Basin-wide water demand estimation using RGB color detection of built-up area from satellite imagery

W K Lee¹, S I Syed Yusoff²

¹ Faculty of Civil Engineering, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia.
² Airis Engineers Sdn. Bhd. A2-876, Leisure Commerce Square, 46150 Petaling Jaya, Selangor, Malaysia.
Corresponding author: leewei994@uitm.edu.my

Abstract. The unavailability of as-built drawings and population data in primarily rural east central Java, Indonesia, presents a challenge to effectively estimate the total demand of a proposed large-scale retrofit water supply scheme. A simple RGB color detection method is employed to analyse satellite imageries extracted from Google Earth. The applicable range of RGB spectrum for typical rooftops is ascertained based on image inspection and the candidate areas identified are filtered, smoothened and manually cleaned-up to produce the total built-up area by pixel calibration. Water demand is then estimated based on assumed population equivalent per standard residential unit and consumption rate. The preliminary estimate is further corroborated with the most recent census and population growth rate. Result show that the approach taken is a cost-effective and efficient way for large-scale areal approximation.

1. Introduction
Satellite imagery is widely accessible these days via online mapping services such as Google Earth (or Google Map). Extraction of valuable information from these still imageries serves a range of purposes in various sectors and its integration with geographical information system (GIS) databases further opens up a wide spectrum of potential useful applications. In particular, satellite image processing allows large areal extent to be examined without the need for physical site visit or aerial surveillance, which may be prohibitive due to cost or accessibility. Besides, the insight afforded from the inspection of aerial imagery over a vast area is unmatched by any other technology currently available.

Cartographic features extraction from satellite images has been an active research topic in photogrammetry and computer vision since 1990s. While some literatures focus on the processing of LIDAR data, i.e. data collected using airborne laser scanning (e.g. [1,2]), most researchers consider high-resolution satellite imagery. [3] extracts urban building from high-resolution satellite imagery by integrating the results from structural, contextual, and spectral information. [4] focuses on building geometry, including boundary delineation and roof compositional polygon segmentation. Hough transform has been widely used for edge detection to extract the lines or curves of the objects [e.g. 5]. [6] utilises shadow and color information to further verify the extracted line segments to generate building hypotheses. Shadow information has also been used to detect building height [7].

Satellite images at appropriate altitude shows rooftops and building shapes in the aerial view of Indonesia which are clear to human eye owing to their distinctive color compared to the surrounding.
Although in some cases the building shapes or color may be ambiguous, or blocked by tree canopy or cloud cover, it is possible to efficiently identify these built-up areas instead of costly site reconnaissance. In this study, an estimation of the water demand in three regions in Lamongan Regency, Indonesia is sought for the preliminary planning of a large-scale retrofit water supply scheme targeted at the existing built-up area. However, population census data is not specific to the area considered and as-built drawings are also unavailable. In view of the large spatial extent of the area covered, an image processing and analysis tool is employed to automate the process to provide quick and dependable estimation of the built-up area, the population equivalent (PE) and hence the total water demand.

2. Study Area

Lamongan Regency in east central Java, Indonesia, is partially located within the downstream lowland section of Bengawan Solo watershed (110° 27’ 16” - 112° 39’ 27”E; 6° 37’ 05” – 8° 17’ 27” S) (Figure 1a) [8]. Bengawan Solo is the longest river in Java Island, measuring approximately 600 km, and has a catchment of size 16.100 sq.km equivalent to 12.3% of Java Island. The river flows through the provinces of central and eastern Java. The area is primarily rural where clusters of towns are scattered over vast agricultural land. The land use comprises 32.4% settlements, 21.5% paddy field, 25.3% other agriculture, and 19.2% forest in year 1992 [8]. The northern coastal plain of Lamongan facing the Java Sea belongs to the coastal watershed and has a higher density of built-up area and population.

Figure 1. Study area.

The region has a mean annual precipitation of 2100 mm and mean annual runoff of 363 cumec at Bojonegoro (111° 53’E; 7° 11’S) (Figure 1a) based on data from 1983-1991. There are seven (7) dams across Bengawan Solo river with various storage capacities, which were built mainly for irrigation purpose [8]. The use of treated water supply is still limited, whereas groundwater utilization is common for domestic and agriculture need.

