade and forecourt, the roof, and the shadow behind. Even if highly accurate segment geometries from cadastral data (e.g. block units) are used, the imaging effects of SAR still impede a perfect match. Thus, we introduce a region-based evaluation as middle course between the pixel-based and the segment-based approach. The region-based strategy is partly pixel-based since the evaluation is performed at any position in the image and partly segment-based because a locally variable environment around the corresponding pixel is considered.

There are mainly two ways to investigate this large entity of measurements per evaluation point. First, a theoretical distribution function is adopted and its parameters are estimated from the samples [29]. In the simplest case of a normal distribution, the number of parameters reduces to the mean and the standard deviation. With respect to SAR images of urban environments the definition of an encompassing distribution, i.e. one statistical distribution that is valid for all settlement types, is not feasible, because different settlement types actually show different kinds of distribution [30]. Hence, the theoretical distribution is commonly chosen from a distribution family and adapted to the local image [31]. Second, the empirical distribution is accepted as it is and evaluated by so-called non-parametric approaches [32]. The latter approach is free from assumptions on the underlying statistics. In general, the computation of such algorithms is very expensive in terms of memory and time because of the usually high sampling rate which is required. The empirical probability density is commonly expressed by histograms. At this point, data preparation can help to reduce the number of bins of the histograms, i.e. the radiometric sampling rate, from several hundreds to just a few by a sophisticated and consistent scaling of the input data. The computing demand thus decreases considerably and the implementation becomes feasible with view to practical applications for the first time. One possible solution is provided by the TANH scaling implemented in the MultiSAR framework of DLR which will be introduced in the following and used as preprocessing environment for our SAR data sets.
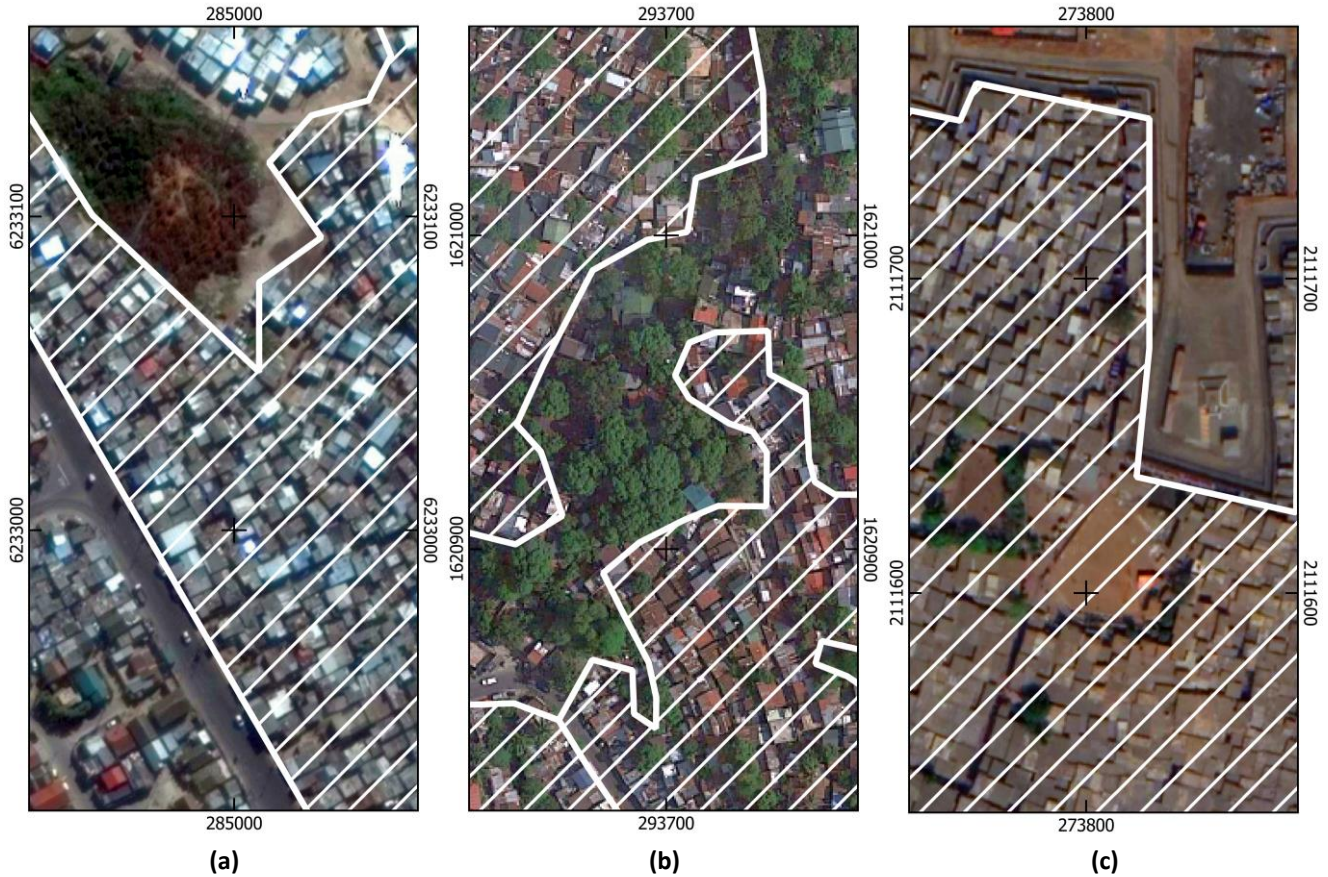
II. STUDY AREA AND DATA SET

Fig. 1. Exemplary maps of the slum segments (white diagonal hatch) for (a) Cape Town (UTM zone 34S), (b) Manila (UTM zone 51N), and (c) Mumbai (UTM zone 43N) with optical image (©GoogleEarth 2017) in the background. These subsets illustrate the complexity of deriving reliable and complete reference data sets for training and validation purposes.

A. Reference data

For our experiments, we select three cities – Cape Town (South Africa), Mumbai (India), and Manila (The Philippines) – because of the following reasons: First, all three cities contain a significant share of slums, e.g. more than 50% of the population lives in slums in Manila. Morphologically these morphologic target areas are in line with the typically physical characteristics (small building extents, complex alignment of buildings, etc.) identified by an expert group [33] as well as with the enhanced ontology presented by [34]. Second, all three cities are very large cities featuring a diverse mixture of different structural types across the city, but at the same time the structural configuration is varying across these cities. The Cape Town test area comprises Khayelitsha which represents a mixture of formal settlements, townships (generally planned settlements, but with very low living conditions), and slums. Further land cover classes in the image are bare soil and vegetation. The Manila test site is characterized by high-rise buildings in the central business district, a large amount of formal residential buildings, the harbor zone with industrial buildings, and several rather small-sized slums. Water and park areas like the Manila Cemetery are the main non-settlements classes. The Mumbai test ground is a very diverse landscape consisting of water, bog, grassland, bare soil, railroads, highways, the airport, and settlements. The built-up area is composed out of chemical factories, the central business district, the Mumbai university campus, formal residential buildings, and large slums like Dharavi.

Thus, this setting allows evaluating transferability of the methodology across structural types of cities. Third, the availability of reference information on slum locations and extents is very scarce which limits the possible selection for test areas. Thus, we take advantage of classifications using EO-data from previous studies delineating slums [35]. This discrimination of morphologic slums is based on manual classification by an expert and thus, the reference data feature a very high accuracy. Fig. 1 shows one subset for each test site that demonstrates the problems in discriminating slums from other settlements types and/or land cover classes. From these reference data about a third of the slum segments (929 overall - 224 for Cape Town, 368 for Manila, and 337 for Mumbai) have been randomly chosen as training samples. The random selection ensures a spatial distribution of training samples across the entire city and across slightly varying morphologic characteristics. Only segments with a minimum size of one hectare out of a larger collection are considered for stability reasons. This complies with a minimum patch size of approximately 100-by-100 pixels for the training areas, i.e. each sample histogram is composed out of at least 10,000 measurements.

B. Remotely sensed data

The very high resolution SAR images are acquired by the TerraSAR-X mission in the high resolution spotlight mode. Each image covers an area of about 5 km by 10 km. Four different images have been taken of each test site: two images in ascending mode and two images in descending mode both with a steep and a flat incidence angle respectively. Two polarizations are measured at the same time, namely HH and VV, with a stable phase reference, so that the phase difference between the two co-polarized channels can be evaluated. The co-polarized phase difference is essential for the discrimination of odd from even bounce effects [36]. All images are transformed into Kennaugh elements, geocoded, calibrated, speckle-filtered, and finally normalized [37]. Details on the acquisition time, the geometry, and the sampling rate resulting from the uniform pixel spacing on ground and the varying incidence angle are given in Table I.

Polarimetric SAR data are delivered in linear scale and commonly processed to the coherency or the covariance matrix still in slant range geometry. The main diagonal elements are real intensity values in the range $[0, \infty[$ whereas the off-diagonal elements are complex values ranging in-between $] - \infty, \infty[$ for the real and imaginary part or $[0, \infty[$ for the amplitude and $[0, 2\pi[$ for the phase angle. As the vast majority of intensity measurements can be found in the range $]0, 0.5[$, a high sampling rate is necessary resulting in thousands of bins. The conversion to decibel projects the intensity values onto a range of about $\pm 30 dB$ in general [30]. The sampling can often be lowered down to around one hundred bins, but the theoretical range still is unlimited. In this context, the Kennaugh framework [37] provides the sophisticated hyperbolic tangent (TANH) normalization guaranteeing a closed value range $] - 1, +1[$ which allows for reducing the sampling to only 21 bins, i.e. $[- 1.0, - 0.9, \ldots, - 0.1, 0.0, 0.1, \ldots, 0.9, 1.0]$, or even less, without significant loss of information. Following layers are available for dual-co-polarized images: k_0 (total intensity, amplified by 7dB), k_3 (relation between even-bounce and odd-bounce, e.g. diplane scattering vs. surface scattering), k_4 (relation between HH and VV polarized intensity, e.g. horizontally oriented dipoles vs. vertically oriented dipoles), and k_7 (phase correlation of the co-pol channels) [38], see Figs. 4-6. It is worth to mention that the Kennaugh elements are nothing else than smart

linear combinations of the covariance matrix elements and therewith very fast to compute.

TerraSAR-X high resolution spotlight scenes in dual-co-polarization HH&VV over the three test sites. All images are geocoded to the UTM coordinate frame with a uniform pixel spacing of 1 m. The look factor results from the original resolution in slant range, the targeted pixel spacing on ground, and the incidence angle. A slight oversampling is indicated and accepted.

