We thank the reviewers for the insightful and thoughtful comments, which have allowed us to produce a stronger manuscript. Our responses to the comments are given below, and the corresponding changes are tracked at specific location **(Page X, Line X)** in the revised manuscript.

Response to Referee #1

General comments:

The machine classification of cloud types found in automatically recorded images is an aim of considerable importance. However, it has proved difficult to develop suitable algorithms for this task. This paper combines two approaches, a texture analysis, such as one might expect based on statistical examination of the image structure, plus a manifold analysis such as is found effective, for instance, in facial recognition. The paper demonstrates that this combined process represents an improvement on previous analyses. The paper is interesting, well presented, and should be publishable.

The progress represented by this paper is incremental and the ultimate aim of classifying any cloud image is still distant. The paper demonstrates an ability to analyse images of "high visual quality" and avoids "a complex mixture of cloud types" in its dataset, which is used both for training and analysis when manually analysed and then split into various groups. We are NOT addressing images with clouds of mixed types at various formation levels, which is a not-uncommon occurrence (noted as a next step in the Conclusions). However, the difficult issue of examining clouds away from the zenith, where the aspect of the cloud changes, is addressed with reasonable success.

Specific comments:

1. Comment: The images under study are recorded in the long-wave infra-red. In these cases, the clear sky background brightness (temperature) varies with time and zenith angle (clear in figures 3 b, e). There is no discussion of whether those variations affect the image analysis, particularly the textural features which may have baseline issues.

Response: (Page 3, Line 23-29) Many thanks for the thoughtful commenting. It is true that the clear sky background radiance in 8-14 μm varies with time and zenith angle. The images of the datasets have been preprocessed in the consideration of this important factor. The clear sky radiance threshold in each image is calculated using the radiation transfer model (Liu et al., 2013). The real radiance R at each pixel in each image is converted to the grey value G_{pixel} between [0,255] with $G_{pixel} = R/(R_{temp} - R_{clear}) \times 255$, where R_{clear} is the corresponding clear sky radiance threshold and R_{temp} is the radiance corresponding to the real-time environment temperature. As a result, the effects of the clear sky background brightness temperature can be ignored, which means that this factor has little influence on the feature extraction of the images. In the revised manuscript, a further description of the cloud image preprocessing will be added in Section 2.1.

2. Comment: In the same sense, the camera (if radiometric) provides real information on the apparent temperature of the cloud, and this is unused.

Response: (Page 11, Line 25-27) Thanks for your constructive comments of the manuscript. The real information on the apparent temperature of the cloud is useful for analysing the images. The radiance value at each pixel in the image corresponds to a bright temperature. How to utilize the information of the brightness temperature effectively is worth further studying. At present, we mainly use the texture

and manifold features to identify the cloud type. The addition of the brightness temperature of the physical information, or the combination of the height information obtained from the laser ceilometer might be helpful for the improvement of the cloud type recognition accuracy. We will carry out the research in the following work. Thanks again for the reviewer's advice. We will add this point to the section of Conclusions in the revised manuscript.

3. Comment: So far as presentation is concerned, the paper is clearly written. Tables 4 and 5 should include some explanation of the fractions 1/10, 1 /2, 9/10 even though their meaning is clear from a reading of the text (minor revision).

Response: (Page 18, 19) Many thanks for the careful reading of our manuscript. We are sorry for our negligence. We will add some explanation of the fractions 1/10, 1 /2, 9/10 in Tables 4 and 5 in the revised manuscript to make it clear: 1/10, 1/2 and 9/10 are the certain proportions of the training set selected randomly from each category, and the rest part forms the testing set correspondingly.

Response to Referee #2

General comments:

In the article "Cloud Classification of ground-based infrared images combining manifold and texture features" the authors introduce a new method to identify cloud regimes from ground based cloud imagers. They base their study on a dataset from the Whole-Sky Infrared Cloud Measurement System, which provides zenith and whole sky images. The clouds in the images have been classified by two independent experts, and the image is only used in case they agree. The authors define a feature vector basing on the grey level co-occurrence matrix, which provides the measures energy, entropy, contrast and homogeneity; and on manifold features which are constructed by computing the regional covariance descriptor and mapping it into its tangent space. With their new method, the authors reach a slightly higher accuracy (by 3 to 5 %) compared to earlier methods of Liu and Cheng. The structure and the content of the paper are ok, but it lacks clarity in several places. Also, the language should be improved. I recommend major revisions. Please find my specific comments below.

Response: Many thanks for the constructive and valuable comments of the manuscript. We have tried our best to enrich the paper's content to make it clearer. Besides, we have polished the language as far as we can to satisfy the requirement of publication.

Specific comments:

1. Comment: The article lacks some clarity. With regard to the underlying dataset, the authors explain how it is obtained and how the cases are chosen. However, in Section 3, "Experiments and discussion" the authors talk of "conducting each experiment 50 times on two datasets" (p.6, l.25). It is not clear to me what experiments are meant and in which way it can be repeated 50 times. Please state very clearly what you do and mean by this.

Response: (Page 8, Line 29-Page 9, Line 6) We are sorry not to make it clear. In Section 3, "Experiments and discussion" validates which features are chosen, presents the results and gives the discussion. We first adopt 10-fold cross validation to determine which features among texture features, manifold features and combined features perform best. 10-fold cross validation means that each dataset is divided into 10 subsets with the same size at random in turn, then one single subset is used for validation and the other 9 parts are taken as the training set. To test the performance of the algorithm, 10-fold cross validation is conducted 50 times and average values are taken as final results.

The combined features are used for the cloud type recognition experiments. Different from a deterministic case, the training samples of the experiments are chosen randomly. The purpose that we conduct the experiments is to test the performance of the algorithm. In each experiment, the training set is selected at random and the rest of the dataset forms the testing set, so we need to repeat this process many times and calculate the mean value in order to reduce the accidental bias and to measure the algorithm well.

2. Comment: You mention a "Support Vector Machine" (SVM), which is used to perform the cloud classification. It is not clear what that actually is. Please extend the respective part a little bit, or give a citation at the very least.

Response: (Page 8, Line 8-21) Many thanks for the constructive comments of the manuscript. In the area of machine learning, Support Vector Machines (SVMs) are supervised learning models (Cristianini and Shawe-Taylor, 2000). An SVM model is a representation of the examples as points in the Reproducing Kernel Hilbert Space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

As Fig.1 shows, given a set of two-class training examples (denoted by × and o, respectively), the key problem is to find the optimal hyperplane to do the separation: $w^T x + b = 0$, where w is a weight vector and b is a bias, and an SVM training model with the largest margin ($2/\sqrt{w^T w}$) is built. The support vectors are the samples on the dotted lines. The optimization classification hyperplane is determined by the solid line. The test examples are assigned to one category or the other based on this model, making it a non-probabilistic binary linear classifier. In this work, we apply a simple linear function as the mapping kernel.

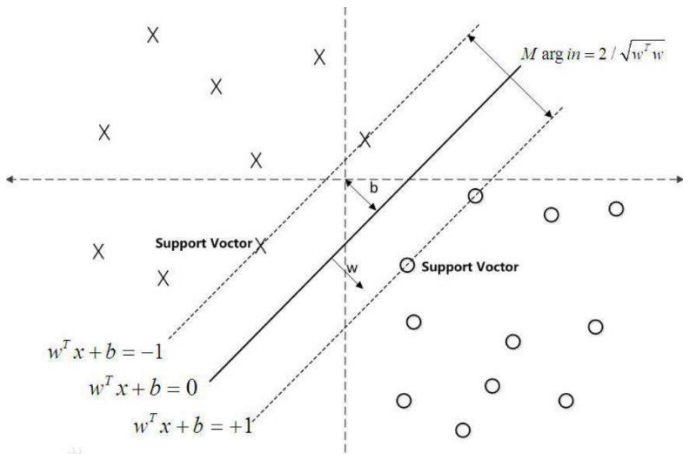

Figure 1: The decision boundary of support vector machine with the largest margin. × and o denote two-class training examples, respectively. $w^T x + b = 0$ is the optimal hyperplane to do the separation, where w is a weight vector and b is a bias, and an SVM training model with the largest margin $2/\sqrt{w^T w}$ is built. The support vectors are the samples on the dotted lines. The optimization classification hyperplane is determined by the solid line.

3. Comment:

1) The mathematical framework of a manifold is explained in quite detail. It distracts a little bit from the final result, the feature vector (supposedly Eq. 11).

2) It would add much clarity to extend Sec. 2.2.3 "Combining manifold and texture features" and clearly state what you are now using for a cloud classification.

3) You could also add information about the SVM here.

4) Overall it is difficult to assess why this manifold features play a bigger role, or if for example other parameters would add equally much information.

5) The mathematics behind the manifold feature vector is rather complicated, and a physical interpretation is hardly possible. Why has this vector been chosen?

6) Would the addition of different, more easily physically interpretable parameters also lead to a higher hit rate in the classification?

7) Is this more a "fitting problem" (more parameters → better fit) or is it more physically based? Please justify the choice of your metrics more.

Response: Many thanks for the constructive comments of the manuscript.

1) **(Page 6, Line 9-Page 7, Line 27)** Generally speaking, the manifold is a topological space that is locally equivalent to a Euclidean space. The differential manifold has a globally defined differential structure. Its tangent space $T_X M$ is a space formed by all possible tangent vectors at a given point X on the differential manifold. For the Riemannian manifold M, an inner product is defined in its tangent space. The shortest curve between two points on the manifold is called a geodesic and the length of the geodesic is the shortest distance between two points.

All SPD matrices form a Riemannian manifold. Suppose S^d is a set of all $n \times n$ real symmetric matrices: $S^d = \{A \in M(d): A^{\mathrm{T}} = A\}$, where $M(d)$ represents the set of all $d \times d$ matrices, so that $S_{++}^d = \{A \in S^d: A > 0\}$ is the set of all $d \times d$ SPD matrices, which construct a $d(d+1)/2$ dimensional SPD manifold. According to the operation rules of the matrix, S^d is a vector space while S_{++}^d is a non-Euclidean space. A Riemannian metric should be given to describe the geometric structure of the SPD matrix and to measure the distance of two points on S_{++}^d.

