electronics

Article

Weather Forecast Based on Color Cloud Image Recognition under the Combination of Local Image Descriptor and Histogram Selection

Kiet Tran-Trung [1], Ha Duong Thi Hong [2] and Vinh Truong Hoang [2,*]

[1] Faculty of Fundamental Studies, Ho Chi Minh City Open University, 97 Vo Van Tan, District 3, Ho Chi Minh City 70000, Vietnam
[2] Faculty of Information Technology, Ho Chi Minh City Open University, 97 Vo Van Tan, District 3, Ho Chi Minh City 70000, Vietnam
* Correspondence: vinh.th@ou.edu.vn

Abstract: Numerous researchers have used machine vision in recent years to identify and categorize clouds according to their volume, shape, thickness, height, and coverage. Due to the significant variations in illumination, climate, and distortion that frequently characterize cloud images as a type of naturally striated structure, the Local Binary Patterns (LBP) descriptor and its variants have been proposed as feature extraction methods for characterizing natural texture images. Rotation invariance, low processing complexity, and resistance to monotonous brightness variations are characteristics of LBP. The disadvantage of LBP is that it produces binary data that are extremely noise-sensitive and it struggles on regions of the image that are "flat" because it depends on intensity differences. This paper considers the Local Ternary Patterns (LTP) feature to overcome the drawbacks of the LBP feature. We also propose the fusion of color characteristics, LBP features, and LTP features for the classification of cloud/sky images. Morover, this study proposes to apply the Intra-Class Similarity (ICS) technique, a histogram selection approach, with the goal of minimizing the number of histograms for characterizing images. The proposed approach achieves better performance of recognition with less features in use by fusing LBP and LTP features and using the ICS technique to choose potential histograms.

Keywords: cloud images; preprocessing techniques; classification; histogram selection; ICS; LTP; LBP; color image; weather forecast

check for updates

Citation: Tran-Trung, K.; Duong Thi Hong, H.; Truong Hoang, V. Weather Forecast Based on Color Cloud Image Recognition under the Combination of Local Image Descriptor and Histogram Selection. *Electronics* **2022**, *11*, 3460. https://doi.org/10.3390/electronics11213460

Academic Editors: Yanhui Guo, Deepika Koundal and Rashid Amin

Received: 19 September 2022
Accepted: 13 October 2022
Published: 26 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Any region's weather is closely tied to the presence of clouds. Clouds play a major role in all types of precipitation. Although not all clouds may result in precipitation, they are crucial for controlling the weather in some regions. Different types and heights of clouds exist in various geographical locations, such as the Earth's tropics or poles [1]. The sky always has clouds, and they are ever-changing. Clouds serve as indicators of atmospheric conditions and are crucial for weather forecasting and warnings, as well as for controlling the Earth's energy balance, temperature, and weather [2]. Before creating a weather prediction, meteorologists investigate the specifics of the cloud type since they are constantly changing.

We can forecast changes in weather by studying and categorizing clouds. Future measurements and forecasts may be significantly impacted by changes in cloud classification. Additionally, locating and analyzing clouds can assist meteorologists in modifying weather forecasts, better comprehending the local ecology, and foretelling changes in the world's climate [3]. Ground-based cloud observations are remote sensing image materials, so they will have information such as cloud parameters, spatial resolution, temporal

resolution, spectral resolution, etc. Classification of cloud information (cloud parameters, temporal, and spatial resolution, etc.) is very important. These parameters need to be defined for ground-based cloud observations such as cloud height, cover, and cloud type [4–6], in which the parameter of cloud type is the first and most readily available parameter that humans can obtain.

The cloud is a product of nature, and it reflects the climate and different weather conditions on Earth. Clouds appear very diverse in different atmospheric conditions. Extracting useful information from huge amounts of image data by detecting and analyzing different entities in images is a major challenge. Today, the identification of cloud types is mainly based on the observations of experts. The results are subjective and cannot meet the actual requirements of meteorological observations. The automatic classification of cloud types has become a demanding problem that needs to be solved in this case.

Researchers are increasingly using imagery of the entire sky above the ground for a variety of purposes, including solar radiation, weather forecasting, and aviation. These images are usually higher in resolution than those obtained from [7] the satellite. Additionally, the camera's vertical orientation makes it simple to photograph clouds at various low altitudes. Thus, they enrich satellite images with useful data. In most cases, cloud observations are performed manually by experts from meteorological institutes. They observe these clouds to better understand atmospheric phenomena and classify clouds into various categories according to the World Meteorological Organization standard [8]. Although the results obtained are quite accurate, such manual observations are expensive and time-consuming. Therefore, it is necessary to apply cloud classification algorithms automatically and systematically to save costs.

In recent years, several studies have approached computer vision techniques to identify and classify clouds based on their volume, shape, thickness, height, and coverage. For example, Kliangsuwan and Heednacram [9] used a new method for feature extraction in cloud classification. The authors propose three more types of features based on Fourier transform, namely Fast Fourier Transform Projection on the modified x-axis (k-FFTPX), half k-FFTPX, and $h \times k$-FFT, and use an ANN-based classification technique with a tree algorithm to extract features. Li et al. [10] proposed a new approach to cloud pattern recognition, based on analyzing the image as a set of patches (set of changes), rather than a set of pixels, and through the Support Vector Machine (SVM) classifier for classification. Zhen et al. [11] used spectral and texture feature extraction by tonal statistical analysis and Gray Level Cooccurrence Matrix (GLCM), and an SVM classifier with the Radial Basis Function (RBF) multiplier to classify different clouds from sky images. Taravat et al. [12] used a neural network in conjunction with SVM for automatic cloud classification for the entire color terrestrial image.

Algorithms based on structural features such as cloud scale, edge sharpness, Fourier transform, etc., cannot effectively exploit the useful information of cloud images due to cloud images, as a type of natural striated structure, often possessing variations with very large connotations due to large variations in illumination, climate, and distortion [13]. Cheng and Yu [14] proposed a cloud classification method that deals with mixed cloud types in an image based on the segmentation of images in different blocks. In each block, the texture statistical feature and the Local Binary Patterns (LBP) feature are extracted. These features are then classified using the Bayesian classifier. Liu et al. [15] proposed a new feature extraction technique by improving the LBP technique. This technique is called Salient Local Binary Pattern (SaLBP) for terrestrial cloud image classification. SaLBP utilizes the most frequently occurring patterns (prominent patterns) to obtain descriptive information. Liu and Zhang [13] presented a new feature extraction algorithm called Learning Group Patterns (LGP) to classify seven sky conditions; the proposed algorithm considers the resolution of the texture by using SaLBP and through LGP. Zhang et al. [16] focused both on designing appropriate feature representations and learning distance metrics from sample pairs. The authors also propose a feature extraction technique called Transfer Deep Local Binary Patterns (TDLBP) and learn WML. Wang et al. [17] proposed a powerful

feature extraction method based on the average rank of occurrence frequencies of invariant rotational samples defined in the LBP of the cloud image, called SLBP. In recent years, different image classifications based on deep learning have been proposed and demonstrated for their effectiveness [18,19]. In 2020, Wang et al. [20] proposed a convolutional neural network (CNN) integrated with a neural network with deep learning capabilities, called CloudA, as a ground-based cloud image recognition method. CloudA visualizes cloud features using TensorBoard visualization, and these features can help us to understand the terrestrial cloud classification process.

