Evaluation of CCTV Data For Estimating Rainfall Condition

Sinta Berliana Sipayung, Lilik Slamet, Edy Maryadi, Indah Susanti dan Amalia Nurlatifah

s_berlianasipayung@yahoo.com

Abstract. Rainfall characteristics of Indonesia's tropical climate have high variability according to space and time, so to determine the rainfall pattern of a location, an in situ rainfall measuring instrument (AWS = automatic weather station) is needed with high density. The existence of AWS also requires relatively high maintenance costs and a standard placement location (according to the rules of WMO = World Meteorological Organization) which is relatively broad and is not obstructed by other objects that can make the result of rainfall data is not representative. With the concept of computer vision, research will be carried out to estimate the rainfall condition from the CCTV cameras. The CCTV camera data which have qualitative characteristic into rainfall data which have quantitative characteristics. This research is also motivated by the large number of CCTVs that are placed in a lot of locations by local governments along with the Smart City program in districts and cities throughout Indonesia. The preliminary research was conducted in Center for Atmospheric Science and Technology office in Bandung. Rainfall data from AWS was used to validate CCTV data which placed in same location. The process of converting CCTV data into rainfall data goes through 6 stages. The first is reading the image mapping data and AWS (in rainfall accumulation data form). Second, read the image data in grayscale. Third, extract the features. Fourth, split the reference and sample data. Fifth, conducts the K-NN Mapping Reference Image and rainfall accumulation data. Sixth is to praise K-NN Testing. The accuracy is calculate with comparing the estimated number of CCTV cameras that are correct with the total sample size. The evaluation result states that the highest accuracy is obtained with K = 1. When K=1, the accuracy percentage reaching 94.8%. Accuracy decreases with increasing value of K and drastically decreases with K> 2. In the 1-10 days reference data, the highest accuracy is obtained by the number of reference data for 10 days, which is around 97%, stable until the value of K = 8. While the lowest accuracy is obtained when the reference data is 1 day with an accuracy value of about 43%. Based on the results of this study, it can be concluded that rain data from CCTV can be used to estimate the rainfall data. The best result happened when K-value is equal to 1.

1. Introduction
Hydrometeorological disasters that include floods and landslides are always a frightening impression on the community, especially during the rainy season. The National Disaster Management Agency (BNPB) noted from its 2017 report that disasters in Indonesia had occurred 2,372 times which resulted in 377 deaths / missing and 3.49 million affected / displaced [1]. Flood is a hydrometeorological disaster that has the widest mode and area of impact when compared to other types of disasters. According to the Central Bureau of Statistics in 2018 there were 19875 urban villages affected by flooding in Indonesia [2]. Flood events are hydrometeorological disasters that often occur in cities both large and small, while landslides often occur in rural areas.

Among all natural disasters, the danger of flooding is the most serious in terms of the number of people affected and the deaths caused by it [3][4]. Therefore, the study of flood hazards has attracted...
significant attention in several scientific fields, such as those focusing on water resources and natural disasters. Nowadays, dynamic flood monitoring has been widely used in flood warning systems. The system mainly obtains data from a number of measuring stations, such as water level measurement stations, rainfall stations, and meteorological radar stations in water catchments [5][6][7]. The main data are fluctuations in water levels and the amount and distribution of rainfall. This monitoring system is important as part of the early warning system for heavy rain events that have the potential to become a flood hazard. This warning system monitors hydrological variables and their time derivatives to provide flood prevention information across sectors for flood management and relevant disaster reduction measures. In this way, the impact of flooding can be reduced as much as possible.

The high cost of an in-situ rain monitoring system that can produce high density data requires an alternative method of measuring rainfall that is cheaper and more efficient. Basically, remote sensing technology has now developed which can provide continuous rainfall data and a wide area so that it can show variations in rainfall conditions. However, remote sensing technology is not currently able to provide rainfall data in sufficient spatial resolution. At this time, there is the potential to develop a method of measuring rainfall that is cheaper and more efficient, namely by using a camera. One of the means of the National "Smart City" program for the capitals of City Governments and Regency Governments throughout Indonesia is to install closed circuit television (CCTV) cameras in several corners of the city with busy traffic.

The idea of using camera networks in urban areas to monitor rainfall and make it a 'tool' in hydrometeorological disaster management, is based on:

1. Understanding the working principle of remote sensing, where light waves / radiation that are reflected from an object and received by the sensor, have a certain value and indicate surface characteristics.