Geodesics on the manifold are related to the tangent vectors in the tangent space. Two operators, namely exponential map $\exp_X(\cdot): T_X M \to M$ and the logarithm map $\log_X(\cdot) = \exp_X^{-1}(\cdot): M \to T_X M$, are defined over differentiable manifolds to switch between the manifold and tangent space at the point *X*. As illustrated in Fig. 2, the tangent vector *v* is mapped to the point *Y* on the manifold through the exponential map. The length of v is equivalent to the geodesic distance between X and Y based on the property of the exponential map. The logarithm map is the inverse of the exponential map and maps a point on the manifold to the tangent space $T_X M$. Conversely, a point on the manifold is mapped to the tangent space $T_X M$ through the logarithm map. The exponential and logarithm maps vary as point X moves along the manifold. The details can be referred in Harandi et al. (2012).

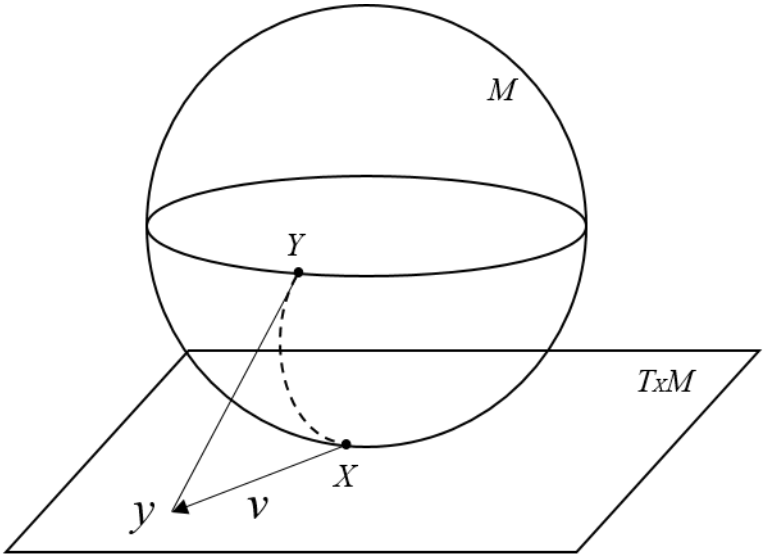

Figure 2: Illustration of the tangent space $T_X M$ at point X on a Riemannian manifold. A SPD matrix can be interpreted as point X in the space of SPD matrices. The tangent vector v can be obtained through the logarithm map, ie. $v = \log_X(Y)$. Every tangent vector in $T_X M$ can be mapped to the manifold through the exponential map, ie. $\exp_X(v) = Y$. The dotted line shows the geodesic starting at X and ending at Y.

For S_{++}^d, the logarithm and exponential maps are defined as:

$$\log_X(Y) = X^{\frac{1}{2}} \log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) X^{\frac{1}{2}}, \tag{1}$$

$$\exp_X(y) = X^{\frac{1}{2}} \exp(X^{-\frac{1}{2}} y X^{-\frac{1}{2}}) X^{\frac{1}{2}}, \tag{2}$$

where $\log(\cdot)$ and $\exp(\cdot)$ are the matrix logarithm and exponential operators, respectively. For SPD matrices, they can be computed through Singular Value Decomposition (SVD). If we let $\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_d)$ be a diagonal matrix formed from real values $\lambda_1, \lambda_2, \ldots, \lambda_d$ on diagonal elements and $X = U \mathrm{diag}(\lambda_i) U^T$ be the SVD of the symmetric matrix X. In Eq. (1) and Eq. (2), $\log(\cdot)$ and $\exp(\cdot)$ are calculated by

$$\log(X) = \sum_{r=1}^{\infty} \frac{(-1)^{r-1}}{r} (X - I)^r = U \mathrm{diag}(\ln(\lambda_i))\, U^T, \tag{3}$$

$$\exp(X) = \sum_{r=0}^{\infty} \frac{1}{r!} X^r = U \mathrm{diag}(\exp(\lambda_i))\, U^T, \tag{4}$$

where I is an identity matrix on manifolds.

The manifold can be embedded into its tangent space at identity matrix I. Thus, based on the bi-invariant Riemannian metric (Arsigny et al., 2008), the distance between two SPD matrices X, Y is $d(X, \mathrm{Y}) = \|\log(X) - \log(Y)\|_2$, where $\log(\cdot)$ is the matrix logarithm operator. Since symmetric matrices (equivalently tangent spaces) form a vector space, the classification tools in the Euclidean space (SVM, KNN and so on) can be seamlessly employed to deal with the recognition problem.

Given an SPD matrix A, its log-Euclidean vector representation $a \in \mathbb{R}^m$, $m = d(d+1)/2$, is unique and defined as $a = Vec(\log(A))$. Let $B = \log(A)$, $B \in S^d$ and

$$B = \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} & \dots & b_{1,d} \\ b_{2,1} & b_{2,2} & b_{2,3} & \dots & b_{2,d} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ b_{d,1} & b_{d,2} & b_{d,3} & \dots & b_{d,d} \end{bmatrix}_{d \times d}, \quad (5)$$

which lies in the Euclidean space. Since *B* is symmetric, we can rearrange it into a vector by vectorizing its upper triangular matrix:

$$a = Vec(B) = \left[b_{1,1}, \sqrt{2}b_{1,2}, \cdots, \sqrt{2}b_{1,d}, b_{2,2}, \sqrt{2}b_{2,3}, \cdots, b_{d,d}\right]^{\mathrm{T}}. \quad (6)$$

Thus, vector a is defined as the manifold features. Since f is a 6-dimensional feature mapping in the experiment, the manifold feature vector a to represent the cloud image is $6 \times (6+1)/2 = 21$ dimensions. The mapped feature vector can reflect the characteristics of its corresponding SPD matrix on matrix manifolds. Thus, manifold features can describe the non-Euclidean property of the infrared image features to some degree.

2) **(Page 8, Line 1-6)** Section 2.2.3 Combining manifold and texture features

As described in Sect. 2.2.1 and 2.2.2, manifold features and texture features can be extracted and integrated to represent the ground-based infrared images. For an image, the four features including energy, entropy, contrast and homogeneity from GLCM, express its texture, while 21-demensional manifold features describe the non-Euclidean geometric characteristics. The manifold and texture features are combined to form a feature vector to represent an image. Thus, the joint features of the infrared image have a total of 25 dimensions.

3) **(Page 8, Line 8-21)** The information of SVM has be added in Section 2.3 Classification.

4) In the experiment, as shown in Table 1 (Corresponding to Table 3 in the manuscript), we first test the effects of different features. It is shown that manifold features perform better than texture features with an increase of at least 10% on the zenithal dataset and about 3.5% on the whole-sky dataset. When these two features are combined, there is an improvement on the 10-fold cross validated classification accuracy compared to using texture or manifold features alone. It's clear on the whole-sky dataset, the accuracy is improved by about 2.7%. Comparing two features alone, manifold features provide more discriminate information than texture features. As a result, manifold features play a bigger role in the combined features in our experiment.

Table 1. The 10-fold cross validated classification accuracy (%) of the proposed method on two datasets.

	Zenithal	Whole-sky
Texture features	83.49	78.01
Manifold features	96.46	82.38
Combined features	**96.50**	**85.12**

5) **(Page 2, Line 21-31)** In this paper, we utilize region covariance matrices, composed from densely sampled features, as the descriptors. This region descriptor has several advantages. Firstly, it calculates the first-order and second-order statistics of the local patch. Secondly, it straightforward fuses various features. Thirdly, it is independent of the region size and has low dimensions. Fourthly, by subtracting

the mean feature vector, the effect of the noisy samples is reduced to some degree. Finally, it is able to speed up the computation in images and videos using efficient methods (Tuzel et al., 2008; Sanin et al., 2013). Covariance matrices are Symmetric Positive Definite (SPD) matrices and naturally form a connected Riemannian manifold. The manifold feature vector can maintain these advantages and describe the images well, so it is chosen for a try on the ground-based cloud classification.

6) In our work, we have tested other features proposed in the literatures. For example, 6-dimensional features including mean, standard deviation, smoothness, third moment, uniformity and entropy in Calbó and Sabburg (2008) and 4-dimensional features including mean, standard deviation, skewness and cloud cover in Heinle et al. (2010) were added with texture features for classification of ground-based cloud types. Table 2 gives the 10-fold cross validation results using these features compared to the proposed method on the zenithal dataset and whole-sky dataset, respectively.

In Table 2, 10-dimensional features are the combination of 4-dimensional texture features, energy, entropy, contrast and homogeneity of GLCM and 6-dimensional features used in Calbó and Sabburg (2008), 8-dimensional features are the combination of these 4-dimensional texture features from GLCM and 4-dimensional features adopted in Heinle et al. (2010) and 25-dimensional features are the proposed method in this manuscript. As shown in Table 2, the results of 25-dimensional features outperform those of the other two combined features. It means that on these two datasets, combination of manifold features and texture features performs better. Other different, more easily physically interpretable parameters may exist and lead to a higher hit rate in the classification. In our manuscript, the addition of manifold features has validated its effectiveness in the ground-based cloud classification on the two datasets.

Table 2. The 10-fold cross validated classification accuracy (%) with different features on two datasets.

	Zenithal	Whole-sky
10-dimenisional features	94.76	82.59
8-dimenisional features	92.91	82.01
25-dimensional features	**96.50**	**85.12**

7) Essentially, cloud classification is not a physically based problem, but a problem of pattern recognition. The classification accuracy is mainly dependent on the features and the classifier. Features are used to represent the image itself, and the classifier acts as a referee to predict its type according to its features. For example, if we can justify all the images based on one feature, there is no need to adopt the others. The best features should represent the most significant properties of the images well. That's not to mean that the more parameters, the better fit.