Therefore, feature extraction plays an important role, affecting the results of the classifier. There are many feature extraction methods have been proposed in the last decade. The LBP feature and its variants have been proposed as an effective feature extraction method for classifying natural texture images [21]. Although the extraction of local features by LBP gives many positive results, there are still noisy and duplicated features. Indeed, many extracted features will decrease the performance due to the curse of dimensionality. Feature selection involves finding a subset of valid features. To improve the accuracy of the classifier and reduce the computational burden, the feature selection step is essential.

This paper presents an approach for color image classifcation based on LTP features and feature selection methods. Since the color information is important to represent texture, we consider different color spaces for extracting local features. The rest of this paper is structured as follows. Section 2 presents the feature extraction by the LBP and LTP descriptor. Section 3 introduces the histogram selection approach via the Intra-Class Simalarity (ICS) score for selecting the most important features. Section 4 presents the cloud image classification process. Then, Section 5 shows the experimental results on the two benchmark datasets. Finally, Section 6 gives the conclusions and perspectives of this work.

2. Feature Extraction

Feature extraction is an important step for multimedia processing. The question of how to extract ideal features that can still reflect, as fully as possible, the contents of the image remains a challenging problem in computer vision. In other words, feature extraction is the process of obtaining the most important data from the raw data. Feature extraction is an important step in building any pattern classification model and aims to extract relevant features. In this process, relevant features are extracted from the objects to form feature vectors.

The most common image features include color, texture, and shape, etc., and most of the feature spaces are built on these features. However, the performance of the model depends heavily on the use of the image features [22]. LBP feature extraction was first proposed by Ojala et al. [23] to describe the texture of the image. It is the comparison of neighboring pixels with the central pixel to obtain a binary sample. This binary pattern is generated as follows: all neighboring pixels will take the value 1 if its pixel value is greater than the central pixel value, and otherwise take the value 0. Then, the pixels are multiplied with the respective weights and summed to obtain the LBP value for the center pixel.

The formula for calculating LBP is determined as follows:

$$\mathrm{LBP}_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \tag{1}$$

where

$$s(g_p - g_c) = \begin{cases} 1 & \text{if } s(g_p - g_c) \geq 0 \\ 0 & \text{if } s(g_p - g_c) < 0 \end{cases} \tag{2}$$

where, x_c and y_c are the coordinates of the center pixel, P is the number of neighboring pixels, R is the neighborhood radius, g_c is the grayscale value of the center pixel, g_p is the grayscale value of the neighboring pixel.

Figure 1 shows the process and encoding of the LBP operator for grayscale images with 3×3 pixels.

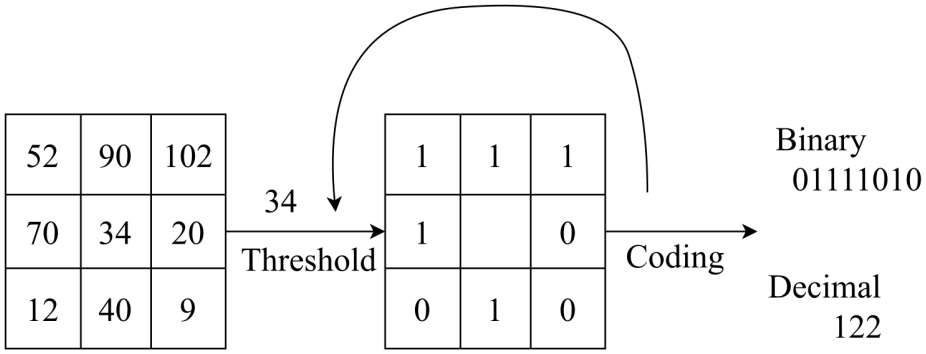


Figure 1. Describes the encoding of the LBP operator.

The LTP operator was developed from LBP and introduced by Tan and Triggs [24]. This proposal offers significantly higher efficiency than LBP and better noise handling than LBP in homogenous regions. In LTP, $s(g_p - g_c)$ is defined as follows:

$$s(g_p - g_c) = \begin{cases} 1 & \text{if } s(g_p - g_c) \geq t \\ 0 & \text{if } \mid s(g_p - g_c) \mid < t \\ -1 & \text{if } s(g_p - g_c) \leq -t \end{cases} \quad (3)$$

where, t is the user-defined threshold. Figure 2 shows the working and encoding of the LTP operator for grayscale images with 3×3 pixels, with the parameter value $t = 5$.

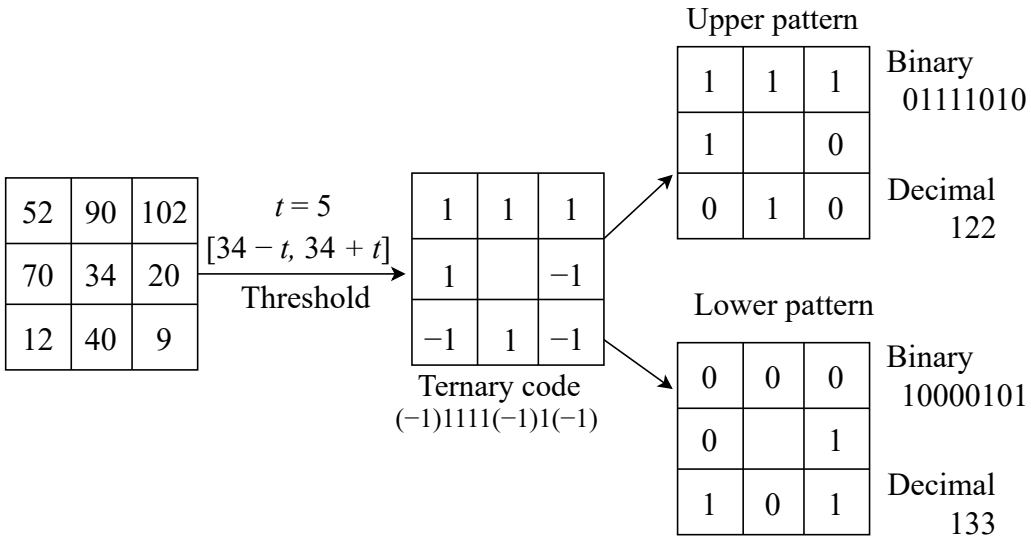


Figure 2. Describes the encoding of the LTP operator.