2. The development of information technology that allows image analysis. At this time, various algorithms used in outdoor vision systems have developed with the assumption that the image intensity is proportional to the brightness of the view. Garg and Nayar [8] have successfully developed a post-processing algorithm to detect and remove rain from images and videos for entertainment purposes. This is important, because for entertainment purposes, rain can greatly reduce the performance of various outdoor vision algorithms, including detecting stereo correspondence features, tracking, segmentation and object experience. This technique is used when there is no control over camera parameters during image / video shooting. Based on this, an idea emerged to modify the algorithm, so that the rain parameter was the object to be taken in the image analysis process.

There are 3 aspects that can support the technique or method of using cameras for monitoring rainfall, namely analysis of rain visibility, camera parameters to eliminate rain and rain gauge based cameras. Rain consists of a large number of droplets that fall at great speed. These droplets produce spatio-temporal high-frequency fluctuations in the video. In this case, an analytical reduction is needed that relates rainfall to camera parameters, rainfall characteristics and image brightness to identify rainfall [9][10]. In the analysis of several studies it has been shown that the visibility of rain increases with the size of the raindrops [11][12]. Rain visibility also decreases linearly with background brightness. In this case, the high speed and small size of the raindrops make the visibility of the rain very dependent on camera parameters, such as exposure time and depth of field. Raindrops fall at a high speed relative to the camera exposure time, resulting in very blurry motion streaks in the image. Also, due to the limited depth of field of the camera, the visibility of rain is significantly affected by blurring. In this case, it is necessary to model the motion-blur and out-of-focus intensity produced by the raindrops. This intensity is then used to obtain the effect due to rain volume [13][14][15].

The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems [16]. So that this method is hopefully good for being used in this research. The use of CCTV cameras as rain quantitative data and the need for rain predictions to mitigate hydrometeorological disasters, especially floods in cities, is analogous to satellite technology. The basic principle of satellite technology is camera
Therefore, in this research activity an effort will be made to convert qualitative data from CCTV cameras to quantitative rain data. In this paper, the study is focused on the city of Bandung.

2. Data and Method

2.1. Data

The data used in this research is CCTV and AWS data in the PSTA LAPAN Bandung Office for the period February, March, April, and September 2020. The processed data includes:

- **AWS Data**
  - Time: February 2020
  - Duration: 24 days
  - Period: 01-11, 20-29
  - Location: PSTA LAPAN Office (Dr. Djunjunan Street No. 133 Bandung)
  - Data unavailable: 34315

- **CCTV Data in Observation Lab**
  - Channel: 3
  - Time: February 2020
  - Duration: 24 days
  - Period: 01-06, 12-29
  - Location: PSTA LAPAN Office (Dr. Djunjunan Street No. 133 Bandung)

AWS and CCTV data collection is done manually, for a system built requires automatic data retrieval, especially CCTV and near real-time. Rainfall data obtained from AWS has a distribution as in table 1 below:

| Frequency | Count |
|-----------|-------|
| 0-5       | 192585|
| 5-10      | 90362 |
| 10-15     | 5125  |
| 15-20     | 7859  |
| 20-25     | 11208 |
| 25-30     | 2502  |
| 30-35     | 866   |
| 35-40     | 1231  |
| 40-45     | 155   |
| 45-50     | 560   |
| 50-55     | 796   |
| 55-60     | 345   |
| 60-65     | 937   |

AWS and CCTV data collection is done manually, for a system built requires automatic data retrieval, especially CCTV and near real-time. Rainfall data obtained from AWS has a distribution as in table 1 below:
2.2. Method

2.2.1. CCTV and AWS data collection

The first step to get a CCTV image model into rain information which is the main indicator in mitigating hydrometeorological disasters, namely collecting CCTV data and AWS data. CCTV data is moving image data that will be analyzed to be converted into qualitative rain information. Automatic Weather Station (AWS) data is data generated from measuring weather parameters including in situ rain data. Rain data from AWS data will be used as a reference in the process of assigning classes or CCTV data labels.

2.2.2. Data processing

To be able to estimate the rainfall using CCTV in urban networks, a model is needed. In this case, the model is built using machine learning based on CCTV image data compared to AWS data. The steps are as follows:

- Sorting acquired image data from CCTV video which is timed according to the available AWS data
- The second step is to classify CCTV image data based on AWS rain data. In this step, the AWS rainfall data will be grouped into light, medium and heavy rainfall values as measured by AWS. The results of this grouping are used as classes in the CCTV image data. Rain intensity attributes and characteristics are applied to each CCTV data which has a time suitability.
- The third step taken is the extraction of CCTV data features. In this step, CCTV data analysis will be carried out to determine what features might be used as a benchmark for determining rain. In addition to identifying features, at this stage an analysis will also be carried out on how much influence these features have in determining rainfall. Extraction of CCTV data model features to convert CCTV image data into qualitative rain information. In this case, the determination of rainfall patterns is carried out and produce weather information from CCTV data by:
  - Calculate similarity. In general, image pixels containing raindrops have 2 peaks in the temporal histogram.
  - Classification using several alternative classification techniques such as K-means cluster
  - Modelling
  - Big Data Analytic analysis produced by CCTV as presented in Figure 1