Test site	Date	Time	Orbit direction	Incidence angle	Look factor
Cape Town	2015-07-27	17:30:18	ascending	50°	0.46
	2015-07-28	04:05:21	descending	45°	0.42
	2015-08-02	04:13:52	descending	30°	0.30
	2015-08-02	17:21:49	ascending	37°	0.36
Manila	2016-02-20	10:06:17	ascending	45°	0.44
	2016-02-26	09:57:46	ascending	27°	0.27
	2016-02-26	21:41:43	descending	54°	0.49
	2016-05-18	21:50:19	descending	39°	0.38
Mumbai	2015-07-21	01:07:21	descending	34°	0.34
	2015-07-27	00:58:50	descending	50°	0.48
	2016-05-18	13:17:13	ascending	46°	0.44
	2016-06-26	13:08:43	ascending	29°	0.28

In the speckle reduction step, the image is filtered by a bank of round-shaped as well as longitudinal kernels (the "Schmittlets") in different scales and azimuth orientations, see Fig. 2. The best-fitting Schmittlet is chosen locally before image reconstruction [39]. Thus, the index image of the best fitting Schmittlets S_b provides a valuable descriptor of texture indicating the ellipticity, size, and absolute orientation of the underlying spatial patterns, see Figs. 4-6. The color thus refers to the azimuth orientation and the brightness to the scale, e.g., bright greenish structures indicate targets with an estimated length of about 16 m and an azimuth orientation of about 45° towards the north direction. Grey tones refer to round-shaped and undirected Schmittlets, e.g. medium grey indicates a circular kernel with a diameter of about 8 m. The Schmittlet image enhancement consists of the convolution of the original intensity image with each of the 35 Schmittlet kernels and the comparison to the original image. Although the convolution is computationally intensive, it can simply be accelerated by parallel computing.
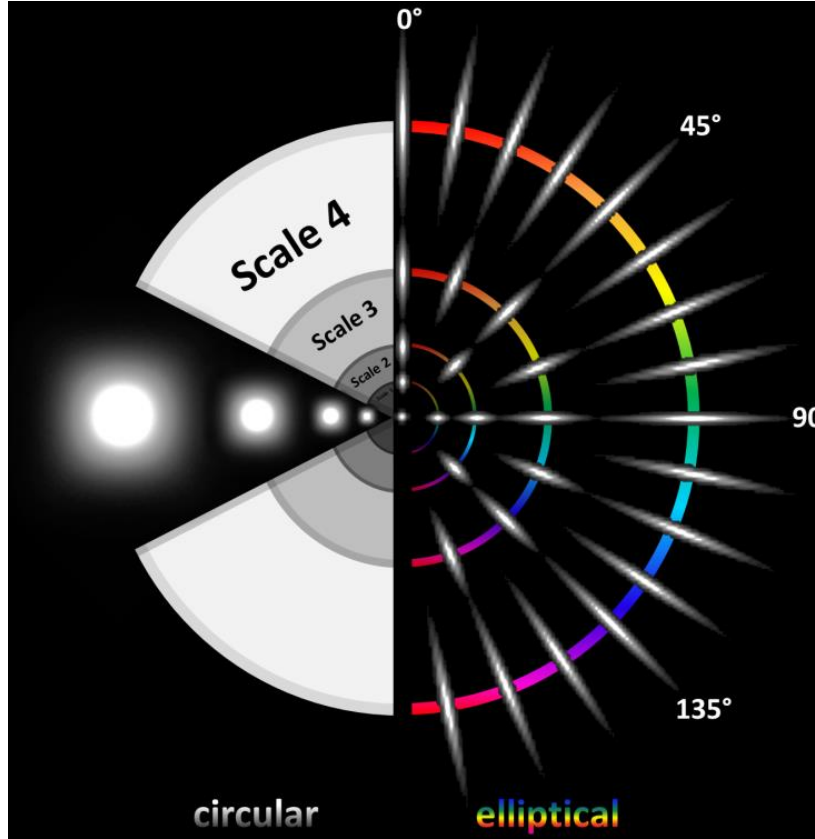
Fig. 2. The 35 Schmittlets of the current implementation [39] according to their characteristics: circular shape (left) or elliptical shape (right), scale from zero to four (concentric circles), and azimuth orientation from 0° to 180° (azimuth of the elongated Schmittlets on the right hand site).

III. METHODOLOGY

The fundamental idea of this study is to describe the image content by the help of representative empirical distributions for slums and to identify similar local distributions in the SAR image. For this purpose, intensity (k_0), polarimetric (k_3, k_4, k_7) and texture layers (S_i) are systematically tested as possible input. The respective distributions are described by their probability density functions (PDF) denoted by $p(\mathrm{x})$ for a specific variable x which are characterized by two properties

$$p(\mathrm{x}) \geq 0 \text{ for } \mathrm{x} \in \mathbb{R} \quad \text{and} \quad \int_{-\infty}^{+\infty} p(\mathrm{x}) \, \mathrm{dx} = 1. \tag{1}$$

The same applies to a two-dimensional probability function like $p_c(x,t)$ as well. Consequently, $p(x)$ and $p_c(x,\mathrm{t})$ both have no negative values and the integral over the whole value range of x (and t) reaches one. The interval $]-\infty, +\infty[$ is omitted in the following for clarity reasons. Three types of PDFs $p(\mathrm{x})$ will be addressed in the following: $p_c(x,t)$ denotes an empirical PDF of the training samples t for a certain class c, $p_c(x)$ stands for an estimated representative PDF of class c derived from $p_c(x,t)$, and $p_k(x)$ describes the local PDF of an pixel environment defined by filter kernel f_k.

All variables and functions are listed in Table II for convenience.

A. PDF estimation

1) Estimation of a representative PDF from a collection of samples

In the ideal case, the representative PDF of a certain class $p_c(x)$ is reached by marginalizing the two-dimensional PDF of the samples $p_c(x, t)$ over the t dimension.

$$p_{c,0}(x) = \int p_c(x, t) \, dy. \tag{2}$$

This implies that all training samples t are equally weighted. As shown in Fig. 1 even reference data are not always reliable, a weighting function $w_{i+1}(t)$ is introduced in order to reduce the impact of possible outliers in the training samples. The reliability of the respective training site $w(t)$ is initialized uniformly as $w_0(t) = 1$ (equal weights) and adapted iteratively.

$$w_{i+1}(t) = \left\{ \int \frac{p_c^2(x,t)}{p_{c,i}(x)} dx \right\}^{-1} \in [0,1]. \tag{3}$$

The similarity measure is identical to the one used for classification later on (see Eq. 8) and derived in a very detailed manner in A.1 and A.2. The weighted representative PDF thus unfolds to

$$p_{c,i}(x) = \int \frac{w_i(t)}{\overline{w_i}} \cdot p_c(x, t) \, dt \tag{4}$$

TABLE II
All variables and functions needed for the derivation of the similarity maps

Variables	
x	an arbitrary continuous variable, e.g. SAR feature
t	training samples of a certain class
i	iteration variable
c	class of interest
k	filter kernel
d	dimension
a	amplification of the SAR intensity in decibel
Probability density functions	
$p(x)$	PDF of an arbitrary entity along x
$p_c(x, t)$	PDF of an arbitrary entity along x and y (training samples)
$p_{c,i}(x)$	estimated PDF of class c in the i-th iteration step
$p_c(x)$	representative PDF of class c after the iteration
$p_0(x)$	PDF of the center pixel
$p_k(x)$	local PDF defined by filter kernel k
f_k	filter kernel to calculate $p_k(x)$ from $p_0(x)$
$p_\alpha(z)$	PDF of the Schmittlets along absolute azimuth orientations
$p_\rho(x)$	PDF of the Schmittlets along relative orientations
Weighting functions	
$w_i(t)$	iterative weight of the trainings samples of iteration i
$\overline{w_i}$	mean weight over the all trainings samples of iteration i
Similarity functions	
$s_{c,k,d}$	similarity between a representative PDF of a certain class of interest c and a local PDF defined by kernel k along a specific dimension d
$s_{c,k}$	similarity between a representative PDF of a certain class of interest c and a local PDF defined by kernel k
s_c	similarity to the class of interest c

where i stands for the iteration step. The iterative estimation of the optimal weights takes into account the similarity of the PDF $p_{c,i}(\mathrm{x})$ to each of the training samples gathered in $p_c(\mathrm{x}, \mathrm{t})$.

These steps (Eqs 3&4) are repeated until the expected change of the weights becomes negligible. In our case, mostly less than ten iterations were already sufficient to reach a mean correction lower than 10^{-6} per iteration. The mean similarity mostly reaches at least 90%. This step guarantees that outliers do not corrupt the representative PDF $p_c(\mathrm{x})$ of the class of interest.

2) Estimation of local PDFs in an image

With view to image classification, PDFs of a local pixel environment $p_k(\mathrm{x})$ are of interest. The starting point is defined by the PDF of one single pixel (equivalent to scale 0) that counts one where the image value v fits the respective bin x, and zero otherwise

$$p_0(\mathrm{x}) = \begin{cases} 1 & \text{if } \mathrm{x} = \mathrm{v} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

This PDF consequently is binary with only one positive entry because only the center pixel is considered. The shape and size of the local environment is given by a two-dimensional filter f_k. It shares the properties of a common PDF defined in Eq. 1. The convolution in spatial domain (denoted by $*$) hence produces a PDF $p_k(\mathrm{x})$ valid for a larger pixel environment

$$p_k(\mathrm{x}) = p_0(\mathrm{x}) * f_k. \tag{6}$$

In that way, multi-kernel PDF evaluation is simply enabled. One could explain it like this: the local histogram is filled by counting pixels with reduced radiometric sampling in a neighborhood. With respect to the Schmittlets, multi-scale and multi-directional processing becomes feasible. As the PDF estimation on varying pixel environments is reduced to a simple convolution, it can be implemented very effectively using parallel computing.

3) Estimation of the relative orientation of Schmittlets

In the case of urban structures, the azimuth direction of Schmittlets is secondary because the orientation of a building block towards the grid north direction generally is arbitrary. By contrast, the relative direction in a local environment, i.e. the presence of parallel or perpendicular pairs of Schmittlets is essential. Therefore, the elongated Schmittlets of one scale are grouped to angle pairs via

$$p_\rho(\mathrm{x}) = \frac{\int_0^\pi p_\alpha(\mathrm{z}) \cdot p_\alpha(\mathrm{x}-\mathrm{z})\, \mathrm{dz}}{\int_0^\pi p_\alpha(\mathrm{z})\, \mathrm{dz}}. \tag{7}$$

The PDFs hence are converted from azimuth orientations $p_\alpha(\mathrm{z})$ to relative directions $p_\rho(\mathrm{x})$. Multiples of π are not further distinguishable and therefore omitted. For instance, the first scale recognizing vertical (north-south) and horizontal (east-west) Schmittlets in $p_\alpha(\mathrm{z})$ now comprises parallel and perpendicular Schmittlet pairs independent of their absolute azimuth orientation in $p_\rho(\mathrm{x})$. This step is applied similarly to the PDF of each training site $p_c(\mathrm{x}, \mathrm{t})$ before the robust estimation of the representative PDF $p_c(\mathrm{x})$ as well as to the local PDFs

after spatial convolution $p_k(x)$.

B. Similarity estimation

1) Similarity of two PDFs

Consistent with the PDFs under study which range in [0,1], we aim at the comparison of two PDFs by a normalized measure. Such a measure can be derived from the X²-similarity test for discrete PDFs in A.1. Alternatively, it can be derived comparably in the continuous domain via the weighted harmonic mean (cf. A.2). Both derivations result in the same equation for the similarity $s_{c,k}$ between a local PDF in the image $p_k(x)$ and a representative PDF of a certain class $p_c(x)$ given by

$$s_{c,k} = \left\{ \int \frac{p_k{}^2(x)}{p_c(x)} \, dx \right\}^{-1} \in [0,1]. \tag{8}$$

The discrete evaluation of this measure requires a closed value range of the input data set as provided by the TANH scaling in the Kennaugh framework for (k_0, k_3, k_4, k_7) [37] or discrete image values with a very low bit-depth like the Schmittlets indices (S_b) [39]. In this case, the implementation reduces to a simple matrix multiplication and element-wise division even in the multiple class case, see A.3. Hence, the computation can easily be parallelized. The similarity ranges in-between 0 for completely different and 1 for identical PDFs and thus, can be interpreted as probability measure again. The closed value range is the main advantage in comparison to other measures, e.g., the Kullback-Leibler-Divergence [40], which mostly range in $[0, \infty[$.

2) Estimation of multi-dimensional similarities

In order to process multi-layer images, the one-dimensional method derived above has to be extended to the evaluation of multi-dimensional data sets. But, the handling of these multi-dimensional PDFs raises serious problems. Assuming a sampling rate of at least 21 bins (which is sufficient for TANH-scaled values), the PDF of a ten layer image (e.g. quad-pol Kennaugh elements) requires at least $21^{10} \approx 1.7 \cdot 10^{13}$ measurements. Apart from the immense memory demand, the number of pixels needed to adequately fill such a PDF is enormous. It is quite comprehensible, that this is not suitable for any practical application. Therefore, the multi-dimensional PDF is reduced to its marginal PDFs in each of the d dimensions and joint afterwards in the geometric mean which bases on the product of the marginal PDFs

$$s_{c,k} = \prod_d \{s_{c,k,d}\}^{-1/d} \in [0,1], \tag{9}$$

see A.4 for the detailed derivation. Using this formulation which is also known as "Naïve Bayes" because of ignoring the conditional probabilities, the number of pixels required is significantly reduced as the PDFs of the single layers can be evaluated separately. Back to the example of TANH-scaled values, the minimum number of pixels to fill a one-dimensional PDF quantized to 21 bins may already be reached by boxcar filter kernel f_k of 5-by-5 pixels. The uncomplicated extension to the multi-channel case is the second advantage of the new similarity measure over standard similarity measures like the Jensen-Shannon-Divergence (e.g.) which cannot easily be applied to marginal distributions separately. It is important to note, that correlation information is

completely ignored in this approach, i.e. classes that only vary in their correlation properties cannot be discriminated. Contrariwise, a weighting of the input layers is simply possible replacing the geometric mean in Eq. 9 by a weighted geometric mean. In that way, layers with a high discriminative potential can be privileged in comparison to other input features.