4. Comment: The preselection of the data used in this study is done by employing two experienced experts, and the images are only chosen if both of them agree on the cloud type. Does that not already mean a very strong constraint on the images with regard to their clarity? Does it affect the study result? How would the algorithm perform under realistic conditions, where images are not preselected? Is there a quality flag involved? Is there a way to further improve the classification?

Response: (Page 3, Line 30–Page 4, Line 10) Many thanks for the constructive comments of the manuscript. It is true that the preselection of the data used in this study is done by employing two

experienced experts, and the images are only chosen if both of them agree on the cloud type. The sentence "The selection criterion is that the chosen images should hold high visual quality and can be recognized by visual inspection" has been changed to "The selection premise is that …". The purpose that the images should hold high visual quality is to make them recognized by visual inspection accurately. If an image is vague, it's hard for experts to justify its type. For the algorithm, it's hard to extract effective features of a vague image, not to mention recognizing its cloud type. In general, the ground-based images are obtained by the specific equipment, like Whole-Sky Infrared Cloud-Measuring System (WSIRCMS), so the quality of the images is similar in most cases. There is no specific quality flag to select images, and it is mainly dependent on the visual judgment. Under realistic conditions, according to the previous data gathered in one place, the SVM models are trained to predict the real-time images there. The results given by automatic cloud classification algorithm are compared with those determined by experts to evaluate the performance of the algorithm. To further improve the classification accuracy, we should gather as many representative samples in per class as possible. When the training images are typical and sufficient enough, there is no doubt that the recognition rate will be improved.

5. Comment: Please also give some insight how to assess your improvement of the classification. Depending on the case and fraction it is somewhere between 2 and 10% it seems. Is that a great improvement? Does it depend on the choice of cases?

Response: (Page 9, Line 29-33; Page 10, Line 17-20) Many thanks for the careful reading of our manuscript. In the manuscript, we give three cases to make comparison with different methods. First, when 1/10 of the dataset is treated as the training set, there is an increase of about 9.2% for the zenithal dataset and about 4.7% for the whole-sky dataset in the overall accuracy. This is a case that the training samples are not sufficient. Second, we conduct the experiment when 1/2 of the dataset is used for training. There is an increase of about 3.7% for the zenithal dataset and about 3% for the whole-sky dataset in the overall accuracy. This is a case that the training samples are enough. Third, we conduct the experiment when 9/10 of the dataset is used for training. There is an increase of about 2.9% for the zenithal dataset and about 3.7% for the whole-sky dataset in the overall accuracy. This is a case that the training samples are quite enough. With different fractions that the training set occupies, it is validated that in most cases the proposed method outperforms the other two methods (Liu et al., 2015; Cheng and Yu, 2015). As a whole, the improvement of the proposed method is between 2% and 10%. To some degree, it may not be a great improvement, but we have validated that the introduction of manifold features is effective and can achieve some success, it is worthy doing more work in this field to promote its development. In general, with the increase in the number of training samples, the overall accuracy will increase until it holds stable.

6. Comment: Abstract: Overall, it already assumes a great background knowledge of the reader.

Response: Many thanks for the constructive comments of our manuscript. We have added some introduction of SVM and other explanations in the revised manuscript.

7. Comment: P1, l10: "the" Support Vector Machine → I think it is an overarching concept. "a" Support Vector Mache.

Response: (Page 1, Line 10) Many thanks for the thoughtful commenting. We agree and have changed in the revised manuscript.

8. Comment: P1, l13: Specify some numbers here (higher by how much?)

Response: (Page 1, Line 14) Many thanks for the thoughtful comment. The proposed method performs higher by 2%-10% than the other two methods in the recognition rate.

9. Comment: P1, l16 and following: Somewhere you should mention and cite CloudNet (http://www.cloud-net.org) which are quite capable of identifying cloud types. Or do you only focus on large scale cloud structures? (The you should clarify that, because cloud classification implies that you look at the cloud type also.).

Response: (Page 1, Line 21) Many thanks for the constructive comments of the manuscript. The categorization products at the CloudNet Web site (http://www.cloud-net.org) form a dataset, aggregated from cloud radar, lidar, a numerical forecast model and optionally a rain gauge and microwave radiometer. They are used to improve the numerical prediction models. The observational datasets and model datasets are used in this procedure. Different from it, we focus on identifying cloud types based on the images, so it is a problem to extract features due to the formats of the input data. Compared to the traditional manual observation, ground-based infrared cloud images can be obtained continuously and have a high spatial resolution at a local scale. Also, the title "Cloud Classification of **ground-based infrared images** combining manifold and texture features" has clarified our purpose. We have added Illingworth et al. (2007) as a reference in the revised manuscript.

10. Comment: P1, l23,24: "weaken their credibility", please check the use of the word "credibility".

Response: (Page 1, Line 24) Many thanks for the careful reading. We have checked the use of the word "credibility", and the original sentence (the last sentence in the first paragraph of Section Introduction) in Tzoumanikas et al. (2012) is "Additionally, the introduction of human factor and a rough measuring system (in octas or tenths of the sky dome) in the estimation of cloud cover and type weaken their credibility.", so we use this word in the manuscript.

11. Comment: P1 l23 – P2, l13: This seems like an itemization of the existing methods. In which way do they connect to your method? Why do you later on choose just two of them (Liu, Cheng) to compare to?

Response: Many thanks for the constructive comments. As mentioned in the manuscript, there are some methods for ground-based cloud classification. Usually, the main steps contain feature extraction and cloud type recognition with the classifier. In the manuscript, we stated what features and classifiers were adopted in the literatures. As a result, we should hand these two sub-problems well to realize our goal. In these existing methods, most features including statistical, texture, and structure features are extracted in the Euclidean space. Different from these methods, besides the texture features, manifold features are extracted on Riemannian manifolds to describe non-Euclidean geometric characteristics in our method.

Most methods are conducted on ground-based colour images, and the extracted features need different channels' information. Infrared images only have one channel, so it's hard to realize these

methods based on colour features. Liu's method uses Weighted Local Binary Pattern (WLBP), which just based on the grayscale images (Liu et al., 2015). Cheng's method uses statistical features and distribution of local texture features (Cheng and Yu, 2015). Differences of R-G, R-B and G-B components are given up to make Cheng's method apply to the infrared images. Then these two methods are adopted for comparison.

12. Comment: P2, l5 "parallelepiped" → typo

Response: (Page 2, Line 6) Many thanks for the careful reading. We have checked the spelling of the word "parallelepiped", and the original sentence (the first sentence in the fourth paragraph of Section Results and discussion) in Calbó and Sabburg (2008) is "The classifier that we developed is based on the supervised **parallelepiped** technique, which has been used elsewhere for similar applications", so we adopt this terminology in our manuscript.

13. Comment: P2, l10: Support Vector Machine needs a citation, it is not generally known.

Response: (Page 2, Line 11; Page 8, Line 9) Many thanks for the careful reading of our manuscript. We have added Cristianini and Shawe-Taylor (2000) as a reference to make it clear.

14. Comment: P.2, l14 – l20: You state that colour images provide more information. Make clearer why infrared images are used anyway.

Response: (Page 2, Line 16-23) Many thanks for the careful reading of our manuscript. It is known that the colour image has 3 channels of R, G and B while the infrared image only has one channel. As a result, from the aspect of the image itself, the RGB images can provide colour information, which is important for analyzing the cloud to some degree. Although colour images have this advantage, it is hard to gather them during the night. Compared to colour images, infrared images can be obtained day and night continuously, which is necessary for practical application. As a result, we investigate the ground-based infrared images.

15. Comment: P.2, l21 – l24: Make it clear why manifolds are chosen. There are many mathematical constructs, it is not obvious why this way is chosen. Very clearly state here what the novelty and potential of your method is.

Response: (Page 2, Line 24- Page 3, Line 2) We are sorry for our inadequate explanation. Nowadays, the Symmetric Positive Definite (SPD) matrix manifold has achieved success in many aspects, such as action recognition, material classification and image segmentation (Faraki et al., 2015; Jayasumana et al., 2015). It is because Euclidean space cannot gain special structures of image features well while Riemannian manifolds form non-Euclidean spaces and are more appropriate to address this problem. Although it proves effective, few researches are pursued for the task of cloud classification with manifold features.

In this paper, we utilize region covariance matrices, composed from densely sampled features, as the descriptors. A region descriptor has several advantages. Firstly, it calculates the first-order and second-order statistics of the local patch. Secondly, it straightforward fuses various features. Thirdly, it

is independent of the region size and has low dimensions. Fourthly, by subtracting the mean feature vector, the effect of the noisy samples is reduced to some degree. Finally, it is able to speed up the computation in images and videos using efficient methods (Tuzel et al., 2008; Sanin et al., 2013). Covariance matrices are SPD matrices and naturally form a connected Riemannian manifold. Covariance matrices are Symmetric Positive Definite (SPD) matrices and naturally form a connected Riemannian manifold. Although it proves effective, few researches are pursued for the task of cloud classification with manifold features. The manifold feature vector can maintain these advantages and describe the images well, so it is chosen for a try on the cloud classification. In this paper, a novel cloud classification method is proposed for ground-based infrared images. Manifold features, representing the non-Euclidean geometric structure of the image features, and texture features, expressing the image texture, are integrated for the feature extraction.

16. Comment: P.2, l 31: "...displayed the best performance in the 10-fold cross validation overall", it is not clear what 10-fold cross validation you mean.

Response: (Page 3, Line 10) We are sorry for our inadequate explanation. Cross validation is a usual way to test the performance of the algorithm (Ripley, 2005), which has been applied by e.g. Heinle et al. (2010), Li et al. (2016) and Gan et al. (2017). In the 10-fold cross validation, the dataset is divided into 10 subsets with the same size at random, then one single subset is used for validation in turn and the other 9 parts are taken as the training set. This process is then repeated 10 times. The average value is used to estimate the capability of the algorithm. In general, 10-fold cross validation is repeated many times to measure the algorithm's performance. We have added Ripley (2005) as a reference and clarified it later (**Page 9, Line 3-5**) in the tracking-changes manuscript.