3. Histogram Selection

A histogram is used to describe discrete or continuous data and is one of the best ways to represent variables. In other words, it provides a visual interpretation of numeric data by displaying the number of data points that fall within a specified range of values (called a bin). In order to select the pertinent features, there are several methods, such as evaluating the individual features or groups of features. Several histogram selection methods based on graph construction or the measurement of similarity are introduced [25,26]. The histogram selection methods can be considered as an evaluation of the groups of features [27]. Histogram selection methods are usually grouped into three approaches: filter method, wrapper method, and embedded method. The latter involves a combination of the reduced processing time of the filtration method and the high efficiency of the encapsulation method. The filtering method is used to calculate the score of each histogram to measure its effectiveness, and then the histogram will be ranked according to the calculated

score. The histogram is evaluated using a specific classification algorithm, and the selected histograms are the histograms that maximize the classification rate.

To improve the classification performance, there are many proposed methods with the goal of reducing the dimensionality of the feature matrix. One such method is dimensionality reduction of the feature matrix based on the feature histogram, as proposed by Porebski et al. in 2013 [28]. In this method, the most important and significant histograms are selected based on the score value of each histogram. The approach to selecting characteristic histograms using ICS techniques has recently been extended to the multicolor space domain. Considering a database with N textured color images, each image $I_i, i \in \{1, 2, \ldots, N\}$ has a characteristic δ histogram. The entire set of data is represented by the matrix $\boldsymbol{H}$ as follows:

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}_1^1 & \ldots & \boldsymbol{h}_1^r & \ldots & \boldsymbol{h}_1^\delta \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \boldsymbol{h}_i^1 & \ldots & \boldsymbol{h}_i^r & \ldots & \boldsymbol{h}_i^\delta \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \boldsymbol{h}_N^1 & \ldots & \boldsymbol{h}_N^r & \ldots & \boldsymbol{h}_N^\delta \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}_1 \\ \ldots \\ \boldsymbol{h}_i \\ \ldots \\ \boldsymbol{h}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}^1 & \ldots & \boldsymbol{h}^r & \ldots & \boldsymbol{h}^\delta \end{bmatrix} \tag{4}$$

in which $\boldsymbol{h}_i^r$ is the rth histogram of the color image with texture i. $\boldsymbol{h}_i^r$ is defined as follows: $\boldsymbol{h}_i^r = h_i^r(1), \ldots, h_i^r(k), \ldots, h_i^r(Q)$, where Q is the bin number of the histogram.

The ICS technique is based on an in-class similarity method to evaluate the similarity between histograms extracted from images of the same class.

Let I_j^k be the k training image of class j, and class j has N_j images. Accordingly, the number of intersections of the histogram is calculated as follows:

$$D(I_j^k, I_j^{k'}) = \sum_{i=1}^{Q} min(h[I_j^k](i), h[I_j^{k'}](i)) \tag{5}$$

To measure the similarity of the class j, let SIM_j be the similarity measure, calculated as follows:

$$SIM_j = \frac{2}{N_j(N_j - 1)} \sum_{k=1}^{N_j - 1} \sum_{k'=k+1}^{N_j} D(I_j^k, I_j^{k'}) \tag{6}$$

Porebski et al., suggested that the higher the SIM_j in a class, the more relevant the histogram $\boldsymbol{h}^r$. Finally, to calculate the ICS score of a histogram $\boldsymbol{h}^r$ by:

$$S_{ICS}^r = \frac{1}{C} \sum_{j=1}^{C} SIM_j \tag{7}$$

where, C is the number of classes to be considered. S_{ICS}^r has a value from 0 to 1. The most distinct histogram is the one with the highest score of S_{ICS}^r.

4. Cloud Image Classification Process

4.1. Data Preparation

To our knowledge, the SWIMCAT dataset [29] is a benchmark for cloud image classification. It contains images taken with WAHRSIS, a calibrated image of the entire sky above ground, and consists of of 784 patches of 5 cloud types from images taken in Singapore between January 2013 and May 2014. The five sky types were identified based on visual features of sky/cloud conditions, consulting experts from the Singapore Meteorological Service. All patch images are 125 × 125 pixels; we take one representative image from each category. This dataset contains 784 images of sky/clouds, classified into 5 categories: clear sky, patterned clouds, thick dark clouds, thick white clouds, and cloud cover (Figure 3).

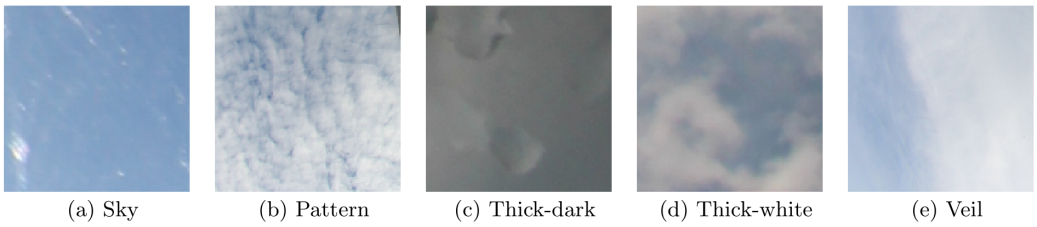


Figure 3. Selected images from five categories of the SWIMCAT dataset.

Although there have been many studies analyzing sky/cloud images captured by ground-based cameras by several research groups, publicly available standard databases are rare. Some datasets are released for cloud detection or segmentation purposes—for example, HYTA [30]. In addition, this paper also evaluates the proposed approach on the Cloud-ImVN 1.0 [31]. This dataset was created based on inspiration from the SWIMCAT dataset, with a larger number of images and more cloud/sky types. Specifically, the Cloud-ImVN 1.0 dataset has 2100 images of sky/clouds, classified into 6 categories: clear blue sky, patterned clouds, thick dark clouds, thick white clouds, thin white clouds, and cloud cover (Figure 4).

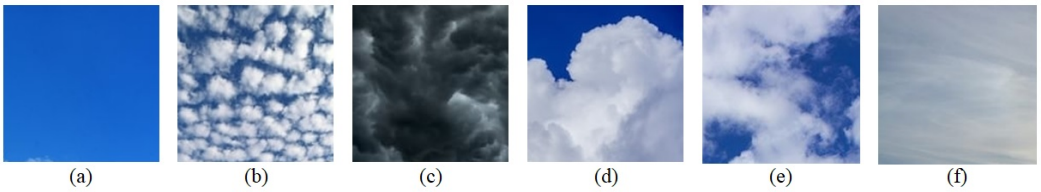


Figure 4. Six categories of cloud/sky images in the Cloud-ImVN 1.0 dataset: (**a**) clear blue sky, (**b**) patterned clouds, (**c**) thick dark clouds, (**d**) thick white clouds, (**e**) thin white clouds, (**f**) cloud cover.