3) Estimation of multi-scale similarities

The only remaining variable in the left hand side of Eq. 9 is the kernel function f_k. The function defining the shape and scale of the pixel environment considered in the estimation of the local PDF can be arbitrarily chosen. Thus, any set of multi-scale kernels f_k (e.g. the 35 Schmittlets) can be adopted. In order to reduce the similarity $s_{c,k}$ to a certain class c estimated in different pixel environments defined by f_k to only one single similarity value s_c, simply the maximum value over all kernels is accepted

$$s_c = MAX_k(s_{c,k}). \qquad (10)$$

This step ensures that the best-fitting kernel is used for a specific class. Consequently, the class maxima might potentially appear in different scales in the multiple class case. This is reasonable because a potential class "building" (e.g.) logically might be part of a class named "residential area". The whole surrounding might be classified as "residential area" whereas single buildings might possibly stick out of the environment. To simplify matters, the single class case detecting slums is investigated exclusively here. The multi-class problem is treated in concurrent studies on land cover classification [41] and multi-sensor wetland monitoring [42] so far and will be subject to future studies as well.

IV. APPLICATION

This section shows application results derived by the methodology presented in the preceding section. It is divided into the introduction of the representative histograms for the slum class, the similarity maps of different scenarios, and the validation of recognized slums vs. other classes.

A. Sample PDFs

The representative PDFs are illustrated in Fig. 3. The dashes lines plot the PDFs for each test site separately whereas the solid line indicates the joint PDF over all test sites for the slum class.

B. Similarity maps

Based on these representative PDFs $p_c(x)$, similarity maps are generated using multi-scale environments. In order to accelerate the calculation a simple boxcar filter covering the window sizes $[5, 11, 25, 51, 101]$ is used as kernel function f_k. Numerous experiments are carried out and validated in the subsequent section. This section illustrates the input data and the derived similarity maps. Subsets of the reference (optical image in background), the similarity maps, the intensity images, polarimetric Kennaugh elements and best-fitting Schmittlets of the respective image acquisitions in ascending pass direction with a flat incidence angle are presented in Fig. 4 for Cape Town, in Fig. 5 for Manila, and in Fig. 6 for Mumbai. The complete similarity maps

are calculated by combining all images per test site together are given in Fig. 7. Similarities in-between 50% and 70% are colored in green tones. Similarities exceeding 70% are colored in ochre tones. The expected similarity ranges around 90% as derived from the reference samples.

C. Validation

In order to numerically validate the similarity maps, their values are correlated with the reference data and the resulting distribution is plotted in Figs 8-10. Common approaches presume a quantization into a low number of equally-frequent classes. Unfortunately, this is not the case in our study because of the rare occurrence of the class slums. Per-class-measures like the completeness (also producer's accuracy) and correctness (or user's accuracy) are more appropriate in order to level out the unequal occurrence of the classes and thus, are better suited for our single-class problem. Additionally, the two measures under study can simply be derived without prior quantization, i.e. instead of thresholding the similarity measure in favor of a binary slum / no slum classification the slum similarities are directly evaluated. Completeness only concerns the samples of the slum class whereas correctness always involves other classes. Therefore, the correctness is checked in relation to mixed settlements, to pure formal settlements and to non-built-up classes, i.e. water, bare soil, forest, etc.
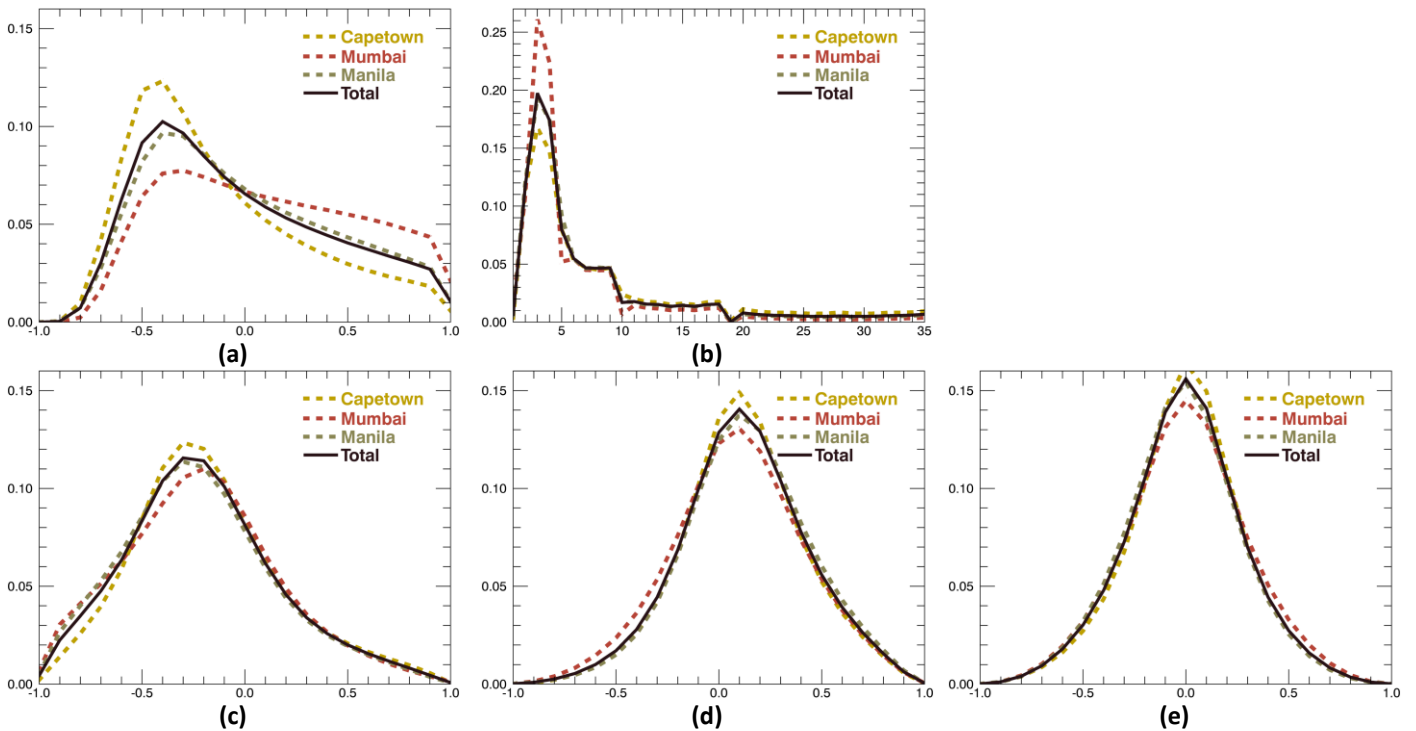

Fig. 3. The representative PDFs $p_c(x)$ for the slum class of (a) the normalized intensity k_0, (b) the Schmittlets S_i and the normalized polarimetric Kennaugh elements (c) k_3, (d) k_4, (e) k_7, derived from the three test sites Cape Town, Mumbai, Manila, and over all test sites. The ordinate shows the relative frequency. The abscissa contains the normalized Kennaugh values]-1,+1[and the 35 Schmittlet indices respectively. Illustrations of the trainings sites can be found in Figs 4-6.
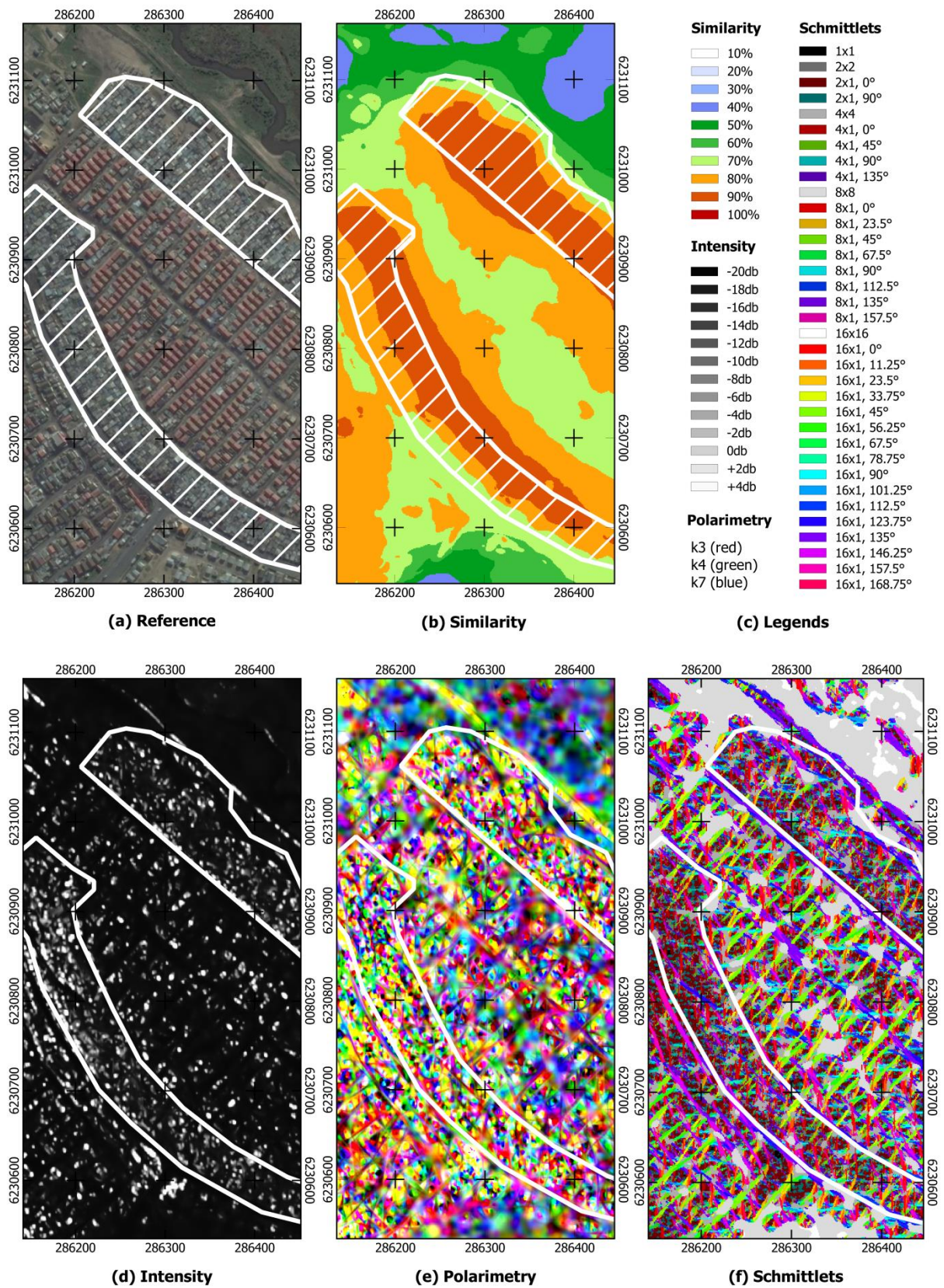
Fig. 4. Subset of the test site Cape Town: (a) optical image from Google Earth, (b) similarity map derived by the presented technique, (c) legends to the illustrations, (d) intensity image, (e) RGB color composition of polarimetric Kennaugh elements, and (f) the best-fitting Schmittlets. Slums are delineated by white lines with (above) or without (below) hatching. The coordinates refer to UTM zone 34S.
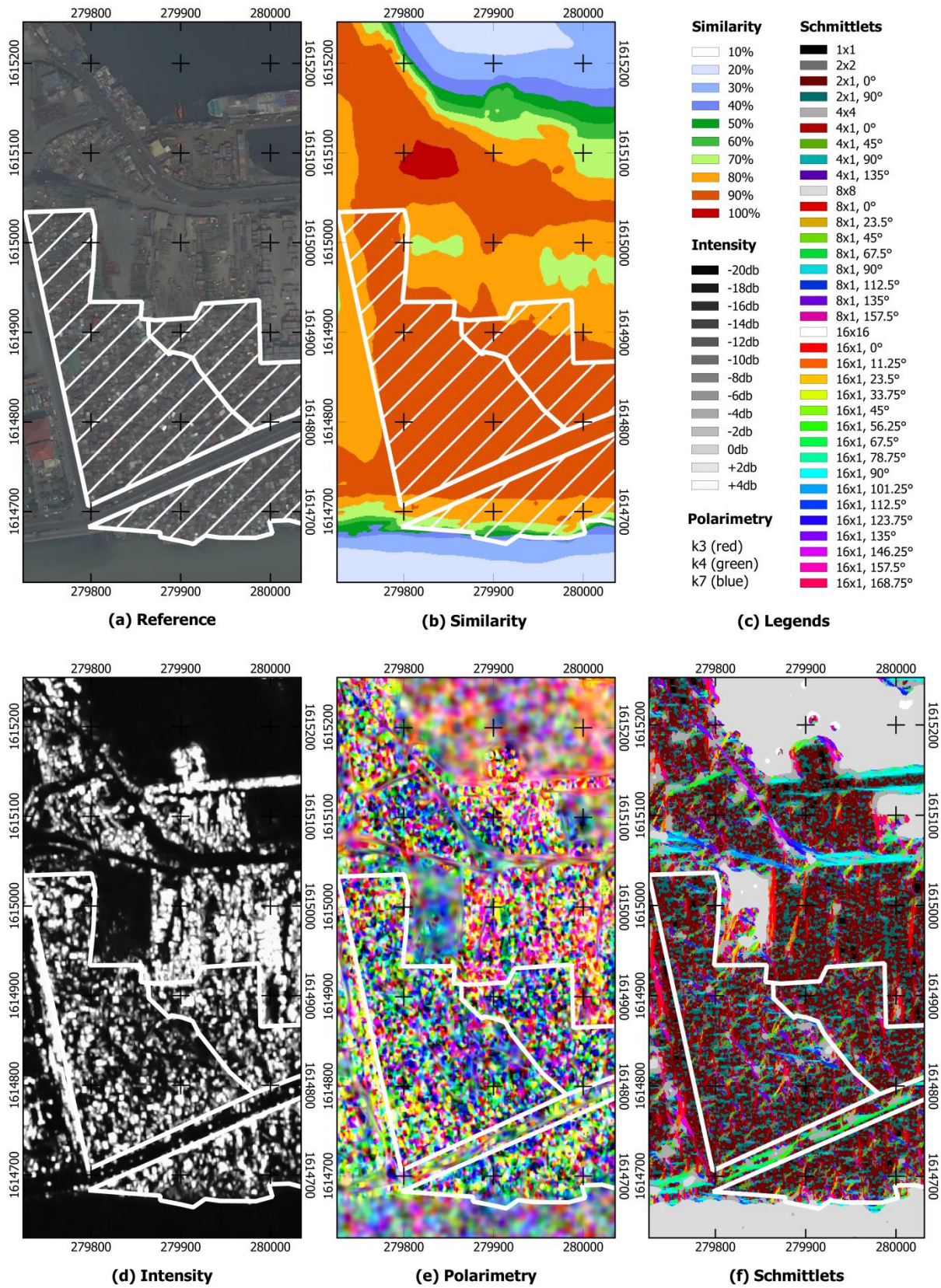
Fig. 5. Subset of the test site Manila: (a) optical image from Bing Maps, (b) similarity map derived by the presented technique, (c) legends to the illustrations, (d) intensity image, (e) RGB color composition of polarimetric Kennaugh elements, and (f) the best-fitting Schmittlets. Slums are delineated by white lines with (above) or without (below) hatching. The coordinates refer to UTM zone 51N.