17. Comment: P.3, l6: "...ground-based passive system that an uncooled microbolometer … is used." → "...ground-based passive system that uses an uncooled microbolometer …"

Response: (Page 3, Line 18) Many thanks for the careful reading of our manuscript. We have made this change in the revised manuscript.

18. Comment: P.3, l 9: Are the pixels size or resolution?

Response: (Page 3, Line 21) Sorry to make it confusing. Here are the pixels' size, not the resolution.

19. Comment: P.3, l20: It is not clear why a historical dataset would not contain a complex mixture of cloud types compared to a dataset of the present.

Response: (Page 4, Line 8-10) We are sorry for our inadequate explanation. The dataset is used to assess the performance of the algorithm. The sky condition is classified into five types: stratiform clouds, cumuliform clouds, waveform clouds, cirriform clouds and clear sky. To guarantee the reliability of true label of each image, we should select the images without mixed cloud types. Indeed, the sky condition with mixed cloud types always exists, but it is complicated. We will investigate the images containing mixed cloud types next.

20. Comment: P.3, l22: comprised of 100 images in each category

Response: (Page 4, Line 11) Many thanks for the careful reading of our manuscript. We have made this change in the revised manuscript.

21. Comment: P.3, l26: the number of cases with stratiform clouds, cumuliform clouds, …

Response: (Page 4, Line 15) We are sorry for our carelessness. We have made this change in the revised manuscript.

22. Comment: P.3, l29: "which is the area of clouds rather than the parts out of the circle" → what do you mean by "parts out of the circle"?

Response: (Page 4, Line 17-19; Page 23) We are sorry to make it confusing. In Fig. 3, it can be seen that the region of interest is the white region while the black region is out of the circle.

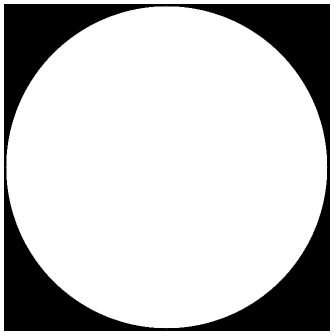

Figure 3: The mask of the whole-sky images.

23. Comment: P.4, l5: Only later clear what "non-Euclidean features" are.

Response: (Page 7, Line25-27) Many thanks for the constructive comment. In the manuscript, the manifold features describe the non-Euclidean geometric characteristics of infrared image features. As each image corresponds to a SPD matrix, to better maintain its manifold geometric structure, it is mapped into its tangent space by the logarithm operation. The mapped feature vector can reflect the characteristics of its corresponding SPD matrix on matrix manifolds. Thus, manifold features can describe the non-Euclidean property of the infrared image features to some degree.

24. Comment: P.4, l22: "...mean values in four directions are obtained as texture feature", mean over what? And what directions?

Response: (Page 5, Line 10-14) We are sorry not to make it clear. The Grey Level Co-occurrence Matrix (GLCM) is calculated in a defined direction θ and a pixel distance d. In the experiment, we get four GLCMs with d=1 and θ=0°, 45°, 90°, 135°. Then four measures energy, entropy, contrast and homogeneity are computed according to Eq. (7) ~ Eq. (10). Since there are 4 GLCMs, the texture features are 16 dimensional. To alleviate the complexity, reduce the dimension and keep rotation invariance, four mean features of four GLCMs with d=1 and θ=0°, 45°, 90°, 135°are obtained as the final texture features.

$$\text{Energy} = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} p(i,j)^2. \quad (7)$$

$$\text{Entropy} = -\sum_{i=0}^{k-1}\sum_{j=0}^{k-1} p(i,j)\log_2 p(i,j). \quad (8)$$

$$\text{Contrast} = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1}(i-j)^2 p(i,j)^2. \quad (9)$$

$$\text{Homogeneity} = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\frac{p(i,j)}{1+|i-j|}. \quad (10)$$

25. Comment: P.5, l10: This is supposedly a d times d matrix. Should the d not show up in this equation as an index or something?

Response: (Page 5, Line 23) Many thanks for the constructive comments. We have checked this equation and there is no mistake for this equation. It is in the form of matrix operation. For a feature image *F*, it contains n=W×H points of d-dimensional features $\{f_k, k=1, 2, \ldots, n\}$. Its Covariance Descriptor (CovD) is a d×d covariance matrix, computed by

$$C = \frac{1}{n-1}\sum_{k=1}^{n}(f_k-\mu)(f_k-\mu)^{\mathrm{T}}, \quad (11)$$

where $\mu = \frac{1}{n}\sum_{k=1}^{n} f_k$, which represents the feature mean vector. In this equation, f_k is a d-dimensional feature vector, and as the feature mean vector, μ is also d-dimensional, so $(f_k-\mu)(f_k-\mu)^{\mathrm{T}}$ is a d×d matrix, and the CovD *C* is a d×d matrix. The (*i*, *j*)-th element of CovD in Eq. (11), $C(i,j)$, can also written as:

$$C(i,j) = \frac{1}{n-1}\sum_{k=1}^{n}\big(f_k(i)-\mu(i)\big)\big(f_k(j)-\mu(j)\big), \quad i,j = 1,2,\ldots,d, \quad (12)$$

where $f_k(i)$, $\mu(i)$ mean the *i*-th elements of the feature vectors f_k and μ, respectively.

26. Comment: P.6, l3: This is not an equation, it lacks a left side. P.6, Section 2.3: Needs some more explanation or at the least citations. SVM not understandable from this.

Response: (Page 7, Line 13-14) Many thanks for the careful reading. This equation has been modified as follows:

$$A^* = \arg\min_A \|C-A\|_F, \text{s.t.}\, A + A^T > 0. \quad (13)$$

27. Comment: P.6, l.21: What do you mean by "voting policy"?

Response: (Page 8, Line 22-27) Many thanks for the thoughtful comments. In a multi-class task, the SVM is conducted between every two classes. In this paper, there are 5 types, so there are 5×(5-1)/2=10 SVM classifiers between every two classes. For an unknown-type image, it will be input into 10 models and get 10 output labels. That is to say, each binary classifier makes its vote to predict the sample's class. According to the voting policy, the most frequent label is this sample's type.

28. Comment: P.6, Section 3, beginning: Here, it should be clarified at the very latest what you mean by "experiment" in your context.

Response: (Page 8, Line 29- Page 9, Line 5) Many thanks for the constructive comment. Different from a concrete case, the experiment we carry out is random. The purpose that we conduct the experiment is to test the performance of the algorithm. In the experiment, the training set is selected at random and the rest of the dataset forms the testing set, so we need to repeat this process many times and calculate the mean value in order to avoid the accidental bias and to measure the algorithm well. In the revised manuscript, we have checked the "experiment" and "cross validation" to make it clear.

29. Comment: P.7, l21: Do you really mean "confusion matrix"?

Response: (Page 10, Line 1-6) Many thanks for the careful comment. The confusion matrix is a way to exhibit the experimental result, which has been applied by e.g. Zhuo et al. (2014), Liu et al. (2015), and Li et al. (2016). In the confusion matrix, each row of the matrix represents an actual class while each column represents the predicted class given by the algorithm. For example, the element in the second row and third column is the percentage of cumuliform clouds misclassified as waveform clouds. Therefore, the recognition rate for each class is in the diagonal of the matrix.

30. Comment: P.7, l24: cululiform → cumuliform

Response: (Page 10, Line 4) We are sorry for our carelessness. In the revised paper, we have corrected the spelling problems.

31. Comment: P.7, l27: "has reached" → "is reached"

Response: (Page 10, Line 8) We are sorry for our carelessness. In the revised manuscript, we have corrected the grammar problems.

32. Comment: P.7, l30: exits → exists

Response: (Page 10, Line 11) We are sorry for our carelessness. In the revised manuscript, we have corrected the spelling problems.

33. Comment: P.8, l4: "when 1/2 for training." → "when 1/2 for training is used."

Response: (Page 10, Line 21) We are sorry for our carelessness. We have corrected this error in the revised manuscript.

34. Comment: P.8, l25: There is indeed improvement, but I would not call it "dramatically".

Response: (Page 11, Line 10; Page 11, Line 17-21) Many thanks for the constructive comment. We have deleted the word "dramatically" In the revised manuscript.

35. Comment: P.8, l26: What do you mean by "the statistical learning method"? I think this hasn't been defined before.

Response: (Page 11, Line 10-11) We are sorry not to make it clear. Statistical learning method is a framework for machine learning drawing from the fields of statistics and functional analysis (Trevor et al., 2009; Mohri et al., 2012). Statistical learning theory deals with the problem of finding a predictive function based on data. Because there is no definition before in the manuscript, we have changed this sentence as: "… on the other hand, the manifold features on the matrix manifold can describe the non-Euclidean geometric structure of the image features and thus …".

36. Comment: P.8, l32: Gabor or wavelet coefficients may need a citation, not generally known.

Response: (Page 11, Line 22) Many thanks for the constructive comment. We have added a reference Liu and Wechsler (2002) to make it clear in the revised manuscript.

37. Comment: P.9, l3: "on the both" → "on both,"

Response: (Page 11, Line 27) We are sorry for our carelessness. In the revised manuscript, we have corrected this problem.

38. Comment: Images and Tables: Please provide more telling captions. Table 4 and 5: The 1/10, 1/2 and so on are not clear.

Response: (Page 18; Page 19) Many thanks for the careful reading of our manuscript. We are sorry for our negligence. We have added some explanation of the fractions 1/10, 1 /2, 9/10 in Tables 4 and 5 in the revised manuscript to make it clear: 1/10, 1/2 and 9/10 are the certain proportions of the training set selected randomly from each category, and the rest part forms the testing set correspondingly.

Special thanks for the reviewers' constructive comments.

We have tried our best to improve the manuscript and made some changes in the manuscript. These changes will not influence the content and framework of the paper. And here we did not list the changes but marked in the revised paper.

