

An Image Similarity Evaluation in Rainfall Forecasting Model

Prattana Deeprasertkul
Technology and Digital Development Division
Hydro – Informatics Institute (Public Organization)
Bangkok, Thailand
Email: prattana[AT] hii.or.th

Abstract—The Global Satellite Mapping of Precipitation or GSMaP data which is used to display the rainfall data was used to analyze and create the rainfall forecasting model. This work is the evaluation of this rainfall forecasting model which is the short-term forecast. The GSMaP forecasting data were matched with the GSMaP history data and calculate their similarity values by applying the original image matching method. The modification of Rainfall Forecasting Model and its evaluation that applied the original image instead of the image hash improve the accuracy of rainfall forecasted results.

Keywords: Image Similarity; Image Matching; Rainfall Forecasting;

I. INTRODUCTION

The improvement in the rainfall forecasting model of Thailand country for flood warning and monitoring have been developed. The Sea Surface Temperature or SST datasets and the Global Satellite Mapping of Precipitation project or GSMaP datasets which are the satellite images were analyzed to predict the rainfall by deep learning method. The User Interface of the model is shown in Fig. 1. This is the improving Big Data Analytics works which has been a project of Hydro – Informatics Institute (Public Organization).

The improvement of image similarity evaluation of the rainfall forecasting model should be measured for representing the system efficiency as well. In the previous work [7], we assessed the model with a perceptual hash technique which generated the hash strings by the hash algorithm. A perceptual

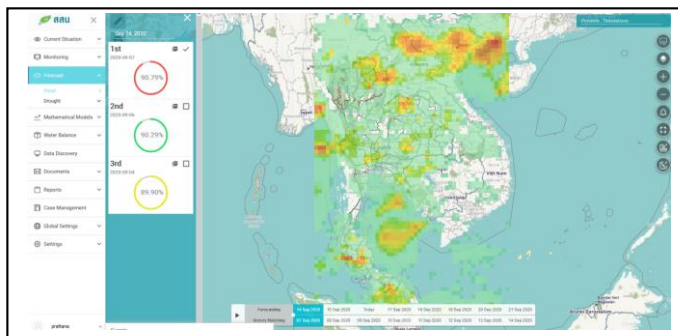


Figure 1. The User Interface of Rainfall Forecasting Model.

hash function calculates similar perceptual hash strings for similar images by comparing and measuring two perceptual hashes. The hash value is much shorter than the original image. However, this work uses the original image values for comparing the similar and dissimilar values. Hamming distance is used to count the number of different bits in string values. Then, the GSMaP forecasting data was matched with the history GSMaP data for finding similarity values by applying the original images.

The remaining of this paper is organized as follows: Section 2 presents the general background information about this accuracy measurement. Section 3 presents the accuracy measurement of rainfall forecasting model. Section 4 shows the result of accuracy measurement. Finally, section 5 is the discussion and conclusion.

II. DATA AND METHODS

A. Data

The data used in this work is the satellite precipitation or GSMaP. The GSMaP [4,5,8,9] or Global Satellite Mapping of Precipitation project was the production of a high-precision, high-resolution global precipitation map using satellite data sponsored by Core Research for Evolutional Science and Technology (CREST) of the Japan Science and Technology Agency (JST) during 2002-2007. Since 2007, GSMaP project activities are promoted by the JAXA Precipitation Measuring Mission (PMM) Science Team. This work uses the GSMaP data sets during 2014-2020.

B. Data Type and Format

The GSMaP as mentioned in previous topic is the satellite data and has a format in the digital image, therefore data is converted from JPEG to ESRI ASCII raster format to overlay them on Map for good performance. The basic structure of the ESRI ASCII raster has the header information at the beginning of the file followed by the cell value data as illustrated in Table I [2].

Example of ESRI ASCII raster file is shown in Fig. 2.

TABLE I. THE HEADER OF ESRI ASCII RASTER FILE.