(a) Reference

(b) Similarity

(c) Legends

Similarity

	10%
	20%
	30%
	40%
	50%
	60%
	70%
	80%
	90%
	100%

Intensity

	-20db
	-18db
	-16db
	-14db
	-12db
	-10db
	-8db
	-6db
	-4db
	-2db
	0db
	+2db
	+4db

Polarimetry

k3 (red)
k4 (green)
k7 (blue)

Schmittlets

	1x1
	2x2
	2x1, 0°
	2x1, 90°
	4x4
	4x1, 0°
	4x1, 45°
	4x1, 90°
	4x1, 135°
	8x8
	8x1, 0°
	8x1, 23.5°
	8x1, 45°
	8x1, 67.5°
	8x1, 90°
	8x1, 112.5°
	8x1, 135°
	8x1, 157.5°
	16x16
	16x1, 0°
	16x1, 11.25°
	16x1, 23.5°
	16x1, 33.75°
	16x1, 45°
	16x1, 56.25°
	16x1, 67.5°
	16x1, 78.75°
	16x1, 90°
	16x1, 101.25°
	16x1, 112.5°
	16x1, 123.75°
	16x1, 135°
	16x1, 146.25°
	16x1, 157.5°
	16x1, 168.75°

(d) Intensity

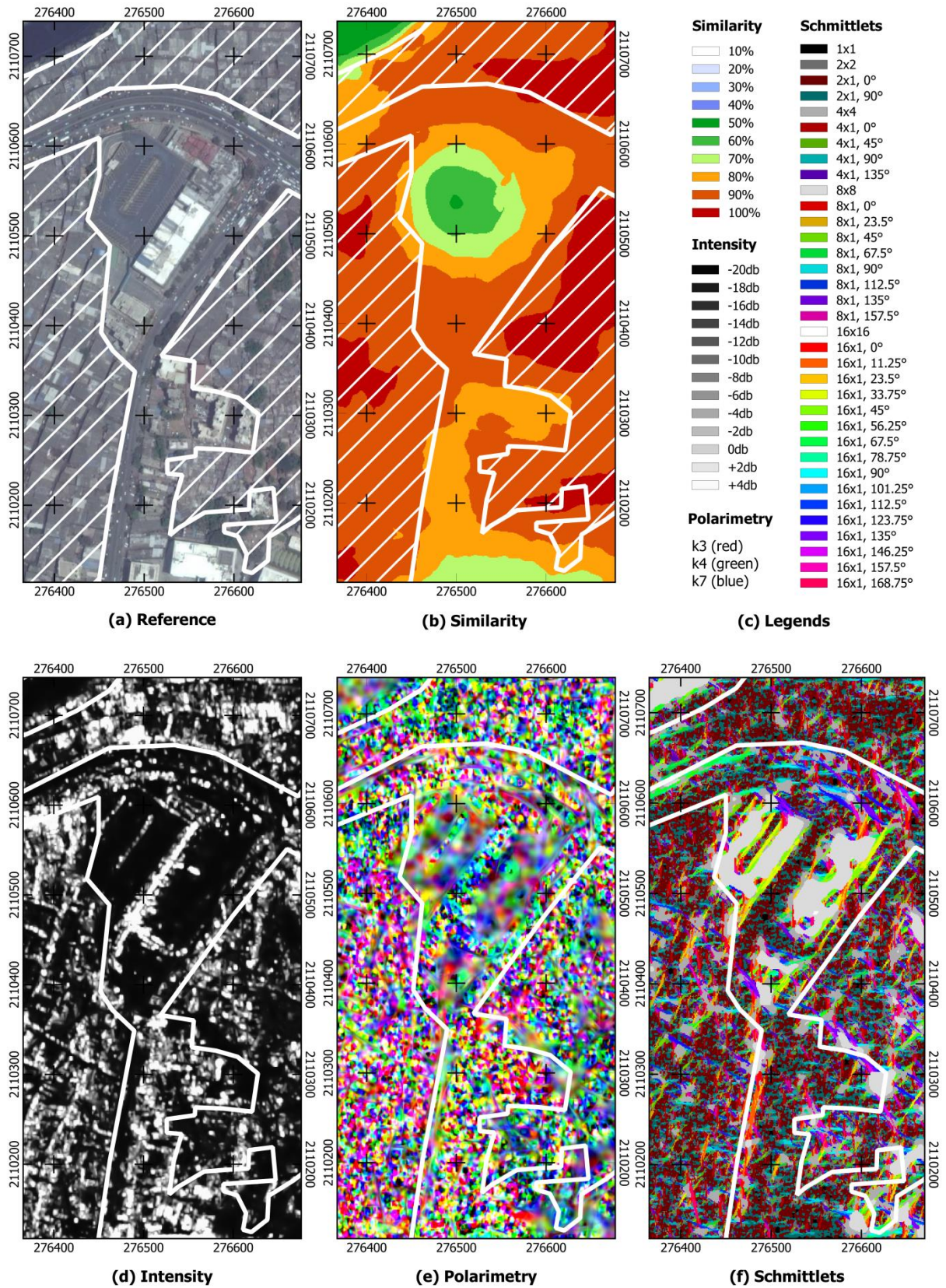**(e) Polarimetry**

(f) Schmittlets

Fig. 6. Subset of the test site Mumbai: (a) optical image from Google Earth, (b) similarity map derived by the presented technique, (c) legends to the illustrations, (d) intensity image, (e) RGB color composition of polarimetric Kennaugh elements, and (f) the best-fitting Schmittlets. Slums are delineated by white lines with (above) or without (below) hatching. The coordinates refer to UTM zone 43N.

The interpretation of Figs 8-10 is quite simple: for a similarity of zero all slums are detected and the completeness logically equals one. When the similarity increases, the completeness decreases and reaches zero when the similarity reaches one, i.e. the higher the threshold the more class members are not detected. Using a very low threshold the correctness is equally low because of a vast overestimation. With a higher threshold, the correctness understandably increases because of a lower false alarm rate. The intersecting point of completeness and correctness marks the similarity value where completeness and correctness vs. the designated class become equal, see Tables III, IV, and V. The denoted tables contain even more experiments than reported in Figs 8-10. The corresponding plots are omitted for clarity reasons.

Taking Fig. 8a as practical example, the completeness and correctness are plotted for the slum similarity derived from a dual-co-polarized high resolution spotlight image of TerraSAR-X in ascending orbit direction with steep incidence angle. A similarity value of 0.5 (see the abscissa) promises a completeness of about 95%, but a correctness of only 60% vs. formal settlements, 40% vs. non-built-up classes, and 20% vs. mixed settlements. In other words, applying a similarity threshold of 0.5 for classification detects 95% of all slum areas, but also delivers 80% of the mixed settlements, 60% of the non-built-up areas, and 40% of the formal settlements. Increasing the similarity threshold increases the correctness and decreases the completeness. A similarity of 0.7 (e.g.) leads to a completeness value of about 80% whereas the correctness vs. formal settlements ranges just above 80% and the correctness vs. non-built-up classes ranges just below 80%. The separability of slums and mixed settlements is understandably very low. The balance between completeness and correctness is reached in 60% which refers to a similarity of 0.8, see Fig. 8a. In the optimal case, the correctness already increases at low similarity values, but the completeness does not decrease until higher similarity values are reached. In consequence, the intersecting point has a very high cumulative probability. The explicit partition of the correctness in correctness towards different classes allows for the estimation of the quality for the distinction of the two classes under consideration which is reported in the Tables III, IV, and V.

