Detecting social groups from space – Assessment of remote sensing-based mapped morphological slums using income data

M. Wurm a* and H. Taubenböck a

aGerman Remote Sensing Data Center (DFD), German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany

*Michael Wurm: michael.wurm@dlr.de; German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany. orcid.org/0000-0001-5967-1894

4446 words including abstract, manuscript and references.

“This is an Accepted Manuscript of an article published by Taylor & Francis Group in Remote Sensing Letters 26/10/2017, available online: http://www.tandfonline.com/10.1080/2150704X.2017.1384586
Detecting social groups from space – Assessment of remote sensing-based mapped morphological slums using income data

Over the last decades, massive urbanization processes in the world lead to the emergence of large slum areas making them home to about a seventh of the global population. Although the variety of morphological characteristics varies significantly within as well as across cities, common determinants exist. Informal, or unplanned settlements in particular, do show very similar morphologies over the world. They are characterized mostly by extremely high building densities and small building sizes, irregular arrangement of buildings and street network and are often located at exposed sites in the city. Based on these characteristics, we deploy satellite images for a systematic mapping of morphological slum areas in the city of Rio de Janeiro, Brazil based solely on physical characteristics and analyse the mapping result with the official census data. Outcomes show first that morphological slums are a semantic and spatial sub-group of all slum areas contained by the Brazilian census and that remote sensing-based mapping yields accuracies of almost 94%. Second, analysis of census-based income data proofs that while almost 45% of all mapped slum blocks are characterized by incomes below the poverty line, as defined by the Organisation for Economic Co-operation and Development (OECD), this holds true for only about 6% of the formal urban neighbourhoods.

Keywords: morphological slums, favelas, informal settlements, census, income

1. Introduction

Cities, as the human habitat, are places of constant motion. While the observed processes of global urban change are mostly positively connotated for developed countries (e.g. productivity, quality of life, access to infrastructure, education), these positive impacts have not only bypassed a large part of the global urban population – about one seventh of the global population suffers from precarious living conditions (Arimah 2010) – their impacts in developing countries are occasionally severe: due to rapid urbanization processes and resulting insufficient provision of housing, inefficient urban planning and low state investments in urban infrastructure, informal settlements emerge to a significant extent in cities in developing countries (Davis 2006). This emergence is reinforced through population growth at a high pace as well as through intensive migration, especially from people with a low socioeconomic status who accept sub-standard accommodation (UN-Habitat 2015). Slums can be seen as the most visible expression of poverty (Amnesty International 2016). However, the ways of treating slums in cities from authorities varies also significantly: formerly, ignorance, eradication, eviction and relocation were common methods to deal with the ‘invisible city’; however, the value in slum upgrading has been recognized more recently as it has been also targeted as one of the Millennium Development Goals by the United Nations (UN 2010). In this sense, spatial knowledge on the physical presence and the geographic locations, extents and structures of slum areas in cities are of crucial importance towards an integration of informal urban areas into governing structures and thus a sustainable urban development.

1.1 The physical face of slums: a remote sensing perspective

The names of slums vary strongly across countries, e.g. favelas, barriadas, shantytowns,
informal/spontaneous/marginal/squatter settlements, gecekondu, ashwa’iyyat, bidonville, township, etc., and so do also their physical manifestations, featuring a wide range of built-up structures (Kuffer et al. 2017). On a general level, however, they basically all describe deprived urban areas with low living standards characterized by verbal descriptions of qualitative conditions such as poor structural quality of housing, overcrowding or inadequate access to clean water and sanitation and also insecure residential status (UN-Habitat 2015). However, the morphology of slums does not follow an agreed concept or definite definition. For a long time only very little systematic research addressing the physical appearance of slums was conducted (Hofmann 2001), although more recently the scientific interest in remote sensing-based mapping slums has experienced a strong increase (Kuffer et al. 2016). Alongside, an ongoing discussion on the physical characteristics of the morphology of slums has been triggered, mostly agreeing that their physical appearance is characterized by high building densities and unstructured arrangement of buildings, both often in sharp morphological contrast to formally built-up areas located in close spatial vicinity (Taubenböck & Kraff, 2014; Figure 1). From a remote sensing perspective, these physical characteristics can be exploited in recently accessible high resolution satellite data. Even though mapping of slum areas using remote sensing data can usually not entirely replace ground-based surveys such as censuses (Kohli et al. 2012) but some relevant features in satellite images can be directly associated with socioeconomic variables (e.g. Sandborn and Engstrom 2016; Jean et al. 2016; Sliuzas et al. 2008, Taubenböck et al. 2009). Typically, significant orientation of buildings and street networks were used as physical proxy information in satellite images for slum mapping, as well as characteristic small building sizes and densities besides derived image descriptors such as textural and morphological features (Wurm et al. 2017a&b).