We appreciate for the reviewers' warm work earnestly, and hope that the correction will meet with approval.

Once again, thank you very much for your careful reading and inspiring suggestions.

References

Arsigny, V., Fillard, P., Pennec, X., and Ayache, N.: Geometric means in a novel vector space structure on symmetric positive-definite matrices, Siam J. Matrix Anal. A., 29, 328-347, doi:10.1137/050637996, 2008.

Calbó, J., and Sabburg, J.: Feature extraction from whole-sky ground-based images for cloud-type recognition, J. Atmos. Ocean. Techn., 25, 3, doi:10.1175/2007JTECHA959.1, 2008.

Cheng, H. Y., and Yu, C. C.: Block-based cloud classification with statistical features and distribution of local texture features, Atmos. Meas. Tech., 7, 1173-1182, doi:10.5194/amt-8-1173-2015, 2015.

Cristianini, N. and Shawe-Taylor, J.: An introduction to support vector machines and other kernel-based learning methods, Cambridge university press, 2000.

Gan, J., Lu, W., Li, Q., Zhang, Z., Yang, J., Ma, Y. and Yao, W.: Cloud type classification of total-sky images using duplex norm-bounded sparse coding, IEEE J. Sel. Top. Appl., 10, 3360-3372, doi:10.1109/JSTARS.2017.2669206, 2017.

Harandi, M. T., Sanderson, C., Wiliem, A., and Lovell, B. C.: Kernel analysis over Riemannian manifolds for visual recognition of actions, pedestrians and textures, In Proceedings of 2012 IEEE Workshop on the Applications of Computer Vision, Breckenridge, CO, USA, 433-439, 2012.

Heinle, A., Macke, A., and Srivastav, A.: Automatic cloud classification of whole sky images, Atmos. Meas. Tech., 3, 557-567, doi:10.5194/amt-3-557-2010, 2010.

Illingworth, A. J., Hogan, R. J., O'connor, E. J., et al.: Cloudnet: Continuous evaluation of cloud profiles in seven operational models using ground-based observations, B. Am. Meteorol. Soc., 88, 883-898, 2007.

Li, Q., Zhang, Z., Lu, W., Yang, J., Ma, Y. and Yao, W.: From pixels to patches: a cloud classification method based on a bag of micro-structures, Atmos. Meas. Tech., 9, 753-764, doi:10.5194/amt-9-753-2016, 2016.

Liu, C. and Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition, IEEE T. Image Process., 11, 467-476, doi:10.1109/tip.2002.999679, 2002.

Liu, L., Sun, X., Gao, T., and Zhao, S.: Comparison of cloud properties from ground-based infrared cloud measurement and visual observations, J. Atmos. Ocean. Tech., 30, 1171-1179, doi:10.1175/JTECH-D-12-00157.1, 2013.

Liu, S., Zhang, Z., and Mei, X.: Ground-based cloud classification using weighted local binary patterns, J. Appl. Remote Sens., 9, 095062, doi:10.1117/1.JRS.9.095062, 2015.

Mohri M., Rostamizadeh, A., and Talwalkar, A.: Foundations of machine learning, MIT press, Massachusetts, USA, 2012.

Ripley, B. D.: Pattern recognition and neural networks, Cambridge University Press, 8 edn., 2005.

Sanin, A., Sanderson, C., Harandi, M. T., and Lovell, B. C.: Spatio-temporal covariance descriptors for action and gesture recognition, In Proceedings of 2013 IEEE Workshop on Applications of Computer Vision (WACV), Tampa, FL, USA, 103-110, 2013.

Trevor, H., Robert, T., and Jerome, F.: The elements of statistical learning, Springer New York, 2009.

Tuzel, O., Porikli, F., and Meer, P.: Pedestrian detection via classification on Riemannian manifolds, IEEE T. Pattern Anal., 30,1713-1727, 2008.

Tzoumanikas, P., Kazantzidis, A., Bais, A. F., Fotopoulos, S., and Economou, G.: Cloud detection and classification with the use of whole-sky ground-based images, Atmos. Res., 113, 80-88, doi:10.1016/j.atmosres.2012.05.005, 2012.

Zhuo, W., Cao, Z., and Xiao, Y.: Cloud classification of ground-based images using texture-structure features, J. Atmos. Ocean. Techn., 31, 79-92, doi:10.1175/JTECH-D-13-00048.1, 2014.

Cloud classification of ground-based infrared images combining manifold and texture features

Qixiang Luo[1], Yong Meng[1], Lei Liu[1], Xiaofeng Zhao[1], and Zeming Zhou[1]

[1]College of Meteorology and Oceanology, National University of Defense Technology, Nanjing, 211101, China

Correspondence to: Zeming Zhou (zhou_zeming@yahoo.com)

Abstract. Automatic cloud type recognition of ground-based infrared images is still a challenging task. A novel cloud classification method is proposed to group images into five cloud types based on manifold and texture features. Compared with statistical features in the Euclidean space, manifold features extracted on Symmetric Positive Definite (SPD) matrix space can describe the non-Euclidean geometric characteristics of the infrared image. The proposed method comprises three stages: pre-processing, feature extraction and classification. Cloud classification is performed by ~~the~~ a Support Vector Machine (SVM). The datasets are comprised of the zenithal and whole-sky images taken by the Whole-Sky Infrared Cloud-Measuring System (WSIRCMS). Benefiting from the joint features, compared to the recent cloud type recognition methods, the experimental results illustrate that the proposed method acquires a higher recognition rate ~~and exhibits a more competitive classification result~~with an increase of 2%-10% on the ground-based infrared datasets.

1. Introduction

The cloud has an essential impact on the absorption, scattering, emission of atmosphere, the vertical transport of heat, moisture and momentum (Hartmann et al., 1992; Chen et al., 2000). Cloud cover and cloud type can affect the daily weather and climate change through its radiation and hydrological effects (Isaac and Stuart, 1996; Liu et al., 2008; Naud et al., 2016). Therefore, accurate cloud detection and classification is necessary for meteorological observation. Nowadays, cloud cover changes and cloud type determination have been available through the ground-based sky imaging systems (Souzaecher et al., 2006; Shields et al., 2003; Illingworth et al., 2007). Different from traditional manual observation, ground-based sky-imaging devices can obtain continuous information of sky condition at a local scale with a high spatial resolution.

However, due to subject factors and a rough ground-based measuring system, the estimation of cloud cover and type may weaken their credibility (Tzoumanikas et al., 2012). Some attempts have been made to develop algorithms for cloud classification of ground-based images (Buch and Sun, 1995; Singh and Glennen, 2005; Cazorla et al., 2008; Heinle et al., 2010; Ghonima et al., 2012; Taravat et al., 2014; Zhuo et al., 2014). Wang and Sassen (2001) developed a cloud detection algorithm by combining ground-based active and passive remote sensing data to illustrate how extended-time remote sensing datasets can be converted to cloud properties of concern to climate research. Li et al. (2002) proposed a method for automatic classification of surface and cloud type using Moderate Resolution Imaging Spectro-radiometer (MODIS) radiance measurements, whose advantage lied in its independence of radiance or brightness temperature threshold criteria,

and its interpretation of each class was based on the radiative spectral characteristics of different classes. Singh and Glennen (2005) adopted the k-nearest neighbour (KNN) and neural network classifiers to identify cloud types with texture features, including autocorrelation, co-occurrence matrices, edge frequency, Law's features and primitive length. Calbó and Sabburg (2008) extracted statistical texture features based on the greyscale images, pattern features based on the spectral power function of images and other features based on the thresholded images for recognizing the cloud type with the supervised parallelepiped classifier. Heinle et al. (2010) chose 12 dimensional features mainly describing the colour and the texture of images for automatic cloud classification, based on the KNN classifier. Besides the statistical feature like the mean grey value of the infrared image, Liu et al. (2011) explored another six structure features to characterize the cloud structure for classification. Zhuo et al. (2014) validated that cloud classification may not perform well if the texture or structure features were employed alone. As a result, texture and structure features were captured from the colour image and then fed into a trained Support Vector Machine (SVM) (Cristlanini and Shawe-Taylor, 2000) to obtain the cloud type. Different from traditional feature extraction, Shi et al. (2017) proposed to adopt the deep convolutional activations-based features and provided a promising cloud type recognition result with a multi-label linear SVM model.

Automatic cloud classification has made certain achievements; however, the cloud classification of ground-based infrared images poses a great challenge to us. By far, few researches ~~works~~ of cloud classification have been dedicated to the ground-based infrared images (Sun et al., 2009; Liu et al., 2011). Most recent methods conducted on the RGB visible images (Heinle et al., 2010; Zhuo et al., 2014; Li et al., 2016; Gan et al., 2017) cannot directly be exploited on the cloud type classification of infrared images ~~owing~~due to the lack of colour information. Compared to colour images, ~~Ground-based~~ infrared images can be obtained ~~continuously~~ day and night continuously, which is important for practical application and analysis. ~~but lack colour information, so it's hard to reach a perfect performance when the recognition method of colour images is applied to the infrared images. Most recent methods conducted on the RGB visible images (Heinle et al., 2010; Zhuo et al., 2014; Li et al., 2016; Gan et al., 2017) cannot directly be exploited on the cloud type classification of infrared images owing to the lack of colour information.~~

Nowadays, the Symmetric Positive Definite (SPD) matrix manifold has achieved success in many aspects, such as action recognition, material classification and image segmentation (Faraki et al., 2015; Jayasumana et al., 2015). As a representative of SPD matrix, the Covariance Descriptor (CovD) is a powerful tool to extract the feature of the image. It owns several advantages. Firstly, it calculates the first-order and second-order statistics of the local patch. Secondly, it straightforward fuses various features. Thirdly, it is independent of the region size and has low dimensions. Fourthly, by subtracting the mean feature vector, the effect of the noisy samples is reduced to some degree. Finally, it is able to speed up the computation in images and videos using efficient methods (Tuzel et al., 2008; Sanin et al., 2013). Covariance matrices naturally form a connected Riemannian manifold. Although it proves effective, few investigations ~~researches~~ are pursued for the task of cloud classification with manifold features. The manifold feature vector can maintain these advantages of non-Euclidean geometric space and describe the image features comprehensively, so it is chosen for a try on the cloud classification. In this paper, a novel cloud classification method ~~combining manifold and texture features~~ is proposed for

ground-based infrared images. Manifold features, representing the non-Euclidean geometric structure of the image features, and texture features, expressing the image texture, are integrated for the feature extraction.