V. Discussion

A. Sample PDFs

Fig. 3 shows that the training data of slums show nearly identical features in the three test sites although they are from different cities in different cultural areas. The polarimetric Kennaugh elements (Fig. 3 c-e) indicate a central – almost truncated normal – distribution, for all three study sites. The only difference can be found in the spreading: Cape Town delivers training samples with the highest accordance, whereas the samples of Mumbai are characterized by a higher variation. The mode of k_3 (Fig. 3c) is shifted to negative values, i.e. surface scattering is the dominant scattering mechanism over slums equally measured in all test sites.
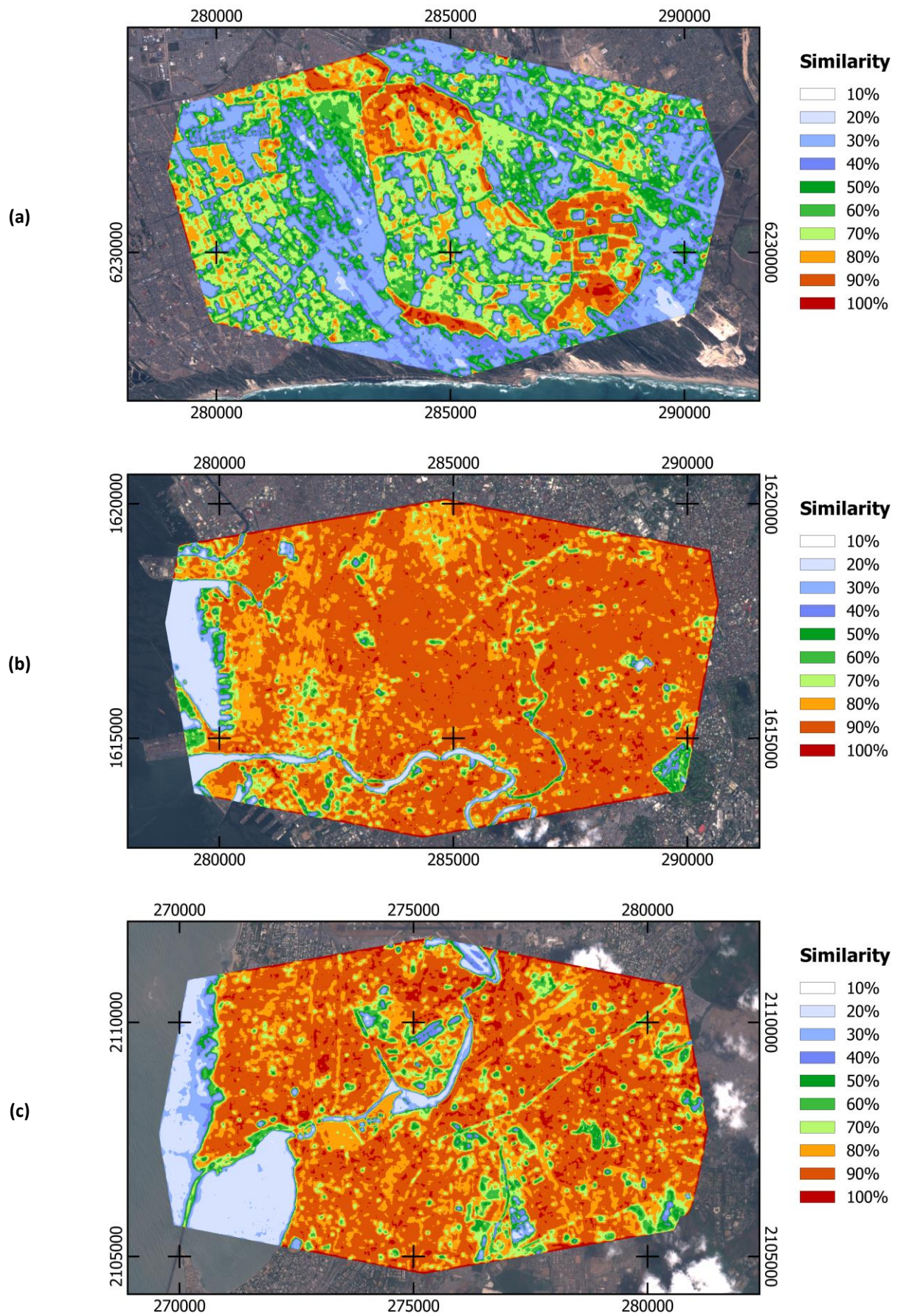
Fig. 7. Complete view of similarity maps s_c derived by the combination of ascending and descending dual-co-polarized high resolution spotlight images of TerraSAR-X (©DLR 2015, 2016) framed by partly cloudy Sentinel-2 multi-spectral images (©ESA 2017) for test site (a) Cape Town (UTM 34S), (b) Manila (UTM 51N), and (c) Mumbai (UTM 43N).
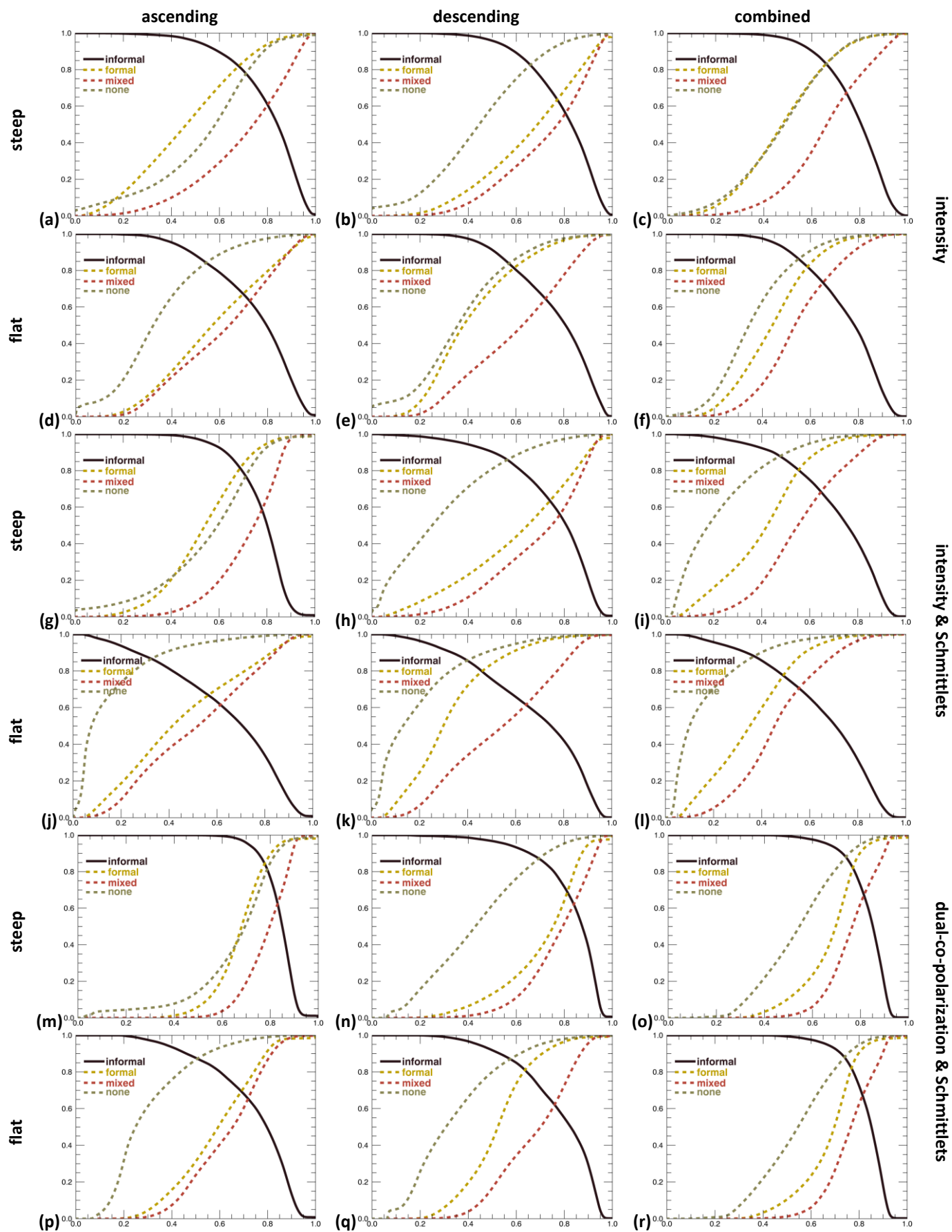
Fig. 8. Validation of the similarity maps s_c from the test site Cape Town: completeness of the slum class and correctness vs. "formal", "mixed", and "non-built-up" areas on the ordinate in relation to the similarity on the abscissa for following layer combinations: (a)-(f) intensity only (k_0), (g)-(l) intensity with Schmittlet index (S_b), and (m)-(r) supplementary polarimetry (k_3,k_4,k_7). All layer combinations are tested using varying passing directions (columns) and varying incidence angles (rows).
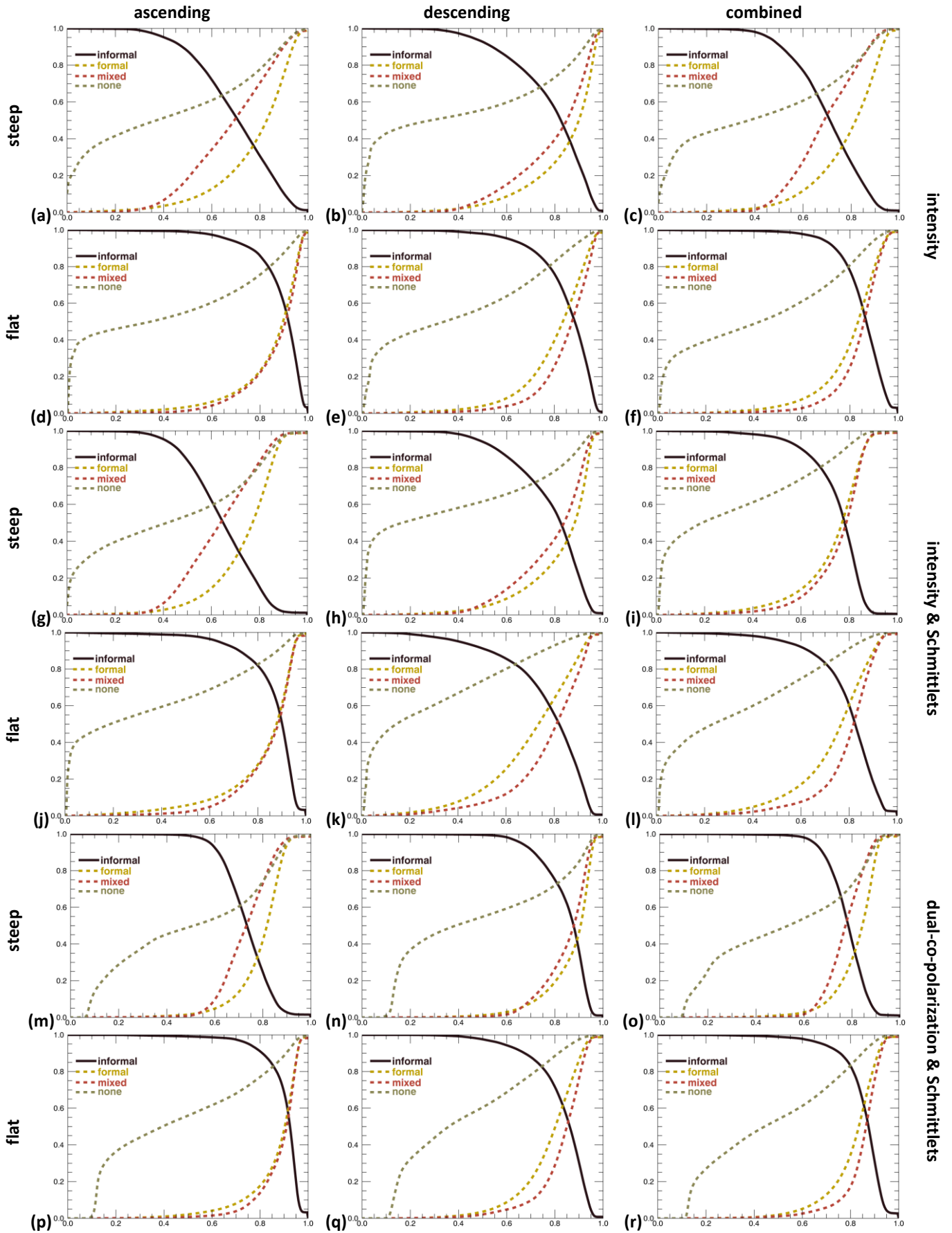
Fig. 9. Validation of the similarity maps s_c from the test site Manila: completeness of the slum class and correctness vs. "formal", "mixed", and "non-built-up" areas on the ordinate in relation to the similarity on the abscissa for following layer combinations: (a)-(f) intensity only (k_0), (g)-(l) intensity with Schmittlet index (S_b), and (m)-(r) supplementary polarimetry (k_3,k_4,k_7). All layer combinations are tested using varying passing directions (columns) and varying incidence angles (rows).
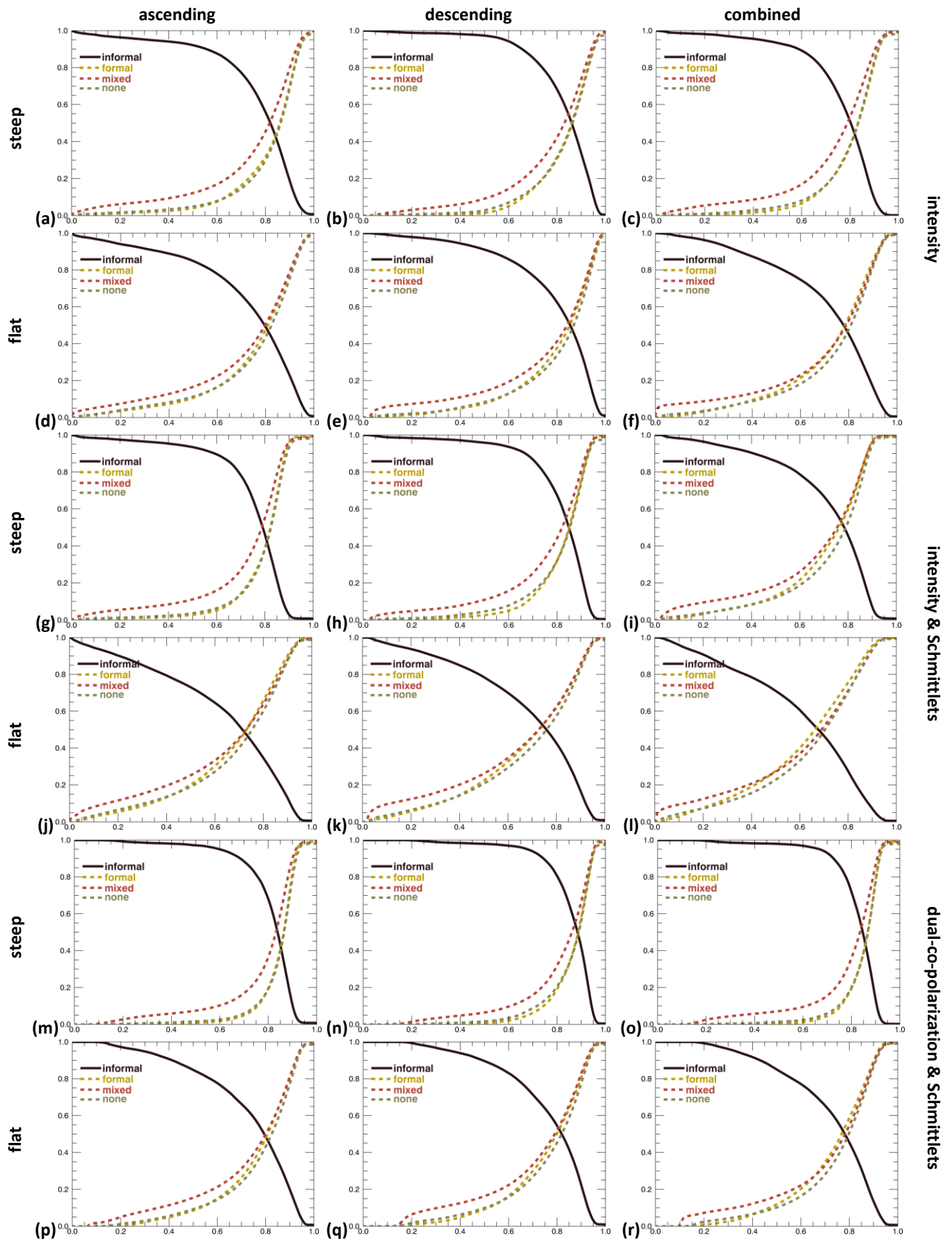
Fig. 10. Validation of the similarity maps s_c from the test site Mumbai: completeness of the slum class and correctness vs. "formal", "mixed", and "non-built-up" areas on the ordinate in relation to the similarity on the abscissa for following layer combinations: (a)-(f) intensity only (k_0), (g)-(l) intensity with Schmittlet index (S_b), and (m)-(r) supplementary polarimetry (k_3,k_4,k_7). All layer combinations are tested using varying passing directions (columns) and varying incidence angles (rows).