![Figure 1. Big Data Analytic Stages](image)

2.2.3. Analysis

The analysis includes model analysis and data analysis. Model analysis is done by validating the model output. There are several scenarios that will be carried out in between, namely train / test splits, cross validation and other scenarios. The modeling and validation stages will be carried out iteratively to
obtain a model with a high degree of accuracy. The standard accuracy rate is 80%. To achieve high accuracy several classification methods will be tried.

2.2.4. Disaster mitigation system development

The mitigation system development stage is the stage for presenting the model output developed from the previous stage in the form of software that can be used easily by general users. At this stage it will be carried out:

- Database design as a persistent data storage system for storing information on rainfall model output. Storage of model output rainfall information on a persistent basis is carried out to meet the needs of rainfall event analysis.
- Interface design to make it easier for users to access information.
- Designing an Application Programming Interface (API), so that the rain information generated can be accessed by existing programs / applications.
- Implementation of the designs that have been made.
- Test the implementation results.

The stages to be taken in developing a disaster mitigation system based on Big Data Analytic are presented in Figure 2.

The architecture of the hydrometeorological disaster mitigation system based on big data analysis is depicted in Figure 3. Different things have been done by (Dong, R., et al. 2017), namely measuring water droplets (raindrop size distribution) and light reflection using video by taking several pictures, while in this activity using one image technique with a lot of data (big data) including image processing (image processing), database, and machine learning then compared with in-situ data.
3. Results and Discussion

One of the important preprocessing to do is video fragmentation into image pieces. Basically, video is a display of a collection of images with fast moving. Fragmentation is more of the taking of one image for a specific time span according to AWS data acquisition. In this case, the fragmented image has a span of 10 seconds. The tool used is ffmpeg with the scripting language bash. Each fragmented image is given a name:

frame_channel_YYY-MM-dd__HH-mm-ss HH-mm-ss xxxx.png

*xxxx incremental number

Example: frame_came3_2020-02-01__00-00-01_00-59-59_0002.png

In addition, each image is stored in folders per day. The video fragmentation script can be seen in the following image (Figure 4):

Figure 4. Script Spliting (bash script)

The next pre-processing step is mapping the CCTV frame data with AWS data. The function of this step is to map or correspond the cctv frame / cct image with the accumulated rainfall from AWS adjusted to the date. Tools and libraries used are python and pandas, numpy and datetime in the python programming language. The result of this step is tabular data consisting of CCTV date, AWS date, CCTV image folder name, CCTV image file name, rainfall, accumulated rainfall in 10 seconds. The total data obtained reached 161811.
The quantification of CCTV data frames into rainfall is carried out in the order shown in Figure 6. Quantification of CCTV data frames into rainfall was carried out with 2 experiments.

- **1st Experiment**:
  - Reference Data : 60000
  - K amount : 1 -10

- **2nd Experiment**:
  - Reference Data : 1 – 10 Hari
  - Amount of data in one day : 8640
  - K amount : 1 – 10

Both experiments produce rainfall estimation data, neighbours and distances for the data classification. The experimental results are then evaluated by:

- Comparison of the estimated rainfall results with the aws rainfall data with the corresponding date
- Calculating the accuracy of each experiment with the formula:

\[
\text{accuracy} = \frac{\text{The estimated amount is corrected}}{\text{Sample amount}}
\]

### 3.1 1st Experiment

In the first experiment, the amount of data that became the reference was 60,000 data. In this experiment, the highest accuracy was 94.8% and was achieved when the value of K = 1. The value of K continues to decrease significantly when K exceeds 2. Figure 7 shows the accuracy of the data between the accumulated rainfall data from AWS and the data from the CCTV model using a K value between 1 and 10. The maximum accuracy obtained with a value of K = 1.
Figure 7. Accuracy value of the AWS rainfall data with the model data using a similarity value of $1-10$

3.2. 2nd Experiment

Reference data covers $1-10$ days. The highest accuracy is obtained by the number of reference data for 10 days, which is about 97%, stable until the value of $K = 8$. While the lowest accuracy is obtained when the reference data is 1 day with an accuracy value of around 43%, the $K$ value has no effect. In the whole experiment 2 the K-NN accuracy was between 43% - 97%. In this second experiment, the amount of data used is around 8,640 data.