Parameter	Description	Requirements
NCOLS	Number of cell columns.	Integer greater than 0
NROWS	Number of cell rows.	Integer greater than 0
XLLCENTER	X coordinate of the origin (by lower left corner of the cell).	Match with Y coordinate type.
YLLCENTER	Y coordinate of the origin (by lower left corner of the cell).	Match with X coordinate type.
CELLSIZE	Cell size.	Greater than 0.
NODATA_VALUE	The input values to be NoData in the output raster.	Optional. Default is -9999.

```
ncols 480
nrows 450
xllcorner 378923
yllcorner 4072345
cellsize 30
nodata_value -32768
43 2 45 7 3 56 2 5 23 65 34 6 32 54 57 34 2 2 54 6
35 45 65 34 2 6 78 4 2 6 89 3 2 7 45 23 5 8 4 1 62 ...
```

Figure 2. The example of ESRI ASCII raster file.

C. Image Matching

Image matching is based on the maximum similar pixel values between two digital images. The image matching methods are classified into two categories: one is gray-based matching; the other is feature-based matching. The gray-based image matching algorithm is based on the grayscale feature of pixels for image matching. The feature-based matching algorithm is based on feature vectors of feature points for image matching [3]. Image similarity is the measurement of similarity between two images. Then some similarity criteria are used to determine the corresponding points in two images. The quality of similarity is defined by the distance. Currently, there are many distance measurement methods such as Euclidean distance, Humming distance, Cosine distance, and etc. the Hamming distance is one of several string metrics for measuring the edit distance between two sequences.

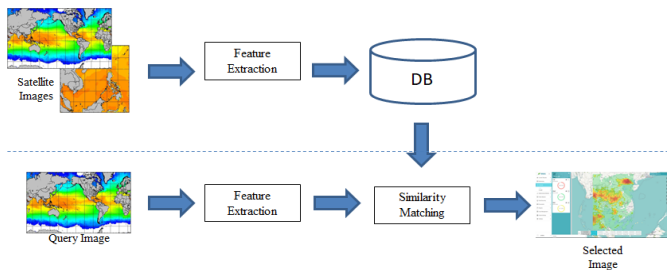


Figure 3. The image matching process flow.

The process flow of image matching is shown in Fig. 3. The query image is extracted the feature, and then search match with the stored images in the database for evaluated the

similarity values. The image which has the highest similarity value is selected.

D. Image Similarity Measure

In this section, the definition of similarity between images is presented. Then the similarity measure is used in this work as a reference for image similarity.

1) Similar Image

The closeness between two images is based on a pixel-to-pixel comparison. Similar image is the two images considered to have more intensity of the same values as compared to the different values.

2) The Similarity Measure

Hamming distance is a metric for comparing two binary data strings [6]. A function on words of fixed length over an alphabet describing the number of changes to the symbols of one word required to reduce it to another. Let A be an alphabet of symbols and C a subset of A^n , the set of words of length n over A . Let $x=(x_1, \dots, x_n)$ and $y=(y_1, \dots, y_n)$ be words in C . The Hamming distance $d(x,y)$ is defined as the number of difference between x and y ; that is, $\#\{i: x_i \neq y_i; i=1, \dots, n\}$. The Hamming distance satisfies [1]:

$$d(x,y) \geq 0;$$

$$d(x,y) = 0 \text{ if and only if } x=y;$$

$$d(x,y) = d(y,x).$$

Hamming Distance between two strings is the number of bits which are different at same position in both strings.

III. EVALUATION OF RAINFALL FORECASTING MODEL

The accuracy measurement of the rainfall forecasting analysis being the main topic of this work has been explained the process in this section. The process flow has been displayed in Fig. 4. In this work, the 1D array string of an image was applied to search the image similarity values. The 1D array is the values in the image pixels to compare and retrieve the most similar image in storage. Even if using the original image is time-consuming process, but it receives the exact similarity values.

The forecasting GSMaP datasets were matched with the history matching GSMaP datasets which were stored in Mongo DB, both. Both 1D arrays of pixel values were generated and then the similarity percentage between each two images was processed and calculated with image similarity measurement, respectively. An example between the forecasting GSMaP dates and the history matching GSMaP dates as shown in Table II and Table III.