The relation of the co-pol intensities in k_4 (Fig. 3d) suggests a slight shift to positive values, i.e. a higher backscattering in HH than in VV. The co-pol correlation or phase shift in k_7 (Fig. 3e) is centered over zero, i.e. no dominant scattering behavior can be observed. One might attribute this observation to the deficiencies of a partial polarimetric acquisition ignoring the cross-polarized component. But a concurrent study published recently proves that polarimetric Kennaugh elements derived from dual-co-polarized images can replace the elements from quad-polarized images nearly without limitation [43]. The total intensity in Fig. 3a underlines that slums are characterized by a rather low backscattering, i.e. lower than 7dB in the visualized case. Slums in Cape Town have the lowest backscattering over all four images under study. From Manila to Mumbai the proportion of brighter targets slightly increases. Nevertheless, the representative PDFs are quite similar for all three test sites. The same behavior can be observed for the best-fitting Schmittlet indices S_i in Fig. 3b: Only the proportion of very short Schmittlets slightly varies between 18% (Cape Town) and 26% (Mumbai). This may correspond to the observed variations in the total intensity: a lower total intensity indicates a lower number of bright targets and therewith no necessity for small Schmittlets to be used for image enhancement. One reason might be a lower building density of the slums in the Cape Town data set which is indicated by visual inspection.

B. Similarity maps

The representative PDFs derived from all image acquisitions in all test sites provide the basis for the similarity maps produced by the multi-scale approach. The purely visual inspection of the total intensity in Fig. 4d indicates nearly no difference of the slums from the remaining areas. Only bare soil stands out by its lower intensity and smooth appearance after the Schmittlet enhancement. The best-fitting Schmittlet index image show rather dark red or turquoise colors over slums, which indicates only small structures in contrast to grey tones over unstructured areas (e.g., bare soil) and brighter colors, i.e. longer linear structures (e.g., streets) in the surrounding areas. Regarding Manila in Fig. 5 the overall proportion of short Schmittlets is much higher than in Cape Town, i.e., even formal building agglomerates show very fine structures. Only non-built-up areas like water or bare soil are clearly distinguishable. In the chosen subset of Mumbai (Fig. 6) the areas of short Schmittlets (dark red or turquoise) are interfused with longer linear structures quite contrary to what is observed in Cape Town and Manila. Despite this difference, the similarity map produced by combining all images over the respective test site delivers a high concordance with the reference data set in all three test sites. The similarity over slums provided by the reference data set is generally higher than 70% whereas surrounding areas drop down to less than 50% in the case of bare soil and less than 20% in the case of water, see Fig. 5b. The applied method does not seem to be suitable to trace the borders of slums. Indeed, it is a very difficult and often fuzzy task even for visual interpreters to delineate structural changes in the complex urban environments, cf. Fig. 1. Thus, the method provides a local probability for the existence of slums. Back to the Manila test site where the harbor area shows similar patterns like the slums: a container terminal for sure cannot be distinguished from container settlements. Mono-temporal remote sensing methods can only recognize an agglomerate of containers, but they are not able to recognize the use of the containers: for transportation or as shelter for inhabitants. In other words, earth observation techniques are only able to derive

the syntax of slums, e.g. physical indicators. The semantic classification as slum may potentially deviate from that.

C. Validation

The special validation considers a continuous variation of the similarity in order to recognize a certain local PDF $p_k(x)$ as PDF of the slum class $p_c(x)$. Although the weighting of the completeness measure versus the correctness measure strongly depends on the application, only the case where completeness equals correctness will be addressed in the following for the sake of clarity. The respective value is printed as ordinate value of the intersecting point of the black curve with the colored curves in Figures 8-10. For each test site, combinations of 36 features and acquisitions geometries are tested and reported in Tables II-IV: intensity only (k_0), Schmittlets only (S_b), dual-co-polarized Kennaugh elements (k_0, k_3, k_4, k_7), Intensity and Schmittlets (k_0, S_b), dual-co-polarization and Schmittlets (k_0, k_3, k_4, k_7, S_b) with a flat or a steep incidence angle in ascending or descending orbit direction respectively as well as combinations of the available imaging geometries. Only the most important combinations are illustrated: The plots for Cape Town can be found in Fig. 8. Throughout all combinations the distinction between the structural class slums and non-built-up areas ranges around 80% to 90% and represents the best class distinction except for one image acquisition: the image with a steep incidence angle in ascending orbit direction (see Figs 8a&g&m) provides a better distinction between the structural classes of formal settlements and slums. This can be seen as positive outlier caused by the acquisition geometry and the orientation of the built-up areas. The discrimination between formal settlements and slums generally ranges between 60% and 85% percent except for the Schmittlets in Fig. 8g that show only 40%. The lowest distinction accuracy with around 60% is perceived between slums and the mixed class, which is conceivable because this class contains both formal and slums. Combining ascending and descending image acquisitions usually increases the accuracy by 10 percentage points. Combining backscatter and structure information (see Fig. 8g-l) has a further stabilizing effect. Most interestingly, the use of polarimetric information does not seem to influence the accuracy because the curves using the full dual-co-polarized information do not differ very much from the ones resulting from the use of intensity and Schmittlets only (cf. Figs 8i&o). Looking at Manila in Fig. 9 the distinction between slums and non-built-up areas ranges in-between 60% and 80%. The separability of formal and slums drops down to 50%. The combination of features and acquisition

TABLE III

Accuracies achieved using the denoted layer and acquisition geometry combinations. The assigned value reflects the equilibrium between completeness and correctness vs. following classes: **formal settlements (bold)**, *mixed settlements (italic)*, non-built-up classes as given in Figs 6&7 for Cape Town.

inc. angle	ascending			descending			combined			layers
steep	**82**	*61*	78	**64**	*56*	83	**83**	*67*	83	k_0
flat	**67**	*64*	85	**81**	*65*	84	**82**	*53*	87	
steep	**45**	*40*	73	**62**	*56*	86	**56**	*49*	87	S_b
flat	**66**	*61*	86	**62**	*76*	86	**74**	*69*	88	
steep	**87**	*65*	82	**72**	*64*	86	**87**	*70*	87	k_0
flat	**73**	*70*	88	**84**	*65*	89	**83**	*75*	90	k_3, k_4, k_7
steep	**81**	*59*	78	**64**	*56*	86	**80**	*69*	88	
flat	**67**	*61*	87	**80**	*62*	85	**77**	*70*	88	k_0 S_b
combined	**75**	*65*	89	**80**	*69*	88	**79**	*70*	89	
steep	**85**	*63*	82	**70**	*62*	87	**82**	*68*	88	k_0
flat	**69**	*66*	87	**82**	*62*	87	**80**	*72*	89	k_3, k_4, k_7
combined	**74**	*69*	90	**83**	*70*	89	**82**	*72*	90	S_b

geometries enhances the accuracy of slums vs. non-built-up, but does not change the distinction between the different built-up classes. Again, the polarimetric information seems to be of minor influence. The validation for the test site Mumbai in Fig. 10 indicates class distinction accuracies around 50% for all layer combinations. Combining ascending and descending acquisitions in general improves the accuracy. The incidence angle is playing a minor role in this context. Even combinations of steep and flat incidence angles do not increase the separability considerably.

In order to simplify the interpretation of the results, the accuracies reached in the equilibrium of completeness and correctness vs. formal settlement, mixed settlement and non-built-up areas are summarized in the Tables III, IV, and V. The maximum accuracies of about 90% can be found in the test site Cape Town (Table III) for the discrimination of slums from non-built-up areas. The discrimination from formal settlements ranges around 87% for the combination of ascending and descending geometries of dual-co-polarized images. With view to Manila (Table IV), the highest accuracy again is reported for the discrimination of slums from non-built-up areas in the combination of ascending and descending images using both polarimetric and structural information at once. The maximum discrimination between formal and slums only amounts 62%. In our third test site Mumbai, the class discrimination is extremely low and ranges around 50% only for all class and image combinations, see Table V.