Presently, two (2) new water treatment plants (WTPs) are proposed along Sungai Bengawan Solo at Karanggeneng (112° 21’ 30.38”E; 7° 0’ 3.69”S) and Laren (6° 57’ 48.81”S; 112° 17’ 59.78”E), respectively. The proposed scheme targets at the populated built-up areas of Karanggeneng and its surrounding, East Laren, and the coastal plain in Paciran as demarcated in Figure 1b. Population estimation in 1993 for the entire Bengawan Solo region is 13.5 million. More recently, population census for Year 2002 to 2007 were reported [9]. However, data available is limited to the sub-districts (kecamatan) and district (kabupaten) levels only. Hence, no estimate specific for the targeted area is available.
3. Methodology

3.1. RGB thresholds

For the purpose of this study, the SimpleColorDetection version 1.4.0.0 published by ImageAnalyst in MATLAB Central File Exchange is adopted [10]. The script requires MATLAB Image Processing Toolbox and is tested with MATLAB R2008b and R2010a. The code detects objects of a certain color in the input image by separating the image into its component red, green, and blue (RGB) color bands. By defining a set of suitable low and high RGB threshold values (0 to 255), it finds the corresponding mask and displays objects of the defined color range.

Figure 2a shows the sample image in Google Earth. For consistency, all images are extracted at 1 km altitude (Figure 2b), where pixel density suffices for onscreen identification of the various features. Here, the sample image contains a range of objects with their distinctive color which characterise the aerial view of Indonesia. They comprise greenish tree canopy and meadows, dark blue ponds, a wide spectrum of brown for tilled or cleared ground surface, light and dark grey roads corresponding to aggregate and paved surfaces respectively, and rooftops which appear primarily in varying shades of red or dark brown, corresponding to the common colored concrete roof tiles used in the country.

![Figure 2](image.png)

(a) Aerial view at 6 km altitude. (b) Sample input image at 1 km altitude.

Figure 2. Google Earth image.

The typical range of RGB values for the built-up rooftops is investigated using the MATLAB image viewer tool (Figure 3). Inspection by eye shows that the red color band is high for most rooftops (up to 255). Some rooftops have green and blue colour band up to 180, but any value higher than that are more likely to be green trees or grass, or water body. There are some cases where buildings appear whitish in colour with low RGB values, suggestive of flat concrete type rooftop. However, these are relatively few in occurrence and is ignored herein. If considered, the full range of RGB colour band becomes applicable and the algorithm will pick up roads which has similar appearance.

Based on the observation above, it is concluded that the lower and upper threshold of the three color bands for rooftops are as follow: red (min. 50, max. 255), green (min. 50, max. 180), and blue (min. 50, max. 180). The distribution of the selected red, green and blue spectrum in the image are then identified and plotted in histograms (Figure 4). The distribution shows that the range defined for red, green and blue, respectively, are all widely detected throughout the image. In order to identify the rooftops correctly, all the three color bands defined is applied in conjunction with one another. Furthermore, based on observations of the RGB values in the image viewer, it is further specified that the building rooftops must have R value > G value.
3.2. Building detection

Using the defined threshold and criteria on the RGB values, the candidate rooftops are identified as shown in Figure 5a. Observation shows that the rooftops are reasonably well identified but some open/farmed land areas are also included due to similarity in RGB value range. This is overcome by excluding these zones where the presence of building is clearly non-existence. In addition, objects which are much too small or large compared to a standard residential building is filtered to give the improved results in Figure 5b.

Next, the algorithm smoothen the edge of the objects identified and fill in unidentified gaps in these objects. Smoothening is not very sensitive to the shape chosen, where diamond shape has been chosen to better approximate the shape of building oriented at any arbitrary angle. Large smoothening size leads to identification of smaller number of object but larger total identified area. For smoothening size of 5 pixels, a total of 110 blobs, equivalent to 2.697 million pixels were identified. For size of 10 pixels, 99 blobs equivalent to 2.799 million pixels were identified, where the difference is only +3.77%. Due to the negligible difference, the optimum smoothening size chosen for this study is 5 pixels. Note that the algorithm is not able to separate closely packed building and thus each blob
of area identified may represent a cluster of buildings/houses as seen in Figure 5. False-detection is manually clean-up where identifiable. The algorithm is also not able to reproduce strictly straight building edges, especially when blocked by tree canopies, or shadowed part of the building. For this reason, a 10% factor is applied to correct for shadow and canopy effect loss.

Figure 6a and 6b shows the original input image and the final identified rooftops respectively. The total number of blobs identified and the area of each blob (in terms of number pixels) can then be generated for the estimation of total built-up residential area.