To exhibit the classification performance, we have compared the results with the other two models (Liu et al., 2015; Cheng and Yu, 2015), which are adapted for the classification task of infrared images. To make up for the weakness of the Local Binary Patterns (LBP) that cannot describe the local contrast well, Liu et al. (2015) proposed a new descriptor called Weighted Local Binary Patterns (WLBP) for the feature extraction. And then the KNN classifier based on the chi-square distance was employed for cloud type recognition. Cheng and Yu (2015) incorporated statistical features and local texture features for block-based cloud classification. As Cheng and Yu (2015) reported, the method combining the statistical and uniform LBP features with the Bayesian classifier (Bensmail and Celeux, 1996) displayed the best performance in the 10-fold cross validation (Ripley, 2005) overall.

In this paper, the data and methodology of the method are described in Sect. 2. Section 3 focuses on the experimental results. Conclusions are summarized in Sect. 4.

2. Data and Methodology

In this section, the datasets and the methodology for cloud classification are introduced. The proposed method contains three main steps: pre-processing, feature extraction and classification. The framework is illustrated in Fig. 1.

2.1 Dataset and pre-processing

The datasets include the zenithal images and whole-sky images, which are gathered by the Whole-Sky Infrared Cloud Measuring System (WSIRCMS) (Liu et al., 2013). The WSIRCMS is a ground-based passive system that uses an uncooled microbolometer detector array of 320×240 pixels ~~is used~~ to measure downwelling atmospheric radiance in 8-14μm (Liu et al., 2011). A whole-sky image is obtained after combining the zenithal image and other images at eight different orientations. As a result, the zenithal image has a ~~resolution~~ size of 320×240 pixels while the whole-sky image is of 650×650 pixels. The datasets are provided by National University of Defense Technology in Nanjing, China.

It is true that the clear sky background radiance in 8-14μm varies with time and zenith angle. The images of the datasets have been pre-processed in the consideration of this important factor. The clear sky radiance threshold in each image is calculated using the radiation transfer model (Liu et al., 2013). The real radiance R at each pixel in each image is converted to the grey value G_{pixel} between [0,255] with $G_{pixel} = R/(R_{temp} - R_{clear}) \times 255$, where R_{clear} is the corresponding clear sky radiance threshold and R_{temp} is the radiance corresponding to the real-time environment temperature. As a result, the effects of the clear sky background brightness temperature can be ignored, which means that this factor has little influence on the feature extraction of the images.

The cloud images used in the experiment are selected with the help of two professional meteorological observers with many years of observation experiences. The selection premise ~~criterion~~ is that the chosen images should hold high visual quality and can be recognized by visual inspection. If an image is vague, it's hard for experts to justify its type. For the

algorithm, it's difficult to extract effective features of a vague image, not to mention recognizing its cloud type. All infrared cloud images are labelled to construct the training set and testing set. To guarantee the golden-standard's confidence, only images labelled same by two meteorological observers are finally chosen as the dataset used in this study. Different from traditional cloud classification by observers, automatic cloud classification by the devices needs a new criterion for recognition. According to the morphology and generating mechanism of the cloud, the sky condition is classified into five categories in this study (Sun et al., 2009): stratiform clouds, cumuliform clouds, waveform clouds, cirriform clouds and clear sky. The sky condition and its corresponding description are as shown in Table 1.

The zenithal dataset used in this study is selected from historical dataset to assess the performance of the algorithm. To guarantee the reliability of true label of each image, the images without mixed cloud types are selected. ~~purposely to avoid a complex mixture of cloud types.~~ The typical samples from each category are demonstrated in Fig. 2. As listed in Table 2, the zenithal dataset is comprised of 100 cloud images in each category.

The whole-sky dataset is obtained during July to October in 2014 at Changsha, China. Since the whole-sky image is obtained by combining the nine sub-images at different orientations, the division rules of the whole-sky dataset remain the same as that of the zenithal dataset. The whole-sky samples from each category are exhibited in Fig. 3. As listed in Table 2, the number of cases with stratiform clouds, cumuliform clouds, waveform clouds, cirriform clouds and clear sky is 246, 240, 239, 46 and 88, respectively.

As Fig. ~~3~~ 4 shows, a pre-processing mask is provided on the whole-sky images, which is used to extract the region of interest (ROI) from the images, which is the areas within the circle~~of the clouds~~ rather than the parts out of the circle. Different from the whole-sky images, all parts of the zenithal images are ROI. Thus, we implement the feature extraction directly on the original zenithal images.

2.2 Feature extraction

In addition to the manifold features proposed in this work, the texture features are also combined. The manifold features of the ground-based infrared image are extracted on the SPD matrix manifolds, and after that, they are mapped into the tangent space to form a feature vector in Euclidean space. The texture features represent the statistical information in Euclidean space; on the contrary, the manifold features describe the non-Euclidean geometric characteristics of the infrared image.

2.2.1 Texture features

In this paper, the Grey Level Co-occurrence Matrix (GLCM) is used to extract the texture features, including energy, entropy, contrast and homogeneity (Haralick et al., 1973). Each matrix element in the GLCM represents the joint probability occurrence $p(i,j)$ of pixel pairs with a defined direction θ and a pixel distance d having grey level values i and j in the image.

$$\mathrm{GLCM} = \begin{bmatrix} p(0,0) & p(0,1) & p(0,2) & \dots & p(0,k-1) \\ p(1,0) & p(1,1) & p(1,2) & \dots & p(1,k-1) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ p(k-1,0) & p(k-1,1) & p(k-1,2) & \dots & p(k-1,k-1) \end{bmatrix}_{k\times k} \quad (1)$$

The energy measures the uniformity and texture roughness of the grey level distribution:

$$\mathrm{Energy} = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} p(i,j)^2 \quad (2)$$

The entropy is a measure of randomness of grey level distribution:

$$\mathrm{Entropy} = -\sum_{i=0}^{k-1}\sum_{j=0}^{k-1} p(i,j)\log_2 p(i,j) \quad (3)$$

The contrast is a measure of local variation of grey level distribution:

$$\mathrm{Contrast} = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1}(i-j)^2 p(i,j)^2 \quad (4)$$

The homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal:

$$\mathrm{Homogeneity} = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\frac{p(i,j)}{1+|i-j|} \quad (5)$$

As the number of intensity levels k increases, the computation of the GLCM increases strongly. In this work, k is set with 16 and then the texture features are obtained by calculating four ~~the~~ GLCMs with $\mathrm{d} = 1$ and $\theta = 0°, 45°, 90°, 135°$, respectively. To alleviate ~~avoid~~ the complexity ~~and~~, reduce the dimension and keep rotation invariance, four mean features of four GLCMs with $\theta = 0°,\ 45°,\ 90°,\ 135°$ ~~mean values in four directions~~ are obtained as the final texture features. In the experiments, we find that these texture features are significant for the cloud classification of the ground-based infrared image.

2.2.2 Manifold features

The manifold features are attained by two steps: computing the regional CovD and mapping the CovD into its tangent space to form a feature vector.

Step 1: Computing the regional CovD

Suppose the image I is of the size $W \times H$, its d-dimensional features containing greyscale and gradient at each pixel are computed, which compose the feature image F, whose size is $W \times H \times d$:

$$F(x,y) = f(I,x,y) \quad (6)$$

where the feature mapping f is defined as:

$$f = \left[I(x,y)\ \ |I_x|\ \ |I_y|\ \ \sqrt{|I_x|^2 + |I_y|^2}\ \ |I_{xx}|\ \ |I_{yy}|\right]^{\mathrm{T}} \quad (7)$$

In which (x,y) denotes the location, $I(x,y)$ denotes the greyscale. $|I_x|$, $|I_y|$, $|I_{xx}|$ and $|I_{yy}|$ represent the first and second order derivative in the direction of x and y at each pixel, respectively. $\sqrt{|I_x|^2 + |I_y|^2}$ denotes the modulus of gradient.

For the feature image F, supposing it contains $n = W \times H$ points of d-dimensional features $\{f_k, k = 1,2,\dots,n\}$. Its CovD is a $d \times d$ covariance matrix, computed by Eq. (8):

$$C = \frac{1}{n-1}\sum_{k=1}^{n}(f_k - \mu)(f_k - \mu)^{\mathrm{T}} \quad (8)$$

where $\mu = \frac{1}{n}\sum_{k=1}^{n} f_k$, which represents the feature mean vector.

The CovD can fuse multiple dimensional features of the image and express the correlations between different features. Besides, since the CovD is symmetric, it is only $d(d+1)/2$ dimensional. ~~It removes the mean of the sample features; therefore, it has certain effects of denoising. The CovD is symmetric and its dimension is only $d(d+1)/2$.~~ If we convert the CovD into a feature vector to describe the image, its dimension is $n \times d$, which needs a high computation cost for cloud classification.

Step 2: Obtaining the feature vector by mapping the CovD into its tangent space

Generally speaking, the manifold is a topological space that is locally equivalent to a Euclidean space. The differential manifold has a globally defined differential structure. Its tangent space $T_X M$ is a space formed by all possible tangent vectors at a given point X on the differential manifold. For the Riemannian manifold M, an inner product is defined in its tangent space. The shortest curve between two points on the manifold is called the geodesic and the length of the geodesic is the distance between two points.