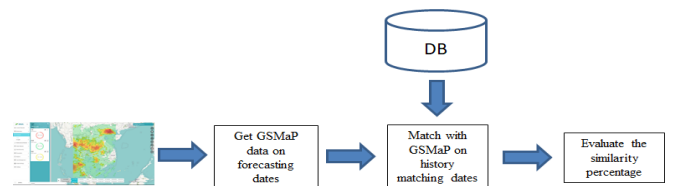


Figure 4. The process flow of the accuracy measurement of rainfall forecasting model.

TABLE II. THE EXAMPLE I OF FORECASTING DATE AND HISTORY MATCHING DATE OF GSMAP DATA.

Forecasting date	History matching date
31 May 2020	06 Oct. 2017
01 Jun. 2019	07 Oct. 2017
02 Jun. 2019	08 Oct. 2017
03 Jun. 2019	09 Oct. 2017
04 Jun. 2019	10 Oct. 2017
05 Jun. 2019	11 Oct. 2017
06 Jun. 2019	12 Oct. 2017
07 Jun. 2019	13 Oct. 2017

TABLE III. THE EXAMPLE II OF FORECASTING DATE AND HISTORY MATCHING DATE OF GSMAP DATA.

Forecasting date	History matching date
01 Apr. 2020	31 Mar. 2019
02 Apr. 2020	01 Apr. 2019
03 Apr. 2020	02 Apr. 2019
04 Apr. 2020	03 Apr. 2019
05 Apr. 2020	04 Apr. 2019
06 Apr. 2020	05 Apr. 2019
07 Apr. 2020	06 Apr. 2019
08 Apr. 2020	07 Apr. 2019

IV. RESULTS

The program to process about the results evaluation of the rainfall forecast model by using image similarity comparison was written by Jupyter or Python as shown in Fig. 5. The GSMaP on the forecasting dates were evaluated the similarity with the history matching dates. The example images of one GSMaP similarity (%) case has been shown in Fig. 6. The similarity values of eight GSMaP forecasting dates and their history matching dates were calculate the average as shown in Table IV. Table IV is the example 1 and the example 2 shows in Table V.

The average of GSMaP similarity since 01 April 2020 to 31 May 2020 had the values during 71.6075% to 75.36%.

```
In [19]: d = "20200401"
         s = "20200331"

In [20]: check_similar_gsm(d, s, imagehash.phash, hashsize=64, isPlot=True, last=7)

Out[20]: [('20200404', '20200403', 70.0),
          ('20200401', '20200331', 73.72),
          ('20200405', '20200404', 73.95),
          ('20200402', '20200401', 74.97),
          ('20200407', '20200406', 76.52),
          ('20200403', '20200402', 76.93),
          ('20200406', '20200405', 77.49),
          ('20200408', '20200407', 79.32)]
```

Figure 5. The example of Python program for GSMaP.

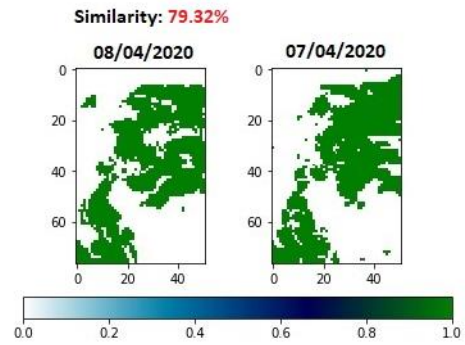


Figure 6. The example images of GSMaP similarity (%) comparison on date 08/04/2020 with 07/04/2020 and getting 79.37% similarity.

TABLE IV. THE EXAMPLE I OF THE AVERAGE OF GSMAP FORECASTING AND HISTORY MATCHING SIMILARITY.