Figure 8. Comparison of model accuracy based on the second experiment using the value of $K = 1-10$

After conducting experiments and calculating accuracy, experiments were also carried out on rainfall conditions in January 2020 which is located at the PSTA LAPAN Office. The experimental results were in the form of a comparison of the AWS rainfall values (ground truth) and the rainfall values modeled (inference). There are two experiments carried out, the first is an experiment involving all January data as model data and the data is tested. The second experiment is an experiment which divides the data into
two, namely model data and test data. From the total data as many as 267,579 (sorted by time of January),
divided by the composition as follows: the first 60,000 data as model data, 207,579 as test data. Experiments were also carried out by changing the value of K.

Figure 9 presents a comparison between the rainfall from AWS and the rainfall from the CCTV model during the first experiment when the value of K = 1. The best model value and the closest to the AWS value is obtained when the K value = 1.

![Figure 9. Comparison of the AWS rainfall value and the output value of the K-NN Model with K = 1](image)

4. Conclusion
Data processing and analysis is carried out to determine the most accurate model value in estimating the rainfall value from AWS. The accuracy calculation data between the AWS rainfall value and the rainfall from the CCTV model shows that the highest accuracy obtained is 97% and the lowest is 43%. The amount of reference data and the K value affects the visible accuracy of experiments 1 and 2, the accuracy obtained varies according to the amount of reference data and its K value and the number of reference data with the highest accuracy is 10 days, equivalent to 86400. Accuracy above 75% is obtained by the number reference data $> 2$ days (2x8640) and K values between 1 and 2, while K values with the highest accuracy are K values between 1 and 2.

Reference
[1] https://bnpb.go.id/uploads/24/buku-data-bencana/buku-data-bencana-2017-compressed.pdf accessed 22 March 2021
[2] https://www.bps.go.id/indicator/168/954/1/banyaknya-desa-kelurahan-menurut-jenis-bencana-alam-dalam-tiga-tahun-terakhir.html accessed 22 March 2021
[3] Borga, M., Anagnostou, E.N., Bloschl, G., Creutin, J.-D., 2011. Flash flooding forecasting, warning, and risk management: THE HYDRATE Project. Environ. Sci. Policy 14(7), 834-844.
[4] Borga, M., Stoffel, M., Marchi, L., Marra, F., Jakob, M., 2014. Hydrogeomorphic response to extreme rainfall in headwater systems: Flash floods and debris flows. Journal of Hydrology 518 (2014) 194–205.
[5] Marin-Perez, R., Garcia-Pintado, J., Gomez, A.S., 2012. A Real-Time Measurement System for Long-Life Flood Monitoring and Warning Applications. Sensors 2012, 12, 4213-4236; doi:10.3390/s120404213.
[6] Li M.W., Li G.L., Jiang Y.Z. 2013. The application of the electrode type water level gauge in reclaimed water treatment control system. Appl. Mech. Mater. 333–335:2297–2300. doi: 10.4028/www.scientific.net/AMM.333-335.2297.
[7] Ji Y.N., Zhang M.J., Wang Y.C., Wang P., Wang A.B., Wu Y., Xu H., Zhang Y.N. 2014. Microwave-photonic sensor for remote water-level monitoring based on chaotic laser. Int. J. Bifurc. Chaos. 24:1450032. doi: 10.1142/S0218127414500321.

[8] Garg, K. and Nayar, S.K. 2006. Photorealistic rendering of rain streaks. ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH), 25:996–1002.

[9] Zheng and Shijie Jiang. 2018. Big Data Challenges for Predictive Modeling of Complex Systems.

[10] S. Shirmohammadi and A. Ferrero, 2017. Camera as the instrument: the rising trend of vision based measurement, IEEE Instru. Meas. Mag., vol.17, no.3, pp.41-47, 201.

[11] Garg. K and S.K. Nayar. 2005, Rain and visibility. Technical Report.

[12] Dong, R., Liao, J., Li, B., Zhou, H., & Crookes, D, 2017. Measurements of Rainfall Rates From Videos. In 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2017): Proceedings IEEE.

[13] Gunn R and G.D. Kinzer. 1949. Terminal velocity for water droplet in stagnant air. J. Metero., 6:243–248.

[14] E. Habib., W.F. Krajewski, and A. Kruger. 2001. Sampling errors of tipping-bucket rain gauge measurements. J. of Hydro. Eng., 6:159.

[15] Horn, B.K.P. 1986. Robot Vision. The MIT Press.

[16] https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761 accessed July, 3 2021

Acknowledgments
The authors thank the Ministry of Research and Technology/National Agency for Research and Technology who have facilitated and supported morally and materially so that this research can be realized.