3.3. Water demand estimation

For consistency of residential areas estimation, all the images extracted from Google Earth are taken at 1 km elevation. For area conversion, a standard residential unit is assumed to be 1400 sq.m, which is equivalent to 3900 pixels on the image upon calibration. Water demand is assumed to be 600 lpd per standard residential unit as defined, equivalent to 5 PE.

Owing to the fact that the code does not necessarily distinguish individual building blocks, the demand is calculated based on the total built-up area identified in lieu of the number of standard residential units identified. Since the code is not designed to distinguish institutional and multi-storey buildings, their water demand is not considered but may be factored in for individual villages or
communities where necessary (or where applicable). Water demand for agriculture and industrial use are not accounted based on the known prevalent practice of groundwater use. Lastly, the estimated demand is based on the most up-to-date (December 2018) satellite imagery from Google Earth, ignoring future population growth and development, which again, may be factored into the estimation accordingly.

4. Results and Discussions
Table 1 shows the summary of the total area inspected, built-up area detected (and its percentage compared to the area inspected), total equivalent residential unit, total population equivalent (PE), and total water demand in million litre-per-day (MLD) for Karanggeneng and its surrounding, East Laren and the coastal plain of Paciran.

| Region          | Area processed (sq.km) | Area detected (sq.km) | Equiv. unit | PE | Demand (MLD) | % |
|-----------------|------------------------|-----------------------|-------------|----|--------------|---|
| Karanggeneng    | 34.1                   | 2.284                 | 6.7         | 17,569 | 87,846       | 10.5 | 43.9 |
| East Laren      | 4.1                    | 0.435                 | 10.6        | 3,346 | 16,730       | 2.0  | 8.4  |
| Coastal Paciran | 14.7                   | 2.481                 | 16.8        | 19,087 | 95,433       | 11.5 | 47.7 |
| **Total**       | **52.9**               | **5.200**             | **9.8**     | **40,002** | **200,008** | **24.0** | **100.0** |

*includes additional areas in neighbouring sub-districts

Due to large extent of the land area involved and the altitude of the images extracted, image processing is carried out in batches only for selected inhabited areas. The total land area of the images processed is 52.9 sq.km. Based on the built-up area detected, the total equivalent standard residential unit is estimated at 40,000 and the total population equivalent is 200,000. Compared to the last census data in year 2007 [9], the population estimated for the coastal plain of Paciran at 95,433 (Table 1) is of the same order as the 92,177 recorded for Paciran sub-district (Table 2). Meanwhile, the population in East Laren is estimated to be 16,730, which is approximately 30% of the entire Laren sub-district. For Karanggeneng, the actual area considered in the present study extends beyond its boundary into three adjacent sub-districts, hence a higher estimate of 87,846, which is 1.8 times the number reported according to [9].

Data in [9] shows that population growth from year 2002 to 2007 in the three sub-districts are generally well below 2% annually except for year 2005 where an explosive growth of 9.0%, 15.0% and 17.9% is recorded in the sub-district of Karanggeneng, Laren and Paciran, respectively. Using year 2007 data, [9] projected the population for year 2011 to 2031 with a growth rate of 10.4% every 5 year. Adjusting for the areal extent of Karanggeneng and East Laren in the present study, the projected population for the three regions for year 2021 is 259,800. Meanwhile, our total population estimate for the three regions for year 2021 using the same growth rate is 212,500 which is lower by 18.2%. This may be attributed to the omission of multi-storey buildings and should be adjusted accordingly for the present purpose.

| Sub-district | Area (sq.km) | Population (population density) |
|--------------|--------------|---------------------------------|
|              | 2002         | 2007               | 2021*                  | 2031*                  |
| Karanggeneng | 96.0         | 42,896 (836)       | 48,643 (948)           | 64,232 (1252)          | 78,299 (1526)  |
| Laren        | 51.3         | 46,977 (489)       | 55,446 (578)           | 73,215 (763)           | 89,249 (930)   |
| Paciran      | 47.9         | 74,212 (1550)      | 92,177 (1925)          | 121,718 (2542)         | 148,373 (3098) |
| **Total**    | **195.2**    | **164,085 (841)**  | **196,266 (1005)**     | **259,165 (1328)**     | **315,921 (1618)** |

*Projection based on year 2007 data with 10.4% growth rate every 5 year
It is noted that the built-up area detected is only a small fraction (9.8% averaged) compared to the area processed because of the spatial scatter of the buildings. This is even smaller compared to the area demarcated in Figure 1, which is 1.5 times larger at approximately 80 sq.km. The greater surrounding agricultural and pristine land has not been accounted though there may be remotely isolated buildings. Based on the total area of 195.2 sq.km for the three sub-districts (Table 2), the average population density in year 2021 is 1,328 per sq.km. The density is relatively high by definition for rural communities but is a relative norm for the populous Java Island. Coastal Paciran has a much higher density of 2,542, whereas Laren is lower at 763.

