Pattern recognition as tools for understanding night sky brightness variabilities

M Rezky\textsuperscript{1,*}, R Priyatikanto\textsuperscript{2}, A G Admiranto\textsuperscript{2} and E Soegiartini\textsuperscript{1}

\textsuperscript{1}Department of Astronomy, Institut Teknologi Bandung, Bandung, Jawa Barat 40132, Indonesia
\textsuperscript{2}Space Science Centre, National Institute for Aeronautics and Space (LAPAN), Bandung, Jawa Barat 40173, Indonesia

Email: mrezky088@gmail.com

Abstract. Night sky brightness (NSB) research related to the artificial light pollution issues has been increasing all over the world, using various measurement techniques and tools. The research produces tens of thousands of data for each month such that proper handling, processing and analysing the data become challenging. In this article, we demonstrate an alternative method for processing the NSB data by utilising pattern recognition techniques: Canny edge detection and Hough transform. These techniques were applied to identify data and extract important parameter from the NSB density plot semi-automatically. Datasets collected from Bandung, Garut, Subang and Sumedang used as the test cases. Three time segments (dusk, night and dawn) became the main focus of the analyses and our method successfully extracted following parameters: the rate of sky brightness change at dusk and dawn, the average NSB at night and the intersections which indicate transition time. This method, along with its many possible improvements, enables us to process data more effectively and encourage more observation campaigns to be conducted in the future.

1. Introduction
The artificial light pollution has become a growing issue on the 21\textsuperscript{st} century, especially within urbanised communities who concerned about its negative impacts on the environment [1] and human health [2] ground-based astronomical observatories that affected by the brighter night sky. Night sky brightness (NSB) measurements are set up as a way to quantify light pollution and mitigate its impact. Since Cinzano et al. (2001) presented their first artificial NSB World Atlas [3] which became a prominent reference for many light pollution research, many in-situ measurements have been conducted all around the globe (e.g. [4], [5]). There is also continuing improvements in measurement methods and tools to gather more accurate and reliable data [6].

NSB measurements can generate thousands of datasets for every night. Depends on the objectives, continuous measurements may be conducted as a short-term or long-term activity. For achieving long-term data (e.g. [7]), researchers are facing challenges such as increasing datasets as well as varying meteorological conditions and ageing of the measurement tools. Posch et al. (2018) have developed several ways to visualise large datasets on many types of diagram [8]. In some cases, we need to investigate the NSB growth rate over time and detecting outliers that may occur during the
measurement period. Added with the necessity to measure the sky brightness transition at dawn or dusk, a researcher may need additional tools for differentiating important data from outliers and also clustering datasets in related time segments (dawn, night and dusk). With many datasets involved, pattern recognition in data computation became a powerful instrument to implement this task.

In this article, we demonstrate an alternative method for NSB data processing by using pattern recognition techniques to extract some features needed to produce a result that corresponds with the growing trend of light pollution phenomenon. This method will allow us to treat the datasets as values that represent particular concentrations rather than as individual points, so we could focus on meaningful data points and extract information about NSB change rate trends.

2. Data source and measurement methods

The NSB datasets analysed in this paper have been carried out at LAPAN's four observation stations in West Java (see table 1). Measurements were conducted using Unihedron Sky Quality Meter (SQM) – LU type that was placed inside weather-proof housing and directed to the zenith. This type of photodiode-based instruments has an effective field of view of ~20° and produce sky brightness data in mag/arcsec². All SQM was set to record zenithal sky brightness magnitude from 5 PM to 7 AM local time with a sampling rate of one minute, regardless of weather conditions and moon phases. Datasets used in this study were obtained from April to July 2018. These raw datasets archived at LAPAN Space Science Database in Bandung.