Especially the advent and the easier accessibility of very high resolution (VHR) optical satellite images since the beginning of the 2000’s triggered remote sensing-based studies on slum mapping because small-structured urban features can be identified in these data sets. For extensive reviews on literature on remote sensing-based slum mapping we refer to related work (e.g. Wurm et al. 2017b; Kuffer et al. 2016). While many of the works related to slum mapping focus on automated methods, lately, visual interpretation of VHR satellite imagery has been used to derive maps of slums and their built-up characteristics in highest detail and accuracy (e.g. Angeles et al. 2009; Gruebner et al. 2014; Gunter 2009, Taubenböck and Kraff 2014). Furthermore, while visual interpretation is considered being very labour-intensive on the one hand, it is also considered less vulnerable to mapping errors than automatic image classification and especially additional morphologic characteristics such as built-up density and average building height can be mapped as well. Nevertheless, one major problem in assessing the quality of automatic satellite image-based slum mapping is the lacking of adequate reference data (Kuffer et al. 2016). First, unbiased reference data does not exist for many urban areas comprising slum settlements because many local governing authorities do not map or keep updated spatial records on slum areas in their cities (UN-Habitat 2003). Second, the definition of the term ‘slum’ underlies strong ambiguity and addresses many non-physical aspects (e.g. legal issues on land property, insecure residential status), respectively. Furthermore, studies on slum mapping using remote sensing data have to deal with the issues of varying performances of automatic image classifiers: the majority of computer-based mapping of slum areas achieves accuracies between 80-90% (Kuffer et al. 2016).
Fig. 1: Characteristic morphology of slum areas in different cities in the world, embedded in formally developed urban neighborhoods: Khayelitsha (Cape Town), Kibera (Nairobi), Dharavi (Mumbai) and Paraisópolis (Sao Paulo).

1.2. Assessing the socioeconomic housing status by physical characteristics

Qualitative evaluation of the status of housing in terms of semantic labelling by means of earth observation and subsequent establishing of a quantitative relationship with socioeconomic characteristics is sought in the scientific literature. Studies are based on e.g. household income (Taubenböck et al. 2009), population data (Sandborn and Engstrom 2016), urban poverty (Engstrom et al. 2017) and local economic outcome (Jean et al. 2016). In the light of these aspects, we propose performing an assessment of the capabilities of VHR remote sensing data to identify a socioeconomic group solely by characteristic morphological properties such as building density and building orientation in contrast to the varying morphologic structure of the surrounding formal urban neighbourhood (Taubenböck and Kraff 2014). To do so, we assess visual mapping capabilities of slum areas for an entire mega city, exemplary by Rio de Janeiro, Brazil and analyse their socioeconomic characteristics in terms of household income. The socioeconomic data are provided via the Brazilian Census (Censo Demográfico) and are available for slum areas and formal residential areas. The analysis specifically aims at providing a quantitative evaluation of mapping capabilities of a distinct socioeconomic class: slums. Only through meaningful assessment of the capabilities of satellite images, their global utilization for slum mapping can be explored.

2. Favelas in Rio de Janeiro: Data and Methods

The presence of favelas – officially named subnormal agglomerations/aglomerados subnormais - in Rio de Janeiro can be traced back in time of more than a century and
today they represent a non-negligible share of the cariocas, the citizens of Rio de Janeiro: out of 6.3 million cariocas in 2010 about a quarter is estimated by the Instituto Brasileiro de Geografia e Estatística (IBGE) to live in slums. A precise counting of slum dwellers is difficult, also because of the ambiguity what exactly constitutes a ‘slum’.