4.2. Classification Process

The classification process of clouds images is illustrated in Figure 5. It consists of four steps as follows:

1. For each dataset, we apply 5-fold cross-validation in order to divide the data into training and testing subsets. The accuracy metric is employed to represent the performance on the testing set with standard deviation.
2. The feature extraction is applied for both testing and training sets using the LBP and LTP descriptors. For LBP, we consider different values of (R, P) parameters with $R \in \{1, 2, ..., 5\}$ and $P \in \{4, 8, 12\}$. For LTP, we use $t = 1$ for the SWIMCAT dataset and $t = 13$ for the Cloud-ImVN 1.0 dataset. In order to create robust features, the fusion of LBP and LTP is applied namely, LBP+LTP.
3. The ICS score is computed for each type of feature, LBP, LTP, and LBP+LTP, of the training set. The scores of each histogram are then arranged in descending order. The histogram selection method consists of selecting, during supervised training, a distinct subspace in which the classifier is used during testing.
4. We apply the SVM classifier to predict the label of the cloud image from the selected histogram in **Step 3**.

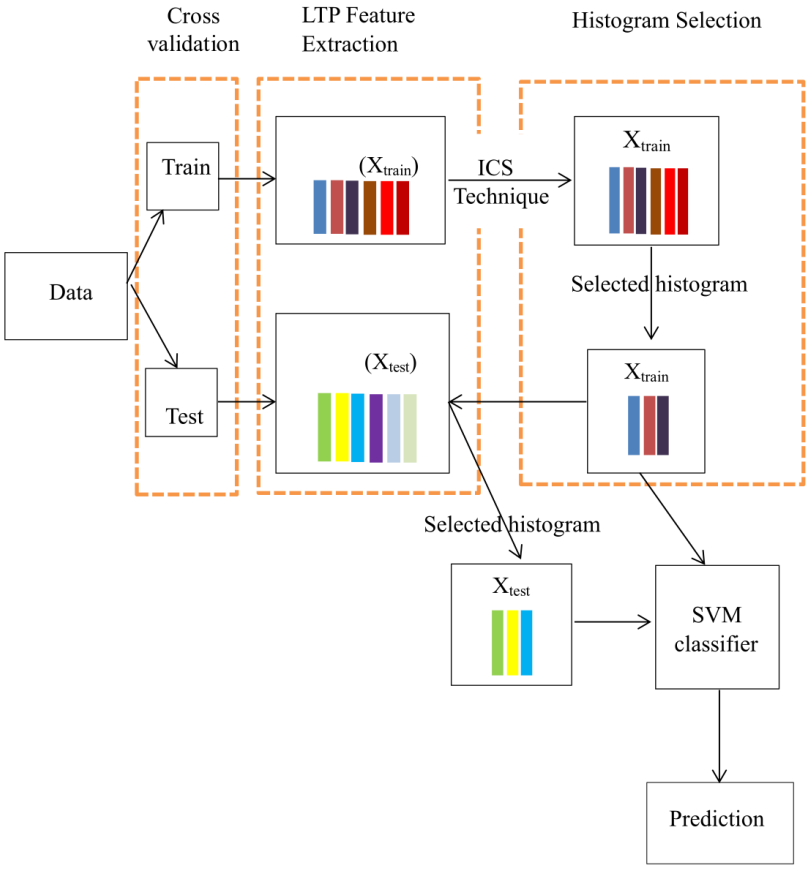


Figure 5. The cloud image classification process based on histogram selection using ICS technique.

5. Experimental Results

Previous studies focus on color features/color information of the RGB color space to classify clouds/sky images. However, the specific color space allows an improvement in the classification performance [32,33]. This work considers 14 different color spaces, such as HLS, HSV, IHLS, Lab, rgb, YIQ, YUV, RGB, bwrgby, XYZ, YCbCr, Luv, I1I2I3, ISH, for extracting features. Each extraction method applied obtains a corresponding feature—in this case, the value of (R, P), with $R \in \{1, 2, 3, 4, 5\}$ and $P \in \{4, 8, 12\}$. Thus, the input parameters for the feature extraction techniques LBP, LTP, and LBP+LTP will have 15 pairs of parameters (R, P), respectively. Each specific feature extraction technique, with a specific pair of input parameters, is applied on the 14 color spaces. Finally, histogram selection is applied for those features. In summary, the parameters used to run the experiment and give the results of cloud/sky image classification include: (1) a pair of parameters (R, P), (2) 14 color spaces, (3) 3 feature extraction methods: LBP, LTP, and LBP + LTP, (4) a selected number of histograms. Moreover, the dimension of each histogram is dependent on the value of P—for example, LBP features with $(1, 8)$ have 3 histograms from three color channels, and each histogram consists of $2^8 = 256$ bins. When applying the LTP descriptor to feature extraction, the number of histograms is doubled ($2 \times 2^8 = 512$ bins) compared with the LBP descriptor.

5.1. Results on the SWIMCAT Dataset

Table 1 is a summary and selective synthesis result obtained when running experiments on the applied SWIMCAT dataset technique to extract LBP, LTP, and LBP + LTP features of 14 color spaces. The highest value for each feature is underlined. We observe that when classifying cloud/sky images in the SWIMCAT dataset, if using more color variables, the results are better than when using grayscale images, specifically with the RGB

color system, with ACC reaching 97.8 ± 1.3 for the LBP (1, 12) technique, ACC reaching 97.9 ± 1.2 for LTP (2, 12), and ACC reaching 98.1 ± 1.3 for technique LBP+LTP (3, 12).

Table 2 presents the results obtained on the SWIMCAT dataset incorporating the histogram selection method. The highest value for each feature is underlined. For color spaces with H color components and S color components, using the LTP feature gives better results than using the LBP feature. Specifically, the HLS color space reaches 98.9 ± 0.8 with the LBP (4, 12) feature and reaches 99.0 ± 0.7 with the characteristic LTP (4, 8); the ISH color system reached 98.6 ± 0.6 with the LBP (4, 12) feature and 99.0 ± 0.4 with the LTP (4, 8) feature.