### Table 1. Geographical coordinates and conditions of 4 NSB observation stations.

| Stations  | Longitude        | Latitude          | Types    |
|-----------|------------------|-------------------|----------|
| Bandung   | 107° 40’ 40.21” E | 6° 55’ 32.03” S  | Urban    |
| Subang    | 107° 46’ 07.43” E | 6° 33’ 44.99” S  | Suburban |
| Sumedang  | 107° 50’ 13.97” E | 6° 54’ 47.08” S  | Suburban |
| Garut     | 107° 41’ 31.97” E | 7° 39’ 00.22” S  | Rural    |

3. Data processing

The purpose of utilising pattern recognition techniques for NSB data processing is to identify features, especially linear trends, from the scotogram (see figure 1(a)). The scotogram is a density plot of sky brightness (or inversely sky darkness) magnitude values as a function of time. The scotograms in this article have a horizontal resolution of 5 minutes and a vertical resolution of 0.1 mag/arcsec². It becomes the primary input of the pattern recognition algorithm which consists of two main parts: (1) Canny edge detection and (2) Hough transform. We will discuss the main idea of each algorithm and its role in the scotogram analysis in the next subsections. We are using OpenCV modules [9] on Python 3.5.2 version to execute both of the methods.

3.1. Canny edge detection

Canny edge detection [10] is a computational procedure developed by John Canny to determine arbitrary edge profiles on the desired image. It was designed to detect a single boundary between contrasting signal-to-noise region in the image. The scotogram image is applied with a $5 \times 5$ Gaussian blur kernel to smoothen the image and reducing noise and $3 \times 3$ Sobel operation to determine segments or boundary lines which are represented by pixel intensity gradient in the horizontal and vertical direction of the image ($G_x$ and $G_y$). Both values determine the overall intensity gradient $G$ and direction $\theta$,

$$G = \sqrt{G_x^2 + G_y^2}$$  \hspace{1cm} (1)
\[ \theta = \tan^{-1} \left( \frac{\partial y}{\partial x} \right). \]  

(2)

Those boundary lines then undergo an edge detection process, which involves two user-defined thresholds (upper and lower), called hysteresis thresholding. Those thresholds play a role in filtering which local maximum pixels can construct an edge line. Pixels with higher intensity than upper limit are automatically accepted as a real local maximum. Pixels with intensity level below the lower limit are automatically denied. Pixels with intensity level between those limits could be accepted as a real local maximum only if such pixel can connect to a nearby real local maximum within few steps, or else it is denied. Those real local maximum pixels will construct a real edge line for the image as shown in figure 1(b). In this article, we used 14-15 variations of threshold range to examine how threshold ranges affect the final results.

3.2. Hough transform

To detect straight or curved lines in the image, Hough transform [11] is employed. This procedure involves pixel coordinate transformation between Cartesian space \((x, y)\) and parameter space \((\theta, \rho)\) which described as equation

\[ x \cos \theta + y \sin \theta = \rho. \]  

(3)

The transformed parameters space with \(\theta\) and \(\rho\) as its axis is called the Hough space. Each active pixel on the original image is represented as a sinusoidal curve in Hough space. Then the algorithm will produce a list of \((\theta, \rho)\) coordinates that correspond to nodes of sinusoidal curves, which defined as local maxima on Hough space. Each node contains information about the total of curves accumulated on the node, determines its weight (counts) against other nodes.

Nodes with the highest weights in the Hough space are selected and converted back into Cartesian space as a representative trend line of the datasets. Additionally, users can also assign how many nodes or how low the counts threshold of the nodes that can be issued as the output of this process. In the context of this article, this means that we could treat all NSB datasets in a Hough space, shown in figure 1(c), and then extract a representative trend line for each of the time segments without split them manually, as seen in figure 1(d).

4. Results and Discussion

Our data processing algorithm had successfully extracted linear feature for each time segment, e.g. dusk, night and dawn. The segments were articulated as three clusters of nodes in Hough space at which the most significant nodes were extracted. For every line, the gradient and y-intercept were stored as the output for further analysis.
Linear features obtained by Hough transform can be translated into several parameters. First, line gradients at dusk (or at dawn) are the gradual change of sky brightness over time as the Sun goes away from (or to) horizon. The average value of sky brightness change at four observing sites is 0.23 mag/arcsec²/min which is in agreement with the result reported by Tyson & Gal [12], but lower than the value obtained at Paranal, Chile, which is 0.26 mag/arcsec²/min for V band [13]. There is no systematic difference between the changes at dusk and dawn, but brightness changes observed at SBG and GRT stations which are close to sea level are steeper compared to the one from BDG station (~750 masl). All values are summarised in table 2. Since the changes of surface brightness at these time segments are related to the scattering of sunlight in the atmosphere, it is plausible to correlate the difference of sky brightness changes to the difference in atmospheric pressures or to the atmospheric column density.