VI. Conclusion

This paper studies the appearance of slums in dual-co-polarized high resolution SAR images based on their empirical distribution. Intensity information is provided by the backscattering strength k_0 in hyperbolic tangent scaling, polarimetric information is kept in the normalized Kennaugh elements k_3, k_4, and k_7. Local Texture is described by the indices of the best-fitting Schmittlets S_b whereas the originally absolute azimuth orientations are reduced to relative directions in order to represent parallel, perpendicular, or diagonal structures (inter alia). The representative PDFs for the slum class are derived from training data sets via a robust estimation approach that reduces the impact of possible outliers iteratively. Local PDFs are calculated by convolving one multi-channel probability image per input layer with a multi-scale filter bank. A simple similarity measure guaranteeing a closed value range and resulting in a new probability measure is applied for the comparison of the local and the representative

TABLE IV

Accuracies achieved using the denoted layer and acquisition geometry combinations. The assigned value reflects the equilibrium between completeness and correctness vs. following classes: **formal settlements (bold)**, *mixed settlements (italic)*, non-built-up classes as given in Figs 8 & 9 for Manila.

inc. angle	ascending			descending			combined			layers
steep	**37**	*52*	64	**39**	*48*	68	**36**	*54*	65	k_0
flat	**56**	*54*	80	**59**	*53*	80	**58**	*53*	80	
steep	**32**	*50*	54	**41**	*50*	74	**34**	*52*	65	S_b
flat	**75**	*76*	83	**60**	*51*	82	**60**	*50*	83	
steep	**35**	*51*	64	**42**	*50*	72	**38**	*75*	67	k_0
flat	**59**	*56*	82	**62**	*55*	81	**60**	*56*	83	k_3, k_4, k_7
steep	**34**	*51*	60	**41**	*48*	72	**53**	*49*	80	k_0
flat	**56**	*55*	83	**60**	*51*	83	**60**	*51*	84	S_b
combined	**39**	*54*	72	**52**	*49*	80	**44**	*53*	75	
steep	**33**	*50*	61	**43**	*50*	74	**36**	*53*	65	k_0
flat	**59**	*57*	83	**62**	*54*	83	**61**	*54*	84	k_3, k_4, k_7
combined	**40**	*53*	73	**56**	*51*	81	**46**	*54*	75	S_b

PDFs to produce similarity maps. Similarity in this context is synonymous with the probability of local occurrence the class of interest. Thanks to the mathematical characteristics of this similarity measure the input layers can be evaluated separately and joined afterwards which simplifies the data processing to a great extent.

The validation is performed by weighting completeness vs. correctness in relation to different reference classes. In general, the measured appearance of slums (based on the morphological reference information) in SAR images is very similar in all three test sites: Cape Town, Manila, and Mumbai. The combination of intensity and structure information provided by the Schmittlets has the greatest impact on the discrimination from other urban structure types. Polarimetric layers in general only deliver a slight enhancement of the discrimination of slums from formal settlements. In contrast, polarimetry turns out to be essential for the distinction of settlements from non-built-up areas. The combination of different acquisition geometries, mainly ascending and descending acquisitions, further enhances the classification accuracy. In summary, local patterns seem to be the crucial feature for the recognition of slums. Although they are similar for three cities under study (even on different continents) the contrast to surrounding formal buildings (or its relation to the surrounding urban morphology) is very special with respect to the specific complex urban morphologic configurations in each location. The Cape Town test site delivers accuracies for the class discrimination of slums from formal settlements up to 87% which is excellent whereas Manila reaches 62% and Mumbai only allows for slightly more than 50%. Thus, the presented method is applicable, but the classification accuracy highly depends on the specific city structure.

This result is supported by a concurrent study on the use of grey-level co-occurrence matrices and random forest classifiers for the identification of urban structure types [22]: Although it is generally possible to separate formal settlements from slums using multi-polarized SAR images and texture or structural information, the impact of the orientation of the quarters inter alia is immense. This fact is confirmed by our study: the combination of ascending and descending acquisitions (and hence varying aspect angles) promises the best discrimination accuracies. However, space-borne SAR sensors today can only provide a narrow range of imaging geometries.

Future SAR missions will provide high resolution wide swath SAR images and enable regional coverages to be derived. Further, experiments with passive SAR receivers on microsatellites are ongoing. The idea behind this

TABLE V
Accuracies achieved using the denoted layer and acquisition geometry combinations. The assigned value reflects the equilibrium between completeness and correctness vs. following classes: **formal settlements (bold)**, *mixed settlements (italic)*, non-built-up classes as given in Figs 10 & 11 for Mumbai.

inc. angle	ascending			descending			combined			layers
steep	**44**	*51*	43	**50**	*55*	50	**46**	*53*	46	k_0
flat	**49**	*50*	47	**51**	*52*	49	**50**	*49*	46	
steep	**46**	*53*	46	**51**	*55*	50	**46**	*55*	46	S_b
flat	**49**	*48*	47	**53**	*52*	50	**52**	*50*	49	
steep	**42**	*52*	42	**50**	*55*	50	**44**	*53*	44	k_0
flat	**48**	*50*	46	**54**	*55*	51	**50**	*50*	49	k_3, k_4, k_7
steep	**43**	*50*	43	**50**	*55*	50	**54**	*55*	50	
flat	**49**	*49*	47	**53**	*53*	50	**50**	*49*	48	k_0
combined	**46**	*49*	45	**54**	*55*	50	**50**	*50*	47	S_b
steep	**42**	*52*	42	**50**	*55*	50	**45**	*53*	45	k_0
flat	**49**	*49*	46	**53**	*54*	50	**51**	*50*	49	k_3, k_4, k_7
combined	**45**	*50*	44	**54**	*55*	50	**51**	*52*	48	S_b

bi-static SAR concept [44] is the illumination of a scene by one emitter and the recording of the backscattered signal by several receivers. Thus, multiple imaging geometries become conceivable. To support this work, a signature database for different urban structure types might be established allowing for a multi-class differentiation of settlement types. As stated before, the presented method is designed to derive class similarities. With respect to the detection of slums this means that local probabilities for the existence of slums are produced. The method thus can only point out possible locations of those by the local patterns. The semantic decision, e.g., if a detected container depot is inhabited or not, cannot be taken automatically. The technique introduced in this paper can carve out regions of interest, even in regional to global data sets, but the final interpretation still has to take into account much more data sources and therefore, will still need human interaction in the near future.

In summary, this study constitutes a further step towards the practical application of SAR remote sensing for a regional inventory of slums and other informal settlements. Together with further data sources and observations it might provide one component to the understanding of self-built environments which is expected to host about two billion people in the coming years.

REFERENCES

[1] UN-HABITAT, "The Challenge of Slums: Global Report on Human Settlement – revised and updated version 2010." http://unhabitat.org/wp-content/uploads/2003/07/GRHS_2003_Chapter_01_Revised_2010.pdf, 2010.

[2] Werthmann, C. & Bridger, J., "Metropolis", nonformal, 2016.

[3] Kuffer, M., Pfeffer, K. and Sliuzas, R.V., "Slums from space: 15 years of slum mapping using remote sensing." Remote Sensing. vol. 8, no. 6, pp. 455ff, 2016.

[4] Taubenböck H.; Wurm M., "Ich weiß, dass ich nichts weiß – Bevölkerungsschätzung in einer Megacity." SpringerSpektrum. pp. 171-178, 2015.

[5] Kohli, D.; Sliuzas, R.V.; Kerle, N. & Stein, A., "An ontology of slums for image-based classification." Comput. Environ. Urban Syst., vol. 36, pp. 154–163, 2012.

[6] Owen, K.K.; Wong, D.W., "An approach to differentiate informal settlements using spectral, texture, geomorphology and road accessibility metrics." Appl. Geogr., vol. 38, pp. 107–118, 2013.

[7] Taubenböck, H.; Kraff, N.J., "The physical face of slums: A structural comparison of slums in Mumbai, India, based on remotely sensed data." J. Hous. Built Environ., vol. 29, pp. 15–38, 2014.

[8] Niebergall, S.; Loew, A & Mauser, M., "Integrative assessment of informal settlements using VHR remote sensing data—The Delhi case study." IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 1, no. 3, pp. 193–205, 2008.

[9] Hofmann, P.; Strobl, J.; Blaschke, T. & Kux, H., "Detecting Informal Settlements From QuickBird Data in Rio De Janeiro Using an Object Based Approach." In: T. Blaschke, S. Lang, and G. J. Hay (Eds) Lecture Notes in Geoinformation and Cartography, Berlin, Heidelberg: Springer, pp. 530-553, http://dx.doi.org/10.1007/978-3-540-77058-9_29, 2008.

[10]Hofmann, P., "Defining Robustness Measures for OBIA Framework - A Case Study for Detecting Informal Settlements." In: Weng, Q. (Ed.) Global Urban Monitoring and Assessment through Earth Observation. Boca Raton, CRC Press Taylor & Francis, pp. 303-324, 2014.

[11]Shekhar, S., "Detecting slums from QuickBird data in Pune using an object-oriented approach." International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 39, no. 8, pp. 519–524, 2012.

[12]Kohli, D.; Warwadekar, P.; Kerle, N.; Sliuzas, R. & Stein, A., "Transferability of object-oriented image analysis methods for slum identification." Remote Sensing. vol. 5, no. 9, pp. 4209-4228, 2013.

[13]Engstrom, E., Sandborn, A., Yu, Q., Burgdorfer, J., Stow, D., Weeks, J. & Graesser, J. (2015). Mapping Slums Using Spatial Features In Accra, Ghana, Joint Urban Remote Sensing Event 2015, Lausanne, Switzerland.

[14]Sandborn, A. & Engstrom, R. N. (2016): Determining the Relationship Between Census Data and Spatial Features Derived From High-Resolution Imagery in Accra, Ghana, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(5), 1970-1977.

[15]Kuffer M. & Barros J (2011) Urban morphology of unplanned settlements: The use of spatial metrics in VHR remotely sensed images. Procedia Environmental Sciences 7 (011) 152-157.

[16]Ferro-Famil, L.; Pottier, E. & Lee, J.-S., "Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha-Wishart classifier" IEEE Transactions on Geoscience and Remote Sensing, 39(11), 2332-2342, 2001.

[17]Marino, A.; Cloude, S.R. & Woodhouse, I.H., "Detecting Depolarizing Targets using a New Geometrical Perturbation Filter", IEEE Transactions on Geoscience and Remote Sensing, 50(10), 3787-3799, 2012.

[18]Ulaby, F.T.; Kouyate, F.; Brisco, B.; Williams, T.H., "Textural Information in SAR Images." IEEE Transactions on Geoscience and Remote Sensing. vol. GE-24, no. 2, pp. 235-245, 1986.

[19]Habermeyer, M. & Schmullius, C., "An algorithm for knowledge based classification of multitemporal remote sensing radar data." (German). Photogrammetrie, Fernerkundung, Geoinformation, vol. 5, pp. 313-323, 1997.

[20]Wen, C.; Zhang, Y.; Deng, K., "Urban Area Classification in High Resolution SAR based on Texture Features." ISPRS Archives – Volume XXXVIII-7/C4, pp. 281-285, http://www.isprs.org/proceedings/XXXVIII/7-C4/281_GSEM2009.pdf., 2009.

[21]Dell'Acqua, F.; Gamba, P., "Texture-based characterization of urban environments on satellite SAR images." IEEE Transactions on Geoscience and Remote Sensing, vol. 41, no. 1, pp. 153-159, doi: 10.1109/TGRS.2002.807754, 2003.

[22]Wurm, M.; Taubenböck, H., Weigand, M.; Schmitt, A., "Slum mapping in polarimetric SAR data using spatial features." Remote Sensing of Environment, vol. 194, pp. 190-204, doi: 10.1016/j.rse.2017.03.030, 2017.

[23]Esch, T.; Marconcini, M.; Felbier, A.; Roth, A.; Heldens, W.; Huber, M.; Schwinger, M; Taubenböck, H.; Müller, A. & Dech, S., "Urban Footprint Processor – Fully Automated Processing Chain Generating

Settlement Masks From Global Data of the TanDEM-X Mission." IEEE Geoscience and Remote Sensing Letters, vol. 10, no. 6, pp. 1617-1621, 2013.

[24] Canadian Space Agency, "The Radarsat Constellation Mission." http://www.asc-csa.gc.ca/pdf/eng/publications/radarsat-constellation-eng.pdf, visited 2016-09-28.

[25] Krieger, G.; Younis, M.; Gebert, N.; Huber, S.; Bordoni, A.; Patychenko, A.; Moreira, A., "Advanced Concepts for High-Resolution Wide-Swath SAR Imaging." Proceedings of EUSAR 2010, pp. 524-527, 2010.

[26] Rodrigues, A.; Corr, D.G.; Pottier, E.; Ferro-Famil, L.; Hoekman, D., "Land Cover Classification using Polarimetric SAR data. Proceedings of the ESA PolInSAR Workshop 2003." http://earth.esa.int/workshops/polinsar2003/papers/10_rodrigues.pdf, visited 2016-09-28, 2003.

[27] Novack, T. & Stilla, U., "Classification of Urban Settlements Types based on Space-Borne SAR Datasets." ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. II-7, pp. 55-60, 2014.

[28] Hänsch, R.; Hellwich, O., "Weighted pyramid linking for segmentation of fully-polarimetric SAR data." The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII, Part B7, pp. 95-99, 2008.

[29] Tison, C.; Nicolas, J.-M.; Tupin, F.; Maître, H., "A New Statistical Model for Markovian Classification of Urban Areas in High-Resolution SAR Images." IEEE Transactions of Geoscience and Remote Sensing, vol. 42, no. 10, pp. 2046-2057, 2004.