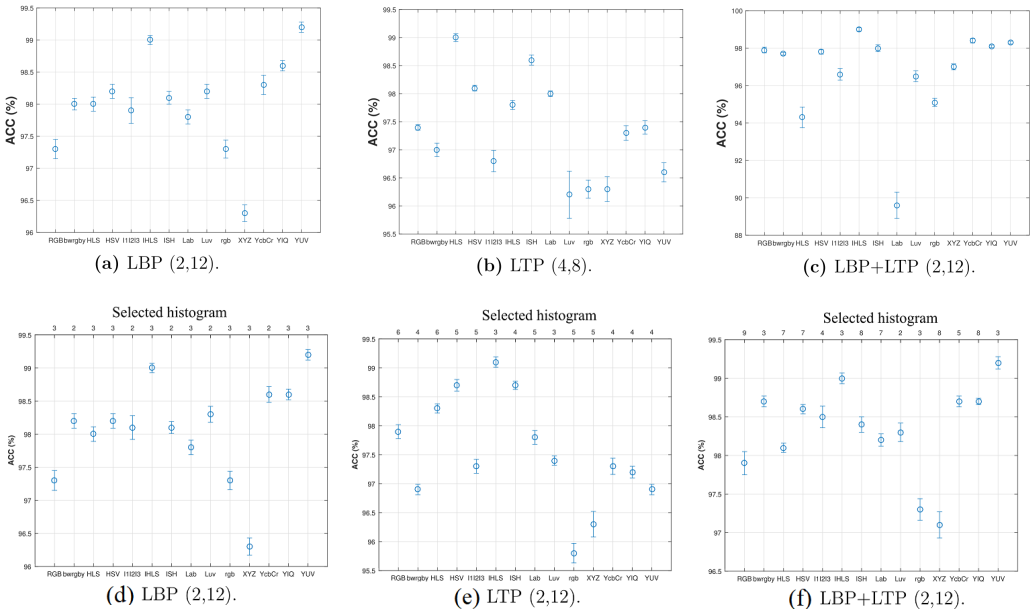
Table 1. The best results obtained for SWIMCAT dataset without using histogram selection method. Some of the best results for each color space, each feature, and the selected parameters (R, P).

	(R, P)	HLS	HSV	IHLS	ISH	Lab	YIQ	YUV
LBP	(1, 8)						98.7 ± 0.4	
	(3, 8)		98.5 ± 0.7					
	(1, 12)					99.0 ± 0.5		
	(2, 12)			99.0 ± 0.7				$\underline{99.2 \pm 0.8}$
	(4, 12)	98.9 ± 0.8			98.6 ± 0.6			
LTP	(2, 8)				98.7 ± 1.0			
	(4, 8)	$\underline{99.0 \pm 0.7}$				98.0 ± 0.5		
	(1, 12)			98.3 ± 0.7				
	(2, 12)		98.7 ± 1.1					
LBP + LTP	(1, 8)						98.3 ± 0.7	
	(3, 8)	98.5 ± 1.2	98.2 ± 0.9		98.6 ± 0.9			
	(1, 12)					98.5 ± 1.0		
	(2, 12)			$\underline{99.0 \pm 0.9}$				98.3 ± 0.8

Table 2. The best results obtained for SWIMCAT dataset while using histogram selection method. Some of the best results for each color space, each feature, and the selected parameters (R, P).

	(R, P)	HLS	HSV	IHLS	ISH	Lab	YIQ	YUV
Number of selected histograms		3 HISTO	3 HISTO	3 HISTO	3 HISTO	3 HISTO	3 HISTO	3 HISTO
LBP	(1, 8)						98.7 ± 0.4	
	(3, 8)		98.5 ± 0.7					
	(1, 12)					99.0 ± 0.5		
	(2, 12)			99.0 ± 0.7				$\underline{99.2 \pm 0.8}$
	(4, 12)	98.9 ± 0.8			98.6 ± 0.6			
Number of selected histograms		6 HISTO	4 HISTO	3 HISTO	4 HISTO	5 HISTO		
LTP	(3, 8)		98.7 ± 1.0					
	(4, 8)	99.0 ± 0.7			99.0 ± 0.4			
	(5, 8)					98.1 ± 0.7		
	(2, 12)			$\underline{99.1 \pm 0.9}$				
Number of selected histograms		5 HISTO	7 HISTO	3 HISTO	7 HISTO	2 HISTO	4 HISTO	3 HISTO
LBP + LTP	(2, 8)	98.5 ± 0.4						
	(4, 8)				98.7 ± 0.4			
	(1, 12)						99.0 ± 0.6	
	(2, 12)			99.0 ± 0.7				$\underline{99.2 \pm 0.8}$
	(3, 12)		98.8 ± 1.1			98.5 ± 0.8		

Figure 6 presents the selected results of the highest accuracy obtained from three types of features (LBP, LTP, LBP+LTP) in the two scenarios: without and with the histogram selection method.



(a) LBP (2,12). (b) LTP (4,8). (c) LBP+LTP (2,12).

(d) LBP (2,12). (e) LTP (2,12). (f) LBP+LTP (2,12).

Figure 6. Highest results obtained on the SWIMCAT dataset without (first row) and with (second row) histogram selection method.

Table 3 presents the comparison of the results obtained on the SWIMCAT dataset with previous studies. Thus, the highest results are obtained on the SWIMCAT dataset for three types of features (Table 3) using the ICS technique, as follows:

- The LBP feature has the highest ACC of 99.2 for the YUV color system at $P = 2$, $R = 12$, with 12,288 features (3 histograms).
- The LTP feature has the highest ACC 99.1 for the IHLS color system at $R = 2$, $P = 12$, with 12,288 features (3 histograms).
- Features LBP+LTP achieved the highest ACC 99.2 for the YUV color system with $R = 2$, $P = 12$, with 12,288 features (3 histograms).

With the characteristic of LBP when using the ICS technique, there is no significant difference in results when not using the ICS technique. However, with the characteristics of LTP and LBP+LTP when using the ICS technique for better results, the number of histograms selected is also smaller than when not using the ICS technique to select histograms. Moreover, with the SWIMCAT dataset, the LTP feature gives better results than the LBP feature (Table 2). The proposed approach clearly outperforms LBP variants such as WLBP, SRBP, and SWOBP. For example, the SaLBP technique is based on the LBP uniform, which has had many of the bins that usually arise eliminated.

Table 3. Comparison of experimental results on SWIMCAT dataset using ICS with previous studies.

Feature	Color Space	Number of Features	Accuracy
LBP (2, 16) [23]	Grayscale	65,536	79.30
LTP (2, 16) [24]	Grayscale	65,536	81.80
DLBP (2, 16) [34]	Grayscale	65,536	84.20
WLBP [35]	RGB	-	84.52
CLBP (2, 16) [36]	Grayscale	65,536	85.50

Table 3. *Cont.*

Feature	Color Space	Number of Features	Accuracy
ISRBP [37]	RGB	3840	88.72
SaLBP (2, 16) [15]	Grayscale	65,536	89.00
Task-based graph CNN [38]	RGB	2560	89.48
ResNet 50 [39]	RGB	150,528	92.80
SRBP [14]	RGB	3840	94.36
Statistical descriptor [20]	RGB	4800	94.87
SWOBP [40]	RGB	996	94.90
ACPS [41]	RGB	768	96.67
LBP + LTP (3, 12) (without ICS) (this paper)	RGB	12,288	98.10
Covariance descriptor [42]	RGB	3328	98.40
Texture and manifold features [43]	RGB	32,768	97.72
SLBP (2, 16) [17]	Grayscale	65,536	97.17
MKSCDL [44]	RGB	9216	98.34
LBP + LTP (2, 12) with ICS (this paper)	RGB	12,288	98.10
Optimized CNN Cloud Network [45]	RGB	9216	98.60
SDCAF [46]	RGB	9216	98.70
LBP (2, 12) with ICS (this paper)	YUV	12,288	**99.20**
LTP (2, 12) with ICS (this paper)	IHLS	12,288	**99.19**
LBP + LTP (2, 12) with ICS (this paper)	YUV	12,288	**99.20**

5.2. Results on the Cloud-ImVN 1.0 Dataset

Table 4 shows the best results obtained with different types of features on the Cloud-ImVN 1.0 dataset. We observe the appearance of the HLS, HSV, and IHLS color spaces. It shows that color components with high dichroism or with H (Hue) and S (Saturation) components can still be good candidates for cloud/sky image classification. Moreover, we observe and confirm that the RGB space is not the best color space for characterizing cloud images. LTP features achieve higher accuracy than LBP features and have lower standard deviation. LBP features achieve the highest accuracy, with 85.4 $\pm$ 4.6, and LTP achieves the highest accuracy at 88.1 $\pm$ 2.2. In total, the combination of LBP and LTP features achieves the best accuracy at 92.2 $\pm$ 2.4.