| Stations | Dusk segment | Dawn segment | Night segment |
|----------|--------------|--------------|---------------|
|          | Gradient     | Gradient     | Average NSB   | Gradient     |
|          | (mag/arcsec²/min) | (mag/arcsec²/min) | (mag/arcsec²) | (mag/arcsec²/min) |
| BDG      | 0.22 (0.00)  | -0.21 (0.00) | 17.25 (0.08)  | 0.77         |
| SBG      | 0.23 (0.04)  | -0.23 (0.00) | 18.55 (0.08)  | 0.21         |
| SMD      | 0.21 (0.04)  | -0.25 (0.00) | 20.39 (0.34)  | -0.69        |
| GRT      | 0.27 (0.06)  | -0.22 (0.00) | 20.86 (0.26)  | -0.01        |

The second parameter produced by the algorithm is related to the NSB itself. In this context, the NSB value is the most statistically significant value, associated with a high density at scotogram. NSB values during cloudy and overcast condition and NSB values with moonlight contamination were excluded by the algorithm. The extracted line indicates the most representative model of NSB during the measurement period. However, there were indeed variations in the obtained NSB values and trends due to the tuning of input parameters required by the algorithm, which we discussed later in this section.

In general, urban area (BDG) has the brightest night sky that reaches 17.25 mag/arcsec² while GRT becomes the station with the darkest sky (20.86 mag/arcsec²). It is important to note that the obtained values were not corrected by -0.25 mag/arcsec² offset due to the housing window. Slight changes of NSB were observed, especially at the BDG station where the sky at early night was approximately 2 times brighter than the sky before dawn (0.77 mag/arcsec² difference). Positive gradient means that the sky is getting darker over time, either by outdoor light reduction that is associated with human activities at night or by alteration of atmospheric condition. The average NSB for any station is defined as the midpoint of night segment trend line between dusk-to-night (d-n) and night-to-dawn (n-d) intersections.

The last parameters to be emphasised are the intersections between the segments. The intersections have notable importance in astronomical observation scheduling and also in the determination of prayers time for Muslims. In this article, the intersections are denoted in the local time since the analysis was done to measurements data obtained in a relatively short duration. During this period, there was a slight change in the setting and rising time of the Sun (~15 minutes). As an insight, the obtained time of d-n and n-d transitions are presented in table 3. For more proper analysis, utilisation of scotogram with Sun altitude as the horizontal axis is highly recommended, though the algorithm explained in this article is still applicable.
Parameters extracted by the algorithm from the scotogram do not always represent the current conditions of light pollution in a particular area, but the models could be updated as time goes by. These models could be our basis to understand the improvement/worsening trend of the pollution periodically. It is worth noted that this kind of evaluation must be treated carefully to produce reliable conclusions as there are many contributing factors affects accuracy, such as the climate of the measurement station or the degrading measurement instrument sensitivity. For example, we could compare the datasets that produced at a particular season that year with the datasets from the similar season a year before. Combining datasets from different seasons will not generate a proper model because of the difference in night duration (especially in temperate regions) or cloud coverage.

It is also important to note that the outputs generated by the algorithm are subject to change due to the hysteresis threshold variations used. As mentioned in subsection 3.1, the variation of input parameters on the Hough transform algorithm (the thresholds) may produce various results. In this case, there are 14-15 different results obtained on each station dataset. Varied input parameters could produce different level of spreads or errors in the output parameters. For NSB, the range was less than 0.3 mag/arcsec$^2$ or 2% relative error. For the brightness gradient at dusk, the maximum spread was 0.06 mag/arcsec$^2$ or approximately 20% relative error. This considerably large error could be associated with the nature of data submitted into the algorithm. The utilisation of scotogram as a function of Sun altitude is expected to suppress the error at dusk and dawn segments.