This ambiguity comes to light when looking at the varying types of slums in Rio: favelas, invasoes, loteamentos and cortiços (UN Habitat 2003). Favelas and invasoes are both self-built and established illegally on occupied land without following any rules of urban planning. They only differ in their state of consolidation, whereas favelas are located all over the city and invasoes can mainly be found on disadvantaged spots like riverbanks, swamps, steep hills or alongside roads. Loteamentos are most likely described by their illegal character and are especially located in the eastern part of the city. All these mentioned types of slums lack of basic infrastructure and services and tend to be located in mainly naturally disadvantaged areas. The fourth group, cortiços, are mostly characterized by their very bad living conditions which can be found in older, shabby buildings, mainly located in the old city centre.

Unambiguous classification of the above described categories of slums in Rio is challenging as they do not consistently feature distinct morphological appearances. In consequence, the visual identification in remote sensing images is challenged by untypical morphological characteristics such as deprived formal buildings in the city centre or informal areas which are built in regular, geometric order. Thus, a remote sensing based mapping of slums can only focus on settlements which are morphologically significant in terms of high building density, irregular building arrangements, and unprivileged, exposed location in the city.

2.1 Remote sensing-based mapping of morphological slums in Rio

The study area is the municipality of Rio de Janeiro with an area of 1,200 km². For the delineation of the morphological slums, visual interpretation of the very high resolution (1 m and higher) Imagery-basemap included in ESRI ArcMap GIS software package was performed. The slums were identified according to the differing morphological characteristics (building density and size, arrangement and location, street network, building materials) compared to their surrounding urban neighbourhoods (Fig. 2).

![Fig. 2. (a) Examples of the mapped slum blocks of favelas/invasoes (white outlines) in Rio de Janeiro and (b) estimated classes of building density (shown as percentages). These classes were subsequently consolidated into three classes: (b-1) 0%-40%; (b-2) 40%-60%; (b-3) 60%-80%.](attachment:image)

Slum-mapping was performed on the spatial level of the city block at an average area of 1.6 ha. Several adjacent urban blocks identified as slums form an entire favela. They are
usually circumscribed by significant street networks or natural boundaries such as water surfaces, vegetated areas or open land. A minimum distance of 10 m was defined for blocks being treated as individual spatial blocks. Mapped slums integrated both Favela and Invasoes which are of a very similar morphologic type and are often located at steep slopes. The categories Loteamento and Cortiço were omitted from mapping since their morphological structure (e.g. regular road networks, structured, regular building orientation) does not differ from surrounding formal urban neighbourhood (Fig. 3-C). In total, 1210 slum blocks were mapped which are grouped to 479 favelas.

After identification and mapping of the slum blocks over the study area, in a subsequent step, building density was derived for each slum blocks. It is defined by the ratio of cumulated building footprint areas to the associated slum block area. Thus, the building density is visually estimated during image interpretation by assigning one of three density classes: 0%-40%, 40%-60% and 60%-100% (Fig. 2). In the same way as building density, also the average building height is derived by the average amount of storeys for buildings within each city block. Additionally, Google Street View images and Google Earth perspective views were used for the building height assessment. Overall, two categories were assigned to every polygon: 1-2 storeys as well as 3-5 storeys. In the remainder of this article we refer to the remote sensing-based identification of slums as morphological slums.

2.2 Brazilian Census Data

Every 10 years, for the last time in 2010, a comprehensive census in Brazil is undertaken with the aim of gaining information on all households about their socioeconomic characteristics such as population, income and labour. For performing the census, the entire country was spatially sub-divided into 317,000 census blocks, which are the smallest spatial unit. Out of them, about 16,000 (5 %) were classified as aglomerados subnormais – slums – incorporating all four slum types of Rio (favelas, invasoes, loteamentos, cortiço). According to IBGE definition (IBGE 2011), slum areas contain at least 51 dwellings (e.g. houses, shacks, tents), they are illegally constructed on public or private land, their morphology is mostly characterized by irregular street networks and shapes as well as they feature small building sizes and deprived areas lack sanitation and waste disposal. Census variables provide detailed information on the household income per census sector. For the subsequent analysis we used the variable ‘Nominal average monthly income of persons responsible for permanent private households (with and without income)’ for modelling the socioeconomic status which is referred to as household income in the remainder of this article.

3. Analysis of urban structure and socioeconomic status

In this section, we aim at establishing a ‘socioeconomy-morphology relationship’ between identified slum areas and their household income.