Table 5 presents the obtained result on the Cloud-ImVN 1.0 dataset while using the histogram selection method on different parameters: color space, features used, (R, P) values. When applying the ICS method, the results of cloud/sky image classification change significantly: the results are higher and the number of features is highly reduced. Considering Table 5, the LBP feature (3,12) in the RGB color system achieved the highest ACC of 89.3 $\pm$ 2.2 with one histogram selected, while, when not using ICS, the LBP feature (5, 12) in the RGB color system achieved the highest ACC of 81.9 $\pm$ 3.3.

Table 4. The best results obtained for Cloud-ImVN 1.0 dataset without using histogram selection method. The best results for each color space, each feature, and the selected parameters (R, P).

	(R, P)	RGB	HLS	HSV	IHLS	ISH	Lab	Luv	YcbCr
LBP	(1, 4)						82.8 ± 8.1	$\underline{85.4 \pm 4.6}$	
	(4, 12)								
	(5, 12)	81.9 ± 3.3							
LTP	(2, 4)					87.6 ± 0.9			
	(2, 8)			$\underline{88.1 \pm 2.2}$	84.5 ± 1.0				
	(5, 8)							82.1 ± 1.5	
	(1, 12)		86.0 ± 1.2						
	(5, 12)								82.3 ± 2.7
LBP + LTP	(1, 4)						$\underline{92.2 \pm 2.4}$	87.3 ± 2.7	

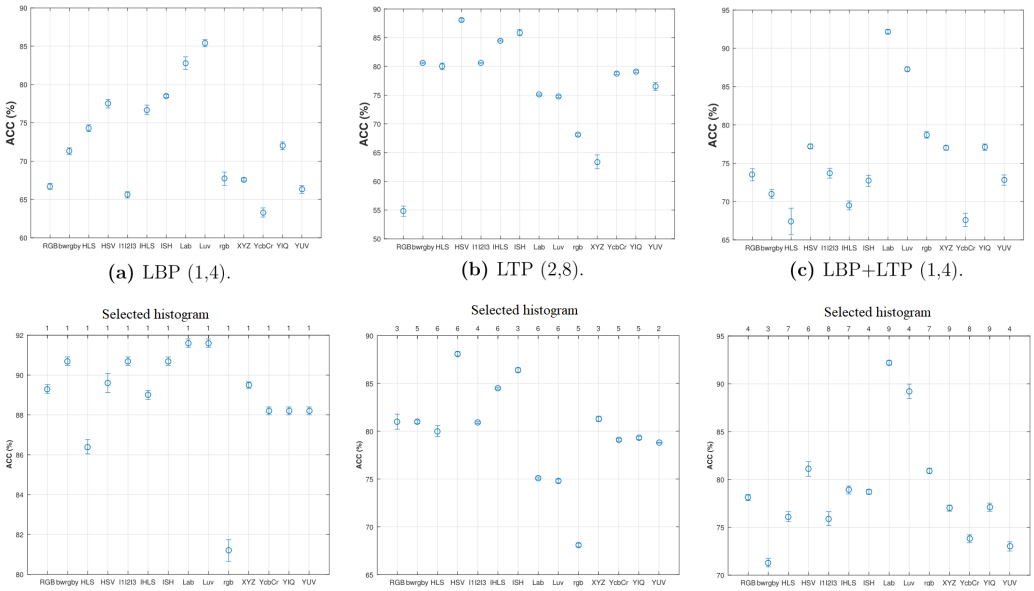
Table 5. The best results obtained for Cloud-ImVN 1.0 dataset while using histogram selection method. The best results for each color space, each feature, and the selected parameters (R, P).

	(R, P)	RGB	bwrgby	HLS	I1I2I3	HSV	IHLS	ISH	Lab	Luv
Number of selected histograms		1 HISTO	1 HISTO	1 HISTO	1 HISTO		1 HISTO	1 HISTO	1 HISTO	1 HISTO
LBP	(1, 8)			89.5 ± 2.0						
	(3, 12)	89.3 ± 2.2	90.7 ± 2.1		90.7 ± 2.1			90.7 ± 2.1	$\underline{91.6 \pm 2.1}$	$\underline{91.6 \pm 2.1}$
	(4, 12)									90.2 ± 1.2
Number of selected histograms				6 HISTO		6 HISTO	6 HISTO	6 HISTO		
LTP	(2, 4)							87.6 ± 0.9		
	(2, 8)					$\underline{88.1 \pm 2.2}$	84.5 ± 1.0			
	(1, 12)			86.0 ± 1.2						
Number of selected histograms		1 HISTO		1 HISTO			1 HISTO		9 HISTO	1 HISTO
LBP + LTP	(1, 4)								$\underline{92.2 \pm 2.4}$	
	(1, 8)			89.5 ± 2.0						
	(1, 12)	88.5 ± 2.4								
	(3, 12)	89.3 ± 2.2					89.0 ± 2.3			91.6 ± 2.1
	(4, 12)									90.2 ± 1.2

Figure 7 presents the selected results of the highest accuracy obtained from three types of features (LBP, LTP, LBP+LTP) in the two scenarios: without and with the histogram selection method.

Table 6 presents the comparison of the results obtained on the Cloud-ImVN 1.0 dataset with previous studies. Thus, the highest results are obtained on this dataset for three types of features using the ICS technique, as follows:

- LBP achieved the highest ACC 91.6 for the Lab and Luv color space at $R = 3$, $P = 12$, with 4096 features.
- The LTP feature has the highest ACC of 88.1 for the HSV color space at $R = 2$, $P = 8$, with 1536 features.
- Features LBP+LTP have the highest ACC of 92.2 for the Lab color space at $R = 1$, $P = 4$, with only 144 features.



(a) LBP (1,4). (b) LTP (2,8). (c) LBP+LTP (1,4).

(d) LBP (3,12). (e) LTP (2,8). (f) LBP+LTP (1,4).

Figure 7. Highest results obtained on the Cloud-ImVN 1.0 dataset without (first row) and with (second row) histogram selection method.