Using the preliminary estimate of the population in Table 1, the total water demand is thus projected to be 24.0 MLD, where nearly half is in the more densely-populated Paciran area (47.7%), followed by the inland Karanggeneng area and its vicinity (43.9%), and the balance is in the adjacent East Laren area (8.4%). To account for the discrepancy due to omission of multi-storey residential units, the value is adjusted to 29.3 MLD (=24/0.82). Lastly, to cater to the future demand, a design factor of 1.2 is recommended for year 2031 population, or 1.4 for year 2041. Despite the unavailability of as-built drawings, the water supply network and final estimated localised demand is proposed based on satellite imagery extracted (Figure 7).

Figure 7. Proposed water supply network for East Laren (showing water demand in x10³ lpd).
5. Conclusions
In this paper, the areal extent of built-up areas in Karanggeneng and its surrounding, East Laren and Paciran in east Java, Indonesia is approximated to estimate the total water demand. A simple RGB color detection method by ImageAnalyst (2018) is employed to analyse satellite imagerys extracted from Google Earth. The applicable range of RGB spectrum is ascertained based on image inspection and the candidate areas identified are filtered, smoothened and manually cleaned-up to produce only the rooftops. Calibrated image pixel to area conversion is applied to obtain the final total built-up area. Water demand is then estimated based on 600 lpd per equivalent standard residential unit. The preliminary estimate is corroborated against year 2007 census data and anticipated population growth rate. The result shows that the approach taken is a cost-effective and efficient way for built-up approximation in large area using accessible online satellite imagerys.

6. References

[1] Rottensteiner F, Trinder J, Clode S, and Kubik K 2003 Building detection using LIDAR data and multi-spectral images. In: Sun C, Talbot H, Ourselin S and Adriaansen T (Eds.) Proc. VIIth Digital Image Comput.: Tech. Appl., 10-12 Dec. 2003. Sydney, Australia.
[2] Awrangjeb M, Ravanbakhsh M and Fraser C S 2010 Building detection from multispectral imagery and LIDAR data employing a threshold-free evaluation system. ISPRS Comm. III Symp. Photogrammetric Comput. Vision & Image Anal. (PCV 2010). Vol. XXXVIII, Part 3A, pp. 49-55. 1-3 September 2010. Paris, France.
[3] Jin X and Davis C H 2005 Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. EURASIP J. Appl. Signal Processing, 14: 2196–2206.
[4] Song Y and Shan J 2008 Building extraction from high resolution color imagery based on edge flow driven active contour and JSEG. Int. Arch. Photogrammetry, Remote Sensing and Spatial Information Sci. Vol. XXXVII. Part B3a. Beijing.
[5] Koc San D and Turker M 2010 Building extraction from high resolution satellite images using Hough transform. Int. Arch. Photogrammetry, Remote Sensing and Spatial Information Sci., Vol. XXXVIII, Part 8. Kyoto, Japan.
[6] Güdülü H V 2008 Building detection from satellite images using shadow and color information. Master Thesis: Middle East Technical Univ.
[7] Saadi S and Bensaib M 2014 Detection of buildings height using satellite monoscopic image. 2nd Europ. Conf. Earthquake Eng. Seismology, Istanbul 25-29 Aug, Istanbul.
[8] Tachikawa Y, James R, Abdullah K and Mohd. Desa M N (eds.) 2004 Catalogue of Rivers for Southeast Asia and the Pacific, Vol. 5. The UNESCO-IHP Regional Steering Committee for Southeast Asia and the Pacific, 285 pp.
[9] Pemerintah Kabupaten Lamongan 2012 Rencana Tata Ruang Wilayah (RTRW) Kabupaten Lamongan Dalam Angka 2007. Lamongan.
[10] Image Analyst 2018. SimpleColorDetection(), MATLAB Central File Exchange. (https://www.mathworks.com/matlabcentral/fileexchange/26420-simplecolordetection), Retrieved November 2018.

Acknowledgments
The author would like to acknowledge Matrade Services Export Fund Committee for the funding on the work. The first author acknowledges the funding by Faculty of Civil Engineering (FKA), Universiti Teknologi MARA for the paper presentation.