3.1 Mapping capabilities of remote sensing-based slum detection

The total area of the census blocks cumulate to 826.3km² ($A_{census}$) out of which 771.1 km² ($A_{formal}$; 93.3%) relate to formal urban census blocks (non-slums) and 55.2 km² relate to census slums ($A_{slums}$; 6.7%). By means of remote sensing identified morphological slums cumulate to 18.6 km² ($A_{morph}$) which have a mutual spatial overlap.
of 17.0 km² (\(\text{AM}_{\text{slums/morph}}\)) with the census slums and a mutual spatial overlap with formal urban areas of 1.6 km² (\(\text{AM}_{\text{census/morph}}\)). As regards with this first spatial assessment, we can state that to a very high degree of 91.6% the morphological slums areas agree with census slums and only a very small share of the morphological slums overlaps with formal urban development (\(\text{AM}_{\text{census/morph}}\); 8.4%). Out of this area, a total of 0.4 km² (25.0%) of \(\text{AM}_{\text{census/morph}}\) are formed by very small geometric fragments (<1200 m²; size defined by the smallest area of a regular census slum tract) of slum block outlines due to ambiguous slum block borders between mapped morphological slums and census slums (Fig. 3(a)). Reducing \(\text{AM}_{\text{census/morph}}\) to only the mis-identified slum blocks with area ≥ 1200m², the re-calculated value for \(\text{AM}_{\text{census/morph}}\) accounts for 1.2 km² (6.3%) indicating an extremely small error of commission or a very high user’s accuracy (93.7%). Concluding this part of the assessment, a remote sensing approach based on the building morphology of urban areas, allows for consequent identification of informal slum areas across the city.

As regards with the share of \(\text{AM}_{\text{slums/morph}}\) from the total area of census slums \(\text{A}_{\text{slums}}\) we find that only 30.8% of its area are detected by remote sensing. Some reasons for this large error of omission can be found in the varying morphological structure of census slums and its sub-categories described in section 2 and 2.2 (favelas, invasoes, loteamentos, cortiço) making it almost impossible to differentiate with formally built-up urban areas (e.g. loteamentos vs. regular urban neighborhoods; Fig. 3-(c)). A second reason can be found in the fact that identified morphological slums follow most of the times sharp boundaries between morphological slum development and e.g. formal urban development or natural boundaries such as rivers, coasts, forests etc. while census

![Fig. 3. Examples of agreement/disagreement between mapped morphological slums (yellow) and census slums (red). Mutual overlap between both data is with grey background. (a): displays variations in ambiguous slum block outlines and apparent non-slum areas which are comprised by the census slum blocks; (b): displays large natural areas comprised by census slum blocks; (c): visualizes the slum category “Loteamenton” in the census slums which cannot be distinguished from formal urban development in direct spatial vicinity solely based on morphological criteria.]
blocks may comprise also large areas of undeveloped land or even industrial areas causing this deviation from morphological slums area (Fig. 3-(b)). As a third reason for the identified spatial deviations between morphological slums and census slums, total accumulation of small spatial ambiguities in between drawn borders of slum blocks (identical to the above described areal differences between formal development and morphological slums) contribute as well to the error of omission. For a graphical depiction of these discussed spatial variations in census slums and morphological slums, see Fig. 3.

3.2 Income analysis

In the following, variations of the household income are analysed with respect to the specific morphology of the urban structure. In a first analysis, the distribution of household income for morphological slums \(n=542\) and formal urban development \(n=7999\) is evaluated using box plots (Fig. 4a). Data reveal a median household income of 673 Brazilian real (R$) and an interquartile range (IR: 25\(^{th}\) quantile-75\(^{th}\) quantile) between 585-787 R$ for morphological slums. In general, the median value indicates a very low income level in morphological slums. But, more significantly, the low spread of IR also reveals a very homogeneous distribution of household income. In contrary, the median income for households in formal urban developments is 1615 R$ with an IR between 1055-3269 R$. This reveals a significantly higher income for households in formal neighbourhoods in contrast to households in morphological slums. It also reveals a large income spread marked by the large IR. Comparing both groups by the household income, it becomes obvious that the population living in morphological slums is much more homogenous in terms of their socioeconomic characteristics than the entire population residing in all kinds of formal urban neighbourhoods.

Fig. 4. Box plots visualizing the distribution of the household income for varying types of urban morphology (morphological slums, formal urban development), density and number of storeys.