Table 6. Comparison of experimental results on the Cloud-ImVN 1.0 dataset.

Features	Color Space	Number of Features	Accuracy
ISRBP [37]	RGB	3840	51.05
SRBP [14]	RGB	3840	62.10
LBP (1, 8) [23]	RGB	768	73.80
SWOBP [40]	RGB	996	76.95
LTP (5, 12) (without ICS) (this paper)	RGB	12,288	81.90
LTP (1, 8) [24]	RGB	1536	84.90
LBP (3, 12) (ICS) (this paper)	Lab, Luv	4096	91.60
LTP (2, 8) (ICS) (this paper)	HSV	1536	88.10
LBP + LTP (3, 12) (ICS) (this paper)	RGB	4096	89.30
LBP + LTP (1, 4) (ICS) (this paper)	Lab	144	**92.20**

There are many image processing algorithms, classifying images mainly based on the RGB color space. However, for the classification of cloud images, it seems that the RGB color space carries many disadvantages and is not a good candidate. A color space with high dichroism and with high luminance components is a good candidate because the luminance in cloudy areas is higher than in others. Similarly to the results obtained on the SWIMCAT dataset, for Cloud-ImVn 1.0, the LTP descriptor gives better results than other LBP variants.

6. Conclusions

The existing cloud features are very useful in determining color space and texture features to classify cloud types. In order to be able to classify sky/clouds, it is essential to distinguish between two types of pixels (sky and clouds); a suitable color space can facilitate this classification. High-dichroism color systems are good candidates for cloud/sky image classification. In addition, color systems with high luminance components are also good candidates because the luminance in cloudy areas is higher than in cloudless areas.

This paper presents and systematically analyzes various features developed for the task of cloud/sky image classification. We found that the LBP and LTP feature extraction techniques generalized well to this objective. We integrate the color space and texture structure with the LBP feature and LTP feature effectively to obtain higher classification accuracy. Integrating color features and texture into cloud/sky image classification also enhances the performance. In the experiment, this work integrates color features to increase the efficiency of the feature extraction process, and using the ICS technique to select potential histograms allows us to enhance the performance clearly, with fewer features. However, the parameter value t, obtained using exhaustive techniques for the RGB color system, may affect the results of other color spaces.

By exploiting different aspects of sky/cloud images through ground-based sky/cloud images, the proposed method has solved the basic problems of color system processing and feature selection. Below are several future directions of the topic:

- In the scope of this study, parameter value t was experimentally obtained by using exhaustive techniques for the RGB color space, so future work should study methods to optimize the value of t for the LTP feature.
- In the future, we will study the scoring method for each feature, using it in combination with the histogram selection method to select the histograms that have the most potential features.

Author Contributions: K.T.-T.: Data curation, Software, Writing—original draft, Investigation, Formal analysis; H.D.T.H.: Data curation, Investigation, Conceptualization, Methodology; V.T.H.: Conceptualization, Methodology, Validation, Writing—review & editing, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by Ho Chi Minh City Open University (HCMCOU) and the Ministry of Education and Training (Vietnam) under grant number B2021-MBS-07.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Satapathy, J.; Thapliyal, P. Retrieval of cloud-cleared radiances using numerical weather prediction model-analysis and forecast fields for INSAT-3D sounder longwave window channel observations. *J. Atmos. Sol. Terr. Phys.* **2021**, *217*, 105602. [CrossRef]
2. Lin, F.; Zhang, Y.; Wang, J. Recent advances in intra-hour solar forecasting: A review of ground-based sky image methods. *Int. J. Forecast.* **2022**, *in press*. [CrossRef]
3. Ahn, H.; Lee, S.; Ko, H.; Kim, M.; Han, S.W.; Seok, J. Searching similar weather maps using convolutional autoencoder and satellite images. *ICT Express* **2022**, *in press*. [CrossRef]
4. Román, R.; Cazorla, A.; Toledano, C.; Olmo, F.; Cachorro, V.; de Frutos, A.; Alados-Arboledas, L. Cloud cover detection combining high dynamic range sky images and ceilometer measurements. *Atmos. Res.* **2017**, *196*, 224–236. [CrossRef]
5. Costa-Surós, M.; Calbó, J.; González, J.; Martin-Vide, J. Behavior of cloud base height from ceilometer measurements. *Atmos. Res.* **2013**, *127*, 64–76. [CrossRef]
6. Wang, M.; Zhou, S.; Yang, Z.; Liu, Z. CloudA: A Ground-Based Cloud Classification Method with a Convolutional Neural Network. *J. Atmos. Ocean. Technol.* **2020**, *37*, 1661–1668. [CrossRef]
7. Dev, S. 2D and 3D Image Analysis and Its Application to Sky/Cloud Imaging. Ph.D. Thesis, Nanyang Technological University, Singapore, 2017; p. 246.
8. WMO. *WMO International Cloud Atlas, Preface to the 2017 Edition*; WMO: Geneva, Switzerland, 2017.
9. Kliangsuwan, T.; Heednacram, A. FFT features and hierarchical classification algorithms for cloud images. *Eng. Appl. Artif. Intell.* **2018**, *76*, 40–54. [CrossRef]
10. Li, Q.; Zhang, Z.; Lu, W.; Yang, J.; Ma, Y.; Yao, W. From pixels to patches: A cloud classification method based on a bag of micro-structures. *Atmos. Meas. Tech.* **2016**, *9*, 753–764. [CrossRef]
11. Zhen, Z.; Wang, F.; Sun, Y.; Mi, Z.; Liu, C.; Wang, B.; Lu, J. SVM based cloud classification model using total sky images for PV power forecasting. In Proceedings of the 2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 18–20 February 2015; pp. 1–5. [CrossRef]
12. Taravat, A.; Del Frate, F.; Cornaro, C.; Vergari, S. Neural Networks and Support Vector Machine Algorithms for Automatic Cloud Classification of Whole-Sky Ground-Based Images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 666–670. [CrossRef]