All SPD matrices form a Riemannian manifold. Suppose S^d is a set of all $n \times n$ real symmetric matrices: $S^d = \{A \in M(d): A^{\mathrm{T}} = A\}$, where $M(d)$ represents the set of all $d \times d$ matrices, so that $S_{++}^d = \{A \in S^d: A > 0\}$ is the set of all $d \times d$ SPD matrices, which construct a $d(d+1)/2$ dimensional SPD manifold. According to the operation rules of the matrix, the set of the real symmetric matrix is a vector space while the real SPD matrix space is a non-Euclidean space. A Riemannian metric should be given to describe the geometric structure of the SPD matrix and to measure the distance of two points on S_{++}^d.

Geodesics on the manifold are related to the tangent vectors in the tangent space. Two operators, exponential map $\exp_X(\cdot): T_X M \to M$ and the logarithm map $\log_X(\cdot) = \exp_X^{-1}(\cdot): M \to T_X M$, are defined over differentiable manifolds to switch between the manifold and its tangent space at X. As illustrated in Fig. 5, the tangent vector v is mapped to the point Y on the manifold through the exponential map. The length of v is equivalent to the geodesic distance between X and Y due to the property of the exponential map. Conversely, a point on the manifold is mapped to the tangent space $T_X M$ through the logarithm map. As point X moves along the manifold, the exponential and logarithm maps change. The details can be referred in Harandi et al. (2012).

For S_{++}^d, the logarithm and exponential maps are given by:

$$\log_X(A) = X^{\frac{1}{2}}\log(X^{-\frac{1}{2}}AX^{-\frac{1}{2}})X^{\frac{1}{2}}, \quad (9)$$

$$\exp_X(y) = X^{\frac{1}{2}}\exp(X^{-\frac{1}{2}}yX^{-\frac{1}{2}})X^{\frac{1}{2}}, \quad (10)$$

where $\log(\cdot)$ and $\exp(\cdot)$ are the matrix logarithm and exponential operators, respectively. For SPD matrices, they can be computed through Singular Value Decomposition (SVD). If we let $\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_d)$ be a diagonal matrix formed from real values $\lambda_1, \lambda_2, \ldots, \lambda_d$ on diagonal elements and $A = U\mathrm{diag}(\lambda_i)U^T$ be the SVD of the symmetric matrix A, then

$$\log(A) = \sum_{r=1}^{\infty} \frac{(-1)^{r-1}}{r} (A - I)^r = U \mathrm{diag}(\ln(\lambda_i)) U^T, \quad (11)$$

$$\exp(A) = \sum_{r=0}^{\infty} \frac{1}{r!} X^r = U \mathrm{diag}(\exp(\lambda_i)) U^T, \quad (12)$$

where I is an identity matrix on manifolds.

The manifold can be embedded into its tangent space at identity matrix I. Thus, ~~B~~based on the bi-invariant Riemannian metric (Arsigny et al., 2008), the distance between A and B on the SPD matrix manifold is $d(A, B) = \|\log(A) - \log(B)\|_2$, where $\log(\cdot)$ denotes matrix logarithm. ~~The logarithmic operation is applied in the elements of the diagonal matrix obtained by the singular value decomposition of the SPD matrix.~~ Since symmetric matrices (equivalently tangent spaces) form a vector space, the classification tools in the Euclidean space (SVM, KNN and so on) can be seamlessly employed to deal with the recognition problem.

The logarithmic operator is valid only if the eigenvalues of the symmetric matrix are positive. When no cloud is observed in the clear sky, the CovD of the image features could be non-negative definite, and in this case, it needs to be converted to a SPD matrix. We can formulate it as an optimization problem (Harandi et al., 2015):

~~$\min_A \|C - A\|_F$~~

~~$\mathrm{s.t.}\ A + A^T > 0$~~ (9)

$$A^* = \arg\min_A \|C - A\|_F, \mathrm{s.t.}\ A + A^T > 0, \quad (13)$$

where C is a CovD and A^*~~A~~ is the closest SPD matrix to C.

For a SPD matrix A, its ~~corresponding feature vector can be represented as Eq. (10):~~ log-Euclidean vector representation $a \in \mathbb{R}^m$, $m = d(d+1)/2$, is unique and can be represented as $a = Vec(\log(A))$. Let $B = \log(A)$, $B \in S^d$ and

$$B = \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} & \cdots & b_{1,d} \\ b_{2,1} & b_{2,2} & b_{2,3} & \cdots & b_{2,d} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ b_{d,1} & b_{d,2} & b_{d,3} & \cdots & b_{d,d} \end{bmatrix}_{d \times d}, \quad (14)$$

which lies in the Euclidean space. Since B is symmetric, we can rearrange it into a vector by vectorizing the upper triangular matrix:

~~$a = Vec(\log(A))$~~ (10)

~~where $Vec(B)$ is a vector in Euclidean space by vectorizing the upper triangular matrix B:~~

$$a = Vec(B) = \left[b_{1,1}, \sqrt{2} b_{1,2}, \cdots, \sqrt{2} b_{1,d}, b_{2,2}, \sqrt{2} b_{2,3}, \cdots, b_{d,d}\right]^{\mathrm{T}}. \quad (~~11~~15)$$

Vector a is defined as the manifold features. Since f is a 6-dimensional feature mapping, the manifold feature vector a to describe the cloud image is $6 \times (6+1)/2 = 21$~~21~~ dimensions. The mapped feature vector can reflect the characteristics of its corresponding SPD matrix on matrix manifolds. Thus, manifold features can describe the non-Euclidean property of the infrared image features to some degree.

2.2.3 Combining manifold and texture features

As described in Sect. 2.2.1 and 2.2.2, manifold and texture features can be extracted and integrated to represent the ground-based infrared images. For an image, its four features including energy, entropy, contrast and homogeneity from GLCM, express its texture, while 21-dimensional manifold features describe the non-Euclidean geometric characteristics. The manifold and texture features are combined to form a feature vector to represent the image. ~~Besides the 4-dimensional texture features~~Thus, the joint features of the infrared image have a total of 25 dimensions.

2.3 Classification

2.3.1 Support vector machine

The classifier used in this paper is the SVM (Cristlanini and Shawe-Taylor, 2000), which exhibits prominent classification performance in the cloud type recognition experiments. (Zhuo et al., 2014; Li et al., 2016; Shi et al., 2017). In machine learning, SVMs are supervised learning models. An SVM model is a representation of the examples as points in the Reproducing Kernel Hilbert Space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. As Fig. 6 shows, given a set of two-class training examples (denoted by × and o, respectively), the key problem is to find the optimal hyperplane to do the separation: $w^T x + b = 0$, where w is a weight vector and b is a bias, and an SVM training model with the largest margin $2/\sqrt{w^T w}$ is built. The support vectors are the samples on the dotted lines. The optimization classification hyperplane is determined by the solid line. The test examples are assigned to one category or the other based on this model, making it a non-probabilistic binary linear classifier. ~~It is a two-class classifier; its basic model is a linear classifier with the largest margin in the feature space. The margin maximization can be formulated as a convex quadratic programming problem.~~ In this work, we apply ~~A~~ a simple linear function ~~is chosen~~ as the mapping kernel, which is validated by the cloud classification experiments.

2.3.2 Multi-class support vector machine method

For a multi-class ~~multiple classification~~ task, ~~the SVM is conducted between every two classes. If there are c types, then the total number of classification operation is $c(c-1)/2$.~~ one binary SVM classifier is constructed for every pair of distinct classes, and so, all together $c(c-1)/2$ binary SVM classifiers are constructed. For an unknown-type sample, it will be input into these binary classifiers and each classifier makes its vote, thus $c(c-1)/2$ independent output labels are obtained. The most frequent label is the sample's type. c is 5 in this paper and the final result is determined by the voting policy.

3. Experiments and discussions

In this section, we validate which features are chosen and report experimental results ~~and~~ to assess the performance of the proposed cloud classification method. Different from a deterministic case, the training samples of the experiments are chosen

randomly. ~~We first validate t~~The effects of the proposed features are first tested by conducting the 10-fold cross validation (Li et al., 2016; Gan et al., 2017)~~each experiment~~ 50 times on two datasets, respectively, and average values are taken as final results. ~~The results of 10-fold cross validation with different features are given in Table 3.~~ In the 10-fold cross validation, ~~Each~~ each dataset is divided into 10 subsets with the same size at random. One single subset is used for validation in turn and the other 9 parts are taken as the training set. The results of 10-fold cross validation with different features are given in Table 3. As Table 3 illustrates, the overall accuracy of texture, manifold and combined features achieves 83.49%, 96.46% and 96.50% on the zenithal dataset while 78.01%, 82.38% and 85.12% on the whole-sky dataset, respectively. It can be seen that the texture or manifold features alone don't achieve a better performance than the joint features, which not only inherit the advantage of the texture features, but also own the characteristic of manifold features. On the whole, the method using the joint features performs best in the cross validation.

Naturally, combined features are used for the cloud type recognition. In the experiment, each dataset is grouped into the training set and testing set. The training set is selected randomly from each category in accordance with a certain proportion 1/10, 1/2 and 9/10, respectively and the rest part forms the testing set. Each experiment is implemented 50 times to reduce the accidental bias and the average accuracy is regarded as the final results of classification to evaluate the performance of the method.

To exhibit the recognition performance of the proposed method, we also compare with the other two models (Liu et al., 2015; Cheng and Yu, 2015) to assess its performance in this experiment. Liu's model employs WLBP feature with the KNN classifier based on the chi-square distance while Cheng's method adopts the statistical and uniform LBP features with the Bayesian classifier. Note that we extract the statistical features from the greyscale images rather than from the RGB images so that the statistical features have only 8 dimension, as a result, without extra colour information provided, both of the two methods are adaptable to the infrared images.