Forecasting date	History matching date	Similarity (%)
2020/05/31	2017/10/06	73.92
2020/06/01	2017/10/07	77.46
2020/06/02	2017/10/08	70.54
2020/06/03	2017/10/09	62.92
2020/06/04	2017/10/10	70.72
2020/06/05	2017/10/11	77.08
2020/06/06	2017/10/12	76.52
2020/06/07	2017/10/13	74.23
Avg.		72.92

TABLE V. THE EXAMPLE II OF THE AVERAGE OF GSMAP FORECASTING AND HISTORY MATCHING SIMILARITY.

Forecasting date	History matching date	Similarity (%)
2020/04/01	2020/03/31	73.72
2020/04/02	2020/04/01	74.97
2020/04/03	2020/04/02	76.93
2020/04/04	2020/04/03	70.0
2020/04/05	2020/04/04	73.95
2020/04/06	2020/04/05	77.49
2020/04/07	2020/04/06	76.52
2020/04/08	2020/04/07	79.32
Avg.		75.36

V. CONCLUSIONS

The SST datasets and GSMaP datasets analysis for rainfall forecasting model was evaluated the accuracy by applying image matching to find out its similarity values. The previous work using image hashing for evaluating model, the accuracy was around 66.46%. For the current work by improving the rainfall forecasting model and the model evaluation applying the original image values, then the accuracy of the rainfall forecasting model was around 73.76%.

However, the forecasting model will be improved the accuracy by considering the other weather parameters over Thailand. The future work, the middle-term and long-term rainfall forecasting will be analyzed and developed in this application as well.

REFERENCES

- [1] Encyclopedia of Mathematics, “Hamming distance,” URL: http://encyclopediaofmath.org/index.php?title=Hamming_distance&oldid=39148.
- [2] ESRI Inc., “ESRI ASCII raster format,” http://resources.esri.com/help/9.3/arcgisengine/java/GPToolRef/spatial_analyst_tools/esri_ascii_raster_format.htm.
- [3] G. Jianhua, Y. Fan, T. Hai, W. Jing-xue, and L. Zhiheng, “Image matching using structural similarity and geometric constraint approaches on remote sensing images,” *Journal of Applied Remote Sensing*, 2016: 10.045007. 10.1117/1.JRS.10.045007.
- [4] K. Aonashi, J. Awaka, M. Hirose, T. Kozu, T. Kubota, G. Liu, S. Shige, S. Kida, S. Seto, N. Takahashi, and Y. N. Takayabu, “GSMaP passive, microwave precipitation retrieval algorithm: Algorithm description and validation,” *J. Meteor. Soc. Japan*, 2009, 87A, 119-136.
- [5] K. Okamoto, T. Iguchi, N. Takahashi, K. Iwanami and T. Ushi, “The Global Satellite Mapping of Precipitation (GSMaP) project,” 25th IGARSS Proceedings, 2005, pp. 3414-3416.
- [6] N. Raut, “What is Hamming Distance?”, 2018, TutorialPoint.
- [7] P. Deeprasertkul, “An Assessment of Rainfall Forecast using Image Similarity Processing,” *ICIGP Proceedings of the 2020 3rd International Conference on Image and Graphics Processing*, 2020, pp. 141–145. <https://doi.org/10.1145/3383812.3383821>.
- [8] T. Kubota, S. Shige, H. Hashizume, K. Aonashi, N. Takahashi, S. Seto, M. Hirose, Y. N. Takayabu, K. Nakagawa, K. Iwanami, T. Ushio, M. Kachi, and K. Okamoto, “Global Precipitation Map using Satelliteborne Microwave Radiometers by the GSMaP Project : Production and Validation,” *IEEE Trans. Geosci. Remote Sens.*, Vol. 45, No. 7, 2007, pp.2259-2275.
- [9] T. Ushio, T. Kubota, S. Shige, K. Okamoto, K. Aonashi, T. Inoue, N. Takahashi, T. Iguchi, M. Kachi, R. Oki, T. Morimoto, and Z. Kawasaki, “A Kalman filter approach to the Global Satellite Mapping of Precipitation (GSMaP) from combined passive microwave and infrared radiometric data,” *J. Meteor. Soc. Japan*, 2009, 87A, 137-151.