13. Liu, S.; Zhang, Z. Learning group patterns for ground-based cloud classification in wireless sensor networks. *EURASIP J. Wirel. Commun. Netw.* **2016**, *2016*, 69. [CrossRef]
14. Boudra, S.; Yahiaoui, I.; Behloul, A. A set of statistical radial binary patterns for tree species identification based on bark images. *Multimed. Tools Appl.* **2021**, *80*, 22373–22404. [CrossRef]
15. Liu, S.; Wang, C.; Xiao, B.; Zhang, Z.; Shao, Y. Salient local binary pattern for ground-based cloud classification. *Acta Meteorol. Sin.* **2013**, *27*, 211–220. [CrossRef]
16. Zhang, Z.; Li, D.; Liu, S.; Xiao, B.; Cao, X. Multi-View Ground-Based Cloud Recognition by Transferring Deep Visual Information. *Appl. Sci.* **2018**, *8*, 748. [CrossRef]
17. Wang, Y.; Shi, C.; Wang, C.; Xiao, B. Ground-based cloud classification by learning stable local binary patterns. *Atmos. Res.* **2018**, *207*, 74–89. [CrossRef]
18. Zhang, R.; Yang, S.; Zhang, Q.; Xu, L.; He, Y.; Zhang, F. Graph-based few-shot learning with transformed feature propagation and optimal class allocation. *Neurocomputing* **2022**, *470*, 247–256. [CrossRef]
19. Zhang, R.; Xu, L.; Yu, Z.; Shi, Y.; Mu, C.; Xu, M. Deep-irtarget: An automatic target detector in infrared imagery using dual-domain feature extraction and allocation. *IEEE Trans. Multimed.* **2021**, *24*, 1735–1749. [CrossRef]
20. Cheng, H.Y.; Yu, C.C. Block-based cloud classification with statistical features and distribution of local texture features. *Atmos. Meas. Tech.* **2015**, *8*, 1173–1182. [CrossRef]
21. Jose, J.A.; Kumar, C.S.; Sureshkumar, S. Tuna classification using super learner ensemble of region-based CNN-grouped 2D-LBP models. *Inf. Process. Agric.* **2022**, *9*, 68–79. [CrossRef]
22. Tian, D. *Advanced Data Mining Techniques*; Springer: Berlin/Heidelberg, Germany, 2008. [CrossRef]
23. Ojala, T.; Pietikäinen, M.; Harwood, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit.* **1996**, *29*, 51–59. [CrossRef]
24. Tan, X.; Triggs, B. Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions. *IEEE Trans. Image Process.* **2010**, *19*, 1635–1650. [CrossRef]
25. Hoang, V.T.; Porebski, A.; Vandenbroucke, N.; Hamad, D. LBP Histogram Selection based on Sparse Representation for Color Texture Classification. In Proceedings of the Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Porto, Portugal, 27 February–1 March 2017; SCITEPRESS—Science and Technology Publications: Setubal, Portugal, 2017; pp. 476–483. [CrossRef]
26. Hoang, V.T. Unsupervised LBP histogram selection for color texture classification via sparse representation. In Proceedings of the 2018 IEEE International Conference on Information Communication and Signal Processing (ICICSP), Singapore, 28–30 September 2018; pp. 79–84.
27. Hu, Z.; Yin, H.; Liu, Y. Locally linear embedding vote: A novel filter method for feature selection. *Measurement* **2022**, *190*, 110535. [CrossRef]
28. Porebski, A.; Vandenbroucke, N.; Hamad, D. LBP histogram selection for supervised color texture classification. In Proceedings of the 2013 IEEE International Conference on Image Processing, Melbourne, Australia, 15–18 September 2013; pp. 3239–3243. [CrossRef]
29. Dev, S.; Savoy, F.M.; Lee, Y.H.; Winkler, S. WAHRSIS: A low-cost high-resolution whole sky imager with near-Infrared capabilities. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXV*; SPIE: Paris, France, 2014; p. 90711L. [CrossRef]
30. Li, Q.; Lu, W.; Yang, J. A Hybrid Thresholding Algorithm for Cloud Detection on Ground-Based Color Images. *J. Atmos. Ocean. Technol.* **2011**, *28*, 1286–1296. [CrossRef]
31. Truong Hoang, V. *Cloud-ImVN 1.0*; Type: Dataset; Ho Chi Minh City Open University: Ho Chi Minh City, Vietnam, 2020. [CrossRef]
32. Simon, A.P.; Uma, B. DeepLumina: A Method Based on Deep Features and Luminance Information for Color Texture Classification. *Comput. Intell. Neurosci.* **2022**, *2022*, 9510987. [CrossRef] [PubMed]
33. Yang, J.; Liu, C.; Zhang, L. Color space normalization: Enhancing the discriminating power of color spaces for face recognition. *Pattern Recognit.* **2010**, *43*, 1454–1466. [CrossRef]
34. Liao, S.; Law, M.; Chung, A. Dominant Local Binary Patterns for Texture Classification. *IEEE Trans. Image Process.* **2009**, *18*, 1107–1118. [CrossRef]
35. Liu, S.; Zhang, Z.; Mei, X. Ground-based cloud classification using weighted local binary patterns. *J. Appl. Remote Sens.* **2015**, *9*, 095062. [CrossRef]
36. Guo, Z.; Zhang, L.; Zhang, D. A Completed Modeling of Local Binary Pattern Operator for Texture Classification. *IEEE Trans. Image Process.* **2010**, *19*, 1657–1663. [CrossRef]
37. Boudra, S.; Yahiaoui, I.; Behloul, A. Bark identification using improved statistical radial binary patterns. In Proceedings of the 2018 International Conference on Content-Based Multimedia Indexing (CBMI), La Rochelle, France, 4–6 September 2018; pp. 1–6.
38. Liu, S.; Li, M.; Zhang, Z.; Cao, X.; Durrani, T.S. Ground-based cloud classification using task-based graph convolutional network. *Geophys. Res. Lett.* **2020**, *47*, e2020GL087338. [CrossRef]
39. Li, M.; Liu, S.; Zhang, Z. Dual guided loss for ground-based cloud classification in weather station networks. *IEEE Access* **2019**, *7*, 63081–63088. [CrossRef]
40. Song, T.; Feng, J.; Wang, S.; Xie, Y. Spatially weighted order binary pattern for color texture classification. *Expert Syst. Appl.* **2020**, *147*, 113167. [CrossRef]

41. Pan, Z.; Hu, S.; Wu, X.; Wang, P. Adaptive center pixel selection strategy in Local Binary Pattern for texture classification. *Expert Syst. Appl.* **2021**, *180*, 115123. [CrossRef]
42. Tang, Y.; Yang, P.; Zhou, Z.; Pan, D.; Chen, J.; Zhao, X. Improving cloud type classification of ground-based images using region covariance descriptors. *Atmos. Meas. Tech.* **2021**, *14*, 737–747. [CrossRef]
43. Luo, Q.; Meng, Y.; Liu, L.; Zhao, X.; Zhou, Z. Cloud classification of ground-based infrared images combining manifold and texture features. *Atmos. Meas. Tech.* **2018**, *11*, 5351–5361. [CrossRef]
44. Luo, Q.; Zhou, Z.; Meng, Y.; Li, Q.; Li, M. Ground-based cloud-type recognition using manifold kernel sparse coding and dictionary learning. *Adv. Meteorol.* **2018**, *2018*, 9684206. [CrossRef]
45. Shi, C.; Wang, C.; Wang, Y.; Xiao, B. Deep convolutional activations-based features for ground-based cloud classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 816–820. [CrossRef]
46. Zhang, J.; Liu, P.; Zhang, F.; Song, Q. CloudNet: Ground-based cloud classification with deep convolutional neural network. *Geophys. Res. Lett.* **2018**, *45*, 8665–8672. [CrossRef]