3.1 Results of the zenithal dataset

The first experiment is performed on the zenithal dataset. Table 4 reports the overall recognition rates of the proposed method and the other methods. The proposed method gets the best results, with at least 2.5% improvement over Liu's method and over 9.5% higher than Cheng's method. Meanwhile, the proposed method demonstrates a more stable and more superior performance than the other two methods, especially when 1/10 of the dataset is treated as the training set. In this case, the proposed method is up to 90.85% on the overall accuracy while the other two methods achieve 81.30% and 81.64%, respectively. That means discriminative features used for classification can be gained even with insufficient training data as well. Although only three cases are given when the fractions of training set are 1/10, 1/2 and 9/10, they can represent most cases when training samples are insufficient, enough and quite enough. In general, with the increase in the number of training samples, the overall accuracy of testing samples will increase until it holds stable, which is in line with the results in Table 4. As a result, when the training images are typical and sufficient enough, there is no doubt that the recognition rate will be improved.

In Fig. ~~4~~7, the classification results of the proposed method are demonstrated in the form of the confusion matrix (Zhuo et al., 2014; Liu et al., 2015; Li et al., 2016) ~~-~~when 1/2 of the dataset constructs the training set while the rest 1/2 is used for testing. In the confusion matrix, ~~Each~~ each row of the matrix represents an actual class while each column represents the predicted class given by SVM. For example, the element in the second row and third column is the percentage of cumuliform ~~cululiform~~ clouds misclassified as waveform clouds. Therefore, the recognition rate for each class is in the diagonal of the matrix. The discrimination rate of stratiform clouds is up to 100%, which indicates that stratiform clouds have the most significant features to be distinguished among five cloud types. Likewise, the results of the other four cloud types achieve over 93%. It is shown that a rather high accuracy of each cloud type ~~has~~ is reached, which means the proposed method performs well in classifying the ground-based infrared zenithal images on the whole.

3.2 Results of the whole-sky dataset

The second experiment is performed on the whole-sky dataset, which is more challenging because there ~~exits~~ exists larger inner-class difference than that of the zenithal dataset. The experimental configuration retains the same in Sec. 3.1. Table 5 lists the results of different methods. It is illustrated that the proposed method gains the overall accuracy of 78.27%, 83.54% and 85.01% as the proportion of the training set varies. In comparison, Liu's method achieves 73.58%, 80.55% and 81.31% while Cheng's method achieves 66.99%, 67.36% and 68.18%, correspondingly. Comparing to the other two methods, the experimental results indicate the effectiveness of the proposed method with an obvious improvement in the accuracy. Similarly, the cases when the fractions of training set are 1/10, 1/2 and 9/10, can represent most cases when training samples are insufficient, enough and quite enough. Generally, the rise in the number of training samples makes the overall accuracy improve, which is in line with the results in Table 5. In a nutshell, a sufficient and representative training set can further promote the classification accuracy.

Figure ~~5~~ 8 displays the confusion matrix of the whole-sky dataset when 1/2 for training is used. The number of each category in the training set is 123, 120, 120, 23 and 44, respectively and the remaining part is treated as the testing set. It is demonstrated that stratiform clouds and clear sky possess obvious characteristics for classification while cumuliform, waveform and cirriform clouds pose a great challenge for a high accuracy of classification. Cirriform clouds are likely to be confused with the clear sky and about 15.22% of cirriform cloud images are misclassified as the clear sky in the experiment. In the whole-sky image, when it is on the condition of cirriform clouds, the area of cirriform clouds may be just a fraction of the whole sky, making it hard to be distinguished correctly. What's more, multiple cloud types could exist in the whole-sky condition, which may result in a relatively low accuracy of the single-type classification, like cumuliform, waveform and cirriform clouds.

There are some misclassifications, just as demonstrated in Fig. ~~6~~9. Figure ~~6~~9(a) shows that stratiform clouds are recognized as waveform clouds. It can be seen that the cloud base has a little fluctuation and makes it similar to the waveform cloud. Figure ~~6~~9(b) shows that cumuliform clouds are recognized as waveform clouds. We can distinguish it as waveform clouds by the shape but the strong vertical motion of cumuliform clouds makes it hard to differ from waveform

clouds. Figure ~~6~~9(c) shows that cumuliform clouds are recognized as cirriform clouds. In this image, besides cumuliform clouds, a little cirriform clouds can also be found. Figure ~~6~~9(d) shows that waveform clouds are recognized as cumuliform clouds. It can be seen that both waveform and cumuliform clouds coexist in the sky. Figure ~~6~~9(e) shows that cirriform clouds are recognized as cumuliform clouds. It is admitted that the whole-sky dataset is more complicated than the zenithal dataset as the weather conditions change.

4. Conclusions

In this paper, a novel cloud classification method of the ground-based infrared images, including the zenithal and whole-sky datasets, is proposed. Besides the texture features computed from the GLCM, manifold features obtained from the SPD matrix manifold are combined together. With the joint features, the proposed method can improve the recognition rate of the cloud types ~~dramatically~~. On ~~the~~ one hand, the joint features can inherit the advantages of the statistical features, which represent texture information in Euclidean space; on the other hand, the manifold features on the matrix manifold~~the statistical learning method on the manifold~~ can describe the non-Euclidean geometric structure of the image features and thus the proposed method can benefit from it for a high classification precision. The CovD is calculated by extracting 6-dimensional features including greyscale, first-order and second-order gradient information, and the mean values are subtracted from the feature vectors, which may improve the recognition performance to some extent, since it can remove the noises of the infrared images. The manifold feature vector is produced by mapping the SPD matrix into the tangent space and afterwards the combined feature vector is adopted for cloud type recognition with SVM. With different fractions that the training set occupies, it is validated that in most cases the proposed method outperforms the other two methods (Liu et al., 2015; Cheng and Yu, 2015). As a whole, the improvement of the proposed method is between 2% and 10%. To some degree, it may not be a great improvement, but we have validated that the introduction of manifold features is effective and can achieve some success, it is worthy doing more work in this field to promote its development.

In future work, more suitable image features like Gabor or wavelet coefficients (Liu and Wechsler, 2002) can be incorporated into the SPD matrix and the classification would be performed directly on the manifolds to improve the recognition rate further. Besides, feature extraction using deep learning method such as convolutional neural networks can be taken into account to increase the classification accuracy. What's more, the addition of the brightness temperature, or the height information obtained from the laser ceilometer might be helpful for the improvement of the cloud type recognition accuracy. It is found that the proposed method is effective to satisfy the requirement of the cloud classification task on ~~the~~ both zenithal and whole-sky datasets. The complex sky condition with multiple cloud types should arise our concern in the next work.

5. Code availability

The code of the proposed method will be available online upon the acceptance of the paper.

6. Data availability

The two ground-based infrared cloud datasets used in this paper will be available upon the acceptance of the paper. This will be helpful for further benchmarking ground-based infrared image classification.

Acknowledgements. This work is in part supported financially by the National Natural Science Foundation of China under Grant No.61473310 and No.41174164.

References

Arsigny, V., Fillard, P., Pennec, X., and Ayache, N.: Geometric means in a novel vector space structure on symmetric positive-definite matrices, Siam J. Matrix Anal. A., 29, 328-347, doi:10.1137/050637996, 2008.

Bensmail, H., and Celeux, G.: Regularized Gaussian discriminant analysis through eigenvalue decomposition, J. Am. Stat. Assoc., 91, 1743-1748, doi:10.1080/01621459.1996.10476746, 1996.

Buch, K. A., Sun Chen-Hui, and Thorne L. R.: Cloud classification using whole-sky imager data, In Proceedings of the 5th Atmospheric Radiation Measurement Science Team Meeting, San Diego, CA, USA, 1995.

Calbó, J., and Sabburg, J.: Feature extraction from whole-sky ground-based images for cloud-type recognition, J. Atmos. Ocean. Techn., 25, 3, doi:10.1175/2007JTECHA959.1, 2008.

Cazorla, A., Olmo, F. J., and Aladosarboledas, L.: Development of a sky imager for cloud cover assessment, J. Opt. Soc. Am. A., 25, 29-39, doi:10.1364/JOSAA.25.000029, 2008.

Chen, T., Rossow, W. B., and Zhang, Y.: Radiative effects of cloud-type variations. J. Climate, 13, 264-286, doi:10.1175/1520-0442(2000)013<0264, 2000.

Cheng, H. Y., and Yu, C. C.: Block-based cloud classification with statistical features and distribution of local texture features, Atmos. Meas. Tech., 8, 1173-1182, doi:10.5194/amt-8-1173-2015, 2015.

Cristianini, N. and Shawe-Taylor, J.: An introduction to support vector machines and other kernel-based learning methods, Cambridge university press, 2000.

Faraki, M., Palhang, M., and Sanderson, C.: Log-Euclidean bag of words for human action recognition, IET Comput. Vis., 9, 331-339, doi:10.1049/iet-cvi.2014.0018, 2015.

Gan, J., Lu, W., Li, Q., Zhang, Z., Yang, J., Ma, Y. and Yao, W.: Cloud type classification of total-sky images using duplex norm-bounded sparse coding. IEEE J. Sel. Top. Appl., 10, 3360-3372, doi:10.1109/JSTARS.2017.2669206, 2017.

Ghonima, M. S., Urquhart, B., Chow, C. W., and Shields, J. E.: A method for cloud detection and opacity classification based on ground based sky imagery, Atmos. Meas. Tech., 5, 4535-4569, doi:10.5194/amt-5-2881-2012, 2012.

Haralick, R. M., Shanmugam, K., and Dinstein, I. H.: Textural features for image classification, IEEE T. Syst. Man Cyb., 3, 610-621, doi:10.1109/TSMC.1973.4309314, 1973.

Harandi, M. T., Hartley, R., Lovell, B., and Sanderson, C.: Sparse coding on symmetric positive definite manifolds using Bregman divergences, IEEE T. Neur. Net. Lear., 27, 1294-1306, doi:10.1109/TNNLS.2014.2387383, 2015.

Hartmann, D. L., Ockert-bell, M. E., and Michelsen M. L.: The effect of cloud type on earth's energy balance: global analysis. J. Climate., 5, 1281-1304, doi:10.1175/1520-0442(1992)005<1281, 1992.

Heinle, A., Macke, A., and Srivastav, A.