[30] Majd, M.S.; Simonetto, E. & Polidori, L., "Maximum Likelihood Classification of Single High-resolution Polarimetric SAR Images in Urban Areas." Photogrammetrie - Fernerkundung - Geoinformation 2012, vol. 4, pp. 395ff, DOI: 10.1127/1432-8364/20/0126, 2012.

[31] Freitas, C.C.; Frery, A.C. & Correia, A.H., "The Polarimetric G Distribution for SAR Data Analysis." Environmetrics, vol. 16, no. 1, pp. 13-31, 2005.

[32] Jäger, M.; Reigber, A. & Hellwich, O., "A Non-Parametric Texture Descriptor for Polarimetric SAR Data with Applications to Supervised Classification." Proceedings of EUSAR 2014, pp. 830-833, 2014.

[33] Sliuzas, R.V.; Mboup, G. & de Sherbinin, A., "Report of the Expert Group Meeting on Slum Identification and Mapping." CIESIN, UN-Habitat, ITC: Enschede, The Netherlands, 2008.

[34] Kohli, D.; Sliuzas, R.V.; Kerle, N. & Stein, A., "An ontology of slums for image-based classification." Comput. Environ. Urban Syst., vol. 36, pp. 154–163, 2012.

[35] Taubenböck, H. & Kraff, N.J., "Das globale Gesicht urbaner Armut? Siedlungsstrukturen in Slums." SpringerSpektrum. pp. 107-119., 2015.

[36] Schmitt, A. & Brisco, B., "Wetland Monitoring Using the Curvelet-Based Change Detection Method on Polarimetric SAR Imagery." Water, vol. 5, pp. 1036-1051, 2013.

[37] Schmitt, A.; Wendleder, A. & Hinz, S., "The Kennaugh element framework for multi-scale, multi-polarized, multi-temporal and multi-frequency SAR image preparation." ISPRS International Journal of Photogrammetry and Remote Sensing, vol. 102, pp. 122-139, 2015.

[38] Moser, L.; Schmitt, A.; Wendleder, A. & Roth, A., "Monitoring of the Lac Bam Wetland Extent Using Dual-Polarized X-Band SAR Data. Remote Sensing." vol. 8 (302), pp 1-31. DOI: 10.3390/rs8040302 ISSN 2072-4292, 2016.

[39] Schmitt, A., "Multiscale and Multidirectional Multilooking for SAR Image Enhancement." IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 9, pp. 5117-5134, 2016.

[40] Sieg, T., "The potential of interferometric and polarimetric SAR data to characterize urban areas at the example of Mumbai and Manila." Master thesis, Global Change Ecology (M.Sc.), University of Bayreuth, *Germany, 2015.*

[41] Schmitt, A., "Histogram Correlation in the Context of Land Cover Classification from Multi-Polarized SAR Data." IEEE International Geoscience and Remote Sensing Symposium 2015. Milano, Italy, 2015.

[42] Schmitt, A., Wendleder, A., Murnaghan, K., Brisco, B. & Poncos, V., "Multi-sensor wetland mapping over the Peace Athabasca Delta." European Conference on Synthetic Aperture Radar (EUSAR) 2016, 7.6.-9.6.2016, Hamburg. VDE Verlag GMBH Berlin Offenbach, Germany. ISBN 978-3-8007-4221-9 ISSN 2198-3909.

[43] Ullmann, T.; Banks, S.N.; Schmitt, A. & Jagdhuber T., "Scattering Characteristics of X-, C- and L-Band PolSAR Data Examined for the Tundra Environment of the Tuktoyaktuk Peninsula, Canada." Applied Sciences, vol. 7(6), p. 595ff, doi: 10.3390/app7060595, 2017.

[44] Walterscheid, I.; Ender, J.H.G.; Brenner, A.R.; Loffeld, O., „Bistatic SAR Processing and Experiments." IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no. 10, pp. 2710-2717, 2006.

VII. APPENDIX

Analytical derivations of the normalized similarity measure are summarized in the appendix. They are indispensable for the justification of the chosen approach. For the sake of simplicity, local PDFs are symbolized by l and model PDFs of a certain class by m whereas n denotes the number of measurements and b the number of bins in the histogram.

A.1 *Derivation of the normalized similarity measure via the χ^2-test*

The χ^2-similarity measure for the comparison of two discrete distributions is defined as sum of the squared differences $(L_i - M_i)^2$ divided by the frequencies of the target distribution M_i. The number of observations n is equal in both histograms and thus, can be excluded. The squared difference is dissolved into the sum of products. As the remaining normalized frequencies l_i and m_i both sum up to one, the last two summands finally result in minus one.

$$\chi^2(b-1) \sim \sum_{i=1}^{b} \frac{(L_i - M_i)^2}{M_i} = \sum_{i=1}^{b} \frac{(n \cdot l_i - n \cdot m_i)^2}{n \cdot m_i} = \sum_{i=1}^{b} \frac{n^2 \cdot (l_i - m_i)^2}{n \cdot m_i} = \sum_{i=1}^{b} \frac{n \cdot (l_i - m_i)^2}{m_i} = n \cdot \sum_{i=1}^{b} \frac{(l_i - m_i)^2}{m_i} = n \cdot$$

$$\sum_{i=1}^{b} \frac{l_i^2 - 2l_i m_i + m_i^2}{m_i} = n \cdot \sum_{i=1}^{b} \frac{l_i^2}{m_i} - 2\frac{l_i m_i}{m_i} + \frac{m_i^2}{m_i} = n \cdot \sum_{i=1}^{b} \frac{l_i^2}{m_i} - 2l_i + m_i = n \cdot \left[\sum_{i=1}^{b} \frac{l_i^2}{m_i} - 2\sum_{i=1}^{b} l_i + \sum_{i=1}^{b} m_i \right] =$$

$$n \cdot \left[\sum_{i=1}^{b} \frac{l_i^2}{m_i} - 2 + 1 \right] = n \cdot \left[\sum_{i=1}^{b} \frac{l_i^2}{m_i} - 1 \right] \tag{A.1}$$

The number of observations is playing a key role in the χ^2-test: the higher the number of independent observations, the higher the reliability of the derived PDF. For low numbers the χ^2-test even should be replaced be the Exact Fisher Test. In image processing, this case occurs very seldom. Furthermore, the similarity of a local and a representative histogram increases with rising number of observations anyways. Thus, the number of observations can be canceled. Adding one to the relation reveals that the following relation where χ^2-distribution with $b - 1$ degrees of freedom, i.e. b bins considered.

$$\frac{\chi^2(b-1)}{n} + 1 \sim \sum_{i=1}^{b} \frac{l_i^2}{m_i} \in [1, \infty[\tag{A.2}$$

Because of the data range of the χ^2-distribution which is $[0, \infty[$, the sum over the quotients of the squared local histograms by the representative sample histograms must obtain positive values higher than one. Reversely, the reciprocal shares the same properties of a typical probability value.

$$s = \left\{ \sum_{i=1}^{b} \frac{l_i^2}{m_i} \right\}^{-1} \in [0,1] \tag{A.3}$$

A.2 *Derivation of the normalized similarity measure via the weighted harmonic mean*

The harmonic mean is defined as the Hölder mean with parameter -1. For positive values, it is the one next to the minimum value out of the typical means (arithmetic, geometric, and harmonic). It is ideally suited for relative measures like normalized frequencies. In the weighted version, is composed of the quotient of the sum over the weights divided by the sum over the weights divided by the values

$$\text{weighted harmonic mean} = \frac{\int \text{weights}}{\int \frac{\text{weights}}{\text{values}}} \tag{A.4}$$

Assuming the similarity measure in Eq. 8 can be interpreted as harmonic mean of the representative sample PDF weighted by the local PDFs which are introduced as squared frequencies in order to enhance typical features in the PDF. A further constant c is necessary to normalize the mean value to one

$$s = \frac{\int l^2(x) dx}{\int \frac{l^2(x)}{m(x)} dx} \cdot c = \begin{cases} 1 & l \equiv m \\ [0,1[& \text{else} \end{cases} \tag{A.5}$$

Let l be identical to m. In this case, the denominator simplifies to the integral over the local PDF which is one. The enumerator stays unchanged. Their product is required to be one.

$$s = \frac{\int l^2(x) dx}{\int \frac{l^2(x)}{l(x)} dx} \cdot c = \frac{\int l^2(x) dx}{\int l(x) dx} \cdot c = \frac{\int l^2(x) dx}{1} \cdot c = \int l^2(x) dx \cdot c = 1 \qquad \text{if } l \equiv m \tag{A.6}$$

Consequently, the constant c must result in the reciprocal of the initial enumerator $\{\int l^2(x) dx\}^{-1}$ and the similarity measure unfolds to

$$s = \left\{ \int \frac{l^2(x)}{m(x)} dx \right\}^{-1} \tag{A.7}$$

which is the same as derived from the χ^2-test in Eq. A.3.

A.3 *Efficient implementation of the normalized similarity measure via matrix multiplication*

The similarity measure has been derived from the harmonic mean by the help of integral notation and from the χ^2-test via sum notation for the comparison of two PDFs. In practice, a set of q local PDFs has to be compared to a set of r representative histograms of different classes. In order to avoid the implementation using inefficient loops, a very simple calculation based on matrix multiplication is derived.

$$s_{q,r} = \left\{ \int \frac{l^2(x)}{m(x)} dx \right\}_{q,r}^{-1} = \left\{ \sum_{i=1}^{b} \frac{l_i^2}{m_i} \right\}_{q,r}^{-1} = \frac{1}{l_{q \times b}^2 \times \frac{1}{m_{b \times r}}} \tag{A.8}$$

The local PDFs are stored in a matrix $l_{q \times b}$ with q the number of pixel environments under study and b the number of bins. The representative sample PDFs are stored in a matrix $m_{b \times r}$ with b again the number of bins and r the number of classes. The diagonal cross $\times$ stands for the matrix multiplication. Divisions are performed entry by entry. In that way, a whole image can be transformed into multiple class properties with minimal time effort.

A.4 *Extension of the normalized similarity measure to multi-channel images*

So far, the whole theory bases on the evaluation of single layer images. With respect to remote sensing applications, multi-channel images or even multi-source data sets are the standard nowadays. Therefore, the formulation is extended exemplarily via sum notation to three dimensions in the following.

$$\frac{1}{s} = \sum_{x=1}^{b_x} \sum_{y=1}^{b_y} \sum_{z=1}^{b_z} \left\{ \frac{l_{x,y,z}^2}{m_{x,y,z}} \right\} = \sum_{x=1}^{b_x} \sum_{y=1}^{b_y} \sum_{z=1}^{b_z} \left\{ \frac{(l_x \cdot l_y \cdot l_z)^2}{m_x \cdot m_y \cdot m_z} \right\} = \sum_{x=1}^{b_x} \sum_{y=1}^{b_y} \sum_{z=1}^{b_z} \left\{ \frac{l_x^2 \cdot l_y^2 \cdot l_z^2}{m_x \cdot m_y \cdot m_z} \right\} =$$

$$\sum_{x=1}^{b_x} \sum_{y=1}^{b_y} \sum_{z=1}^{b_z} \left\{ \frac{l_x^2}{m_x} \cdot \frac{l_y^2}{m_y} \cdot \frac{l_z^2}{m_z} \right\} = \sum_{x=1}^{b_x} \left\{ \frac{l_x^2}{m_x} \right\} \cdot \sum_{y=1}^{b_y} \left\{ \frac{l_y^2}{m_y} \right\} \cdot \sum_{z=1}^{b_z} \left\{ \frac{l_z^2}{m_z} \right\} = \prod_{d=1}^{n_d} \sum_{i_d=1}^{b_{i_d}} \left\{ \frac{l_{i_d}^2}{m_{i_d}} \right\} \tag{A.9}$$

As all layers are assumed to be uncorrelated, the joint PDF can be written as the product of the marginal PDFs along the three dimensions. Thus, the similarity measure over d dimensions equals the product of the similarities of each dimension. The dimensions are completely independent, i.e. the number of bins may arbitrarily vary. As the similarity necessarily lowers with each dimension, a normalization based on the geometric mean s^{1/n_d} is recommended for comparative and normalization purposes, e.g. to compare class probabilities derived from a two layer image to class probabilities derived from a ten layer image.