Considering the definition of the poverty line, as defined by the Organisation for Economic Co-operation and Development (OECD), as being the value of half the median household income of the total population (OECD 2017), we deploy a city-specific poverty line for Rio de Janeiro derived from the census data of 639 R$. This
value agrees widely with the identified median household income for morphological slums. Out of all 479 mapped slums, 214 (44.7 %) are characterized by a lower average household income than the identified poverty line. Additional 155 slums (32.4%) are characterized by incomes just above the poverty line and below Q75.

For the formally developed urban areas, only 5.5% (443 census tracts) of the total of 7999 census tracts comprise an average household income below the poverty line. Thus, a significant prevalence of poor households can be found for the morphological slum areas while only a small number of poor households are found in formal areas. However, it also shows that in formally developed urban neighbourhoods still urban poverty is found, revealing the spatial complexity of societies. As presented in section 2.1, mapping of morphological slum households included not only a semantic labelling of urban neighbourhoods but also a quantitative estimation of two-dimensional building density (2D) and the number of storeys (3D). Both are morphological measures of density and may serve as proxy for population density, which is positively associated with more economic opportunities and hence higher income (Usman and Mahmood, 2011). While the analysis of both measures as depicted in Fig. 4b-4c underlines this assumption that for both aspects of density (2D and 3D) a positive trend for higher densities can be observed regarding higher household income. However, some authors (O’Hare and Barke 2002) question general models of linkages between the age and densities of squatter settlements since the assumption is that older slums are expected to have a greater tendency towards self-improvement.

Overall, these observed correlations between urban morphology and socioeconomic status proof that remote sensing data may be the only and ultimate data source to map urban poverty at global scale. However, understanding of the mutual interweaving complexity of morphologic variations and socioeconomic hybridity across local settings is a prerequisite for remote sensing to fulfil its promises.

4. Conclusions

In this paper, we demonstrated the capabilities for the spatial identification of a specific socioeconomic group using satellite imagery. Mapping of economically disadvantaged urban areas such as slums or favelas (defined by household income) was performed based on their very distinct and morphologically characteristic physical built-up appearance – very high building densities, small building sizes, irregular building orientation, exposed or hazardous locations within the urban areas etc. Official data from the Brazilian Census were used as spatial reference for assessing the mapping capabilities of slums in satellite images. The comparison with the census reveals that remote sensing-based mapping of morphological slums yields very high agreements of 93.7%. The high value of remote sensing in mapping a specific social group is acknowledged in this context. The comparison with census-based household income supports these findings: while 44.7% of the morphological slums have incomes below the poverty line, this holds only true for only 5.5% of the formal urban neighbourhoods. As regards with data reliability of socioeconomic data, the current study focused on publicly available official data provided by national authorities (IBGE) which was considered a good trade-off between trustworthiness and data accessibility. While it can be assumed that the informal sector in economy has a larger impact on the household income in slum areas than in formal urban neighbourhoods, the official census data, however, remain the only available data source on a large scale. Since the physical characteristics used for morphological slums are found to be very similar for slum areas throughout the world (Fig. 1), a remote sensing approach for mapping slum areas
worldwide seems at least from a conceptual aspect feasible. A large-area, or even global inventory of remote sensing-based mapping or monitoring of poverty, however, is not only due the extensive mapping procedures extremely laborious (manual and even automatic), but it is also subject to very large errors of omission. These occur less due to the identified geometric deviations between mapped slums and census slums but because of the very ambiguous definition of slums in general. In the current study we focused on morphologically clearly identifiable urban structures which can be related to poverty, but in contrary, physical structures of poverty can be characterized by a much wider heterogeneity. A solely remote sensing-based approach will be hardly able to overcome these identified obstacles in differentiation of morphologically similar structures between slums and formal urban development, however, promising current and future approaches deal with the integration of social media data for semantic enrichment of geo-spatial data (Klotz et al. 2017).

While this paper proofs that remote sensing data hold the capability to identify urban poverty, research must gain a significantly broader understanding of the mutual interweaving complexity of morphologic variations across cities and socioeconomic hybridity across local settings. Only then, a comprehensive global approach of assessing locations of urban poverty will be feasible.

Disclosure statement

No potential conflict of interest was reported by the authors.

Acknowledgements

The authors would like to thank J. Fricke and M. Weigand for their support.

Funding

This work was supported by the H2020 European Research Council [[714087] - So2Sat].

References