By comparing each varying results, we believe that there is no clear relation between the hysteresis threshold variations and the resulting data from it. The reason is that the lower and upper threshold affected differently on the pixels evaluated in images and the fact that thresholds determination is still subjective, which means it depends on the user’s judgement to assign specific values on the thresholds that satisfy their needs. This subjectivity could hamper the results produced by the Hough transform, because its input already suffers from Canny edge algorithm's poor anti-noise ability [14] and makes possible for the noises to be slipped off the results, increasing its deviation value. It was also found that the lower threshold determination influences the results more than the upper threshold because it sets the tolerance for a pixel in the detection process to be considered as building pixels of the real edge.

However, this data processing algorithm has a potential to be developed further into a semi-supervised machine learning algorithm. By modifying the current methods and made the program learns into and evaluate the results, we could gather more information about threshold variabilities and its effect on the trend lines, which will be given to the algorithm to produce the most acceptable set of rules on threshold determination. Medina-Carnicer et al. even suggest that it is possible to modify it into unsupervised learning [15]. Another problem that could be solved is the boundaries between each time segment. For this article, we set the boundary values based on visual information on the Hough space, which will generate some trouble when we are dealing with more data points. By letting the algorithm clustering the nodes by itself, it will ease the voting procedures when we need to produce trend lines for each time segment.
5. Conclusion
By utilising pattern recognition methods like Canny edge detection and Hough transform to the sky brightness measurement datasets, information about NSB variabilities (trend lines and its characteristic values) for dusk, night and dawn time segment can be extracted. Bandung has the brightest sky (NSB = 17.26 mag/arcsec²) among four stations evaluated, while Garut has the darkest one (NSB = 20.86 mag/arcsec²). Besides, typical decrement/increment of sky brightness at dusk/dawn was detected with an average change of 0.23 mag/arcsec²/min. This information could be used to examine the changing environment effects on light pollution problems in general and long-term scale.

Problems that occurred during implementation of this methods, like the hysteresis threshold on Canny edge detection and boundary value determination on Hough space were identified. Further development and modification on the algorithm into semi-supervised algorithm are expected to resolve these problems.

Acknowledgement
We acknowledge the observers at LAPAN’s observation stations and Nana Suryana at Subang for acquiring the NSB datasets in their respective station. We also acknowledge the help of Siti Maryam and Elyyani for maintaining the archive of NSB data at LAPAN Space Science Center. This article is a product of short-term internship program (Kuliah Kerja Astronomi) supervised by Dr. Hesti Retno Tri Wulandari. We also acknowledge financial support by LAPAN and FMIPA ITB for this publication.

References
[1] Longcore T and Rich C 2004 Front. Ecol. Environ. 2 191–8
[2] Reiter R J, Gultekin F, Manchester L C and Tan D-X 2006 J. Pineal Res. 40 357–8
[3] Cinzano P, Falchi F and Elvidge C D 2001 Mon. Notices Royal Astron. Soc. 328 689–707
[4] Puschnig J, Schwope A, Posch T and Schwarz R J. Quant. Spectrosc. Radiat. Transf. 139 76–81
[5] Bará S 2016 R. Soc. Open Sci. 3 160541
[6] Hänel A, Posch T, Ribas S J, Aubé M, Duriscoe D, Jechow A, Kollath Z, Lolkema D E, Moore C, Schmidt N, Spoelstra H, Wuchterl G and Kyba C C 2018 J. Quant. Spectrosc. Radiat. Transf. 205 278–90
[7] Ochi N and Wuchterl G 2015 Journal of Toyo University 58 1–12
[8] Posch T, Binder F and Puschnig J 2018 J. Quant. Spectrosc. Radiat. Transf. 211 144–65
[9] Bradsky G 2000 Dr. Dobb’s Journal of Software Tools
[10] Canny J 1986 IEEE Trans. on Pattern Anal. Mach. Intell. PAMI-8 679–98
[11] Duda R O and Hart P E 1972 Commun. ACM 15 11–5
[12] Tyson N D and Gal R R 1993 Astron. J. 105 1206–12
[13] Patat F, Ugolnikov O S and Postylyakov O V 2006 Astron. Astrophys. 455 385–93
[14] Feng Y, Zhang J and Wang S AIP Conf. Proc. 1890 040011
[15] Medina-Carnicer R, Munoz-Salinas R, Yeguas-Bolivar E and Diaz-Mas L Pattern Recognit. 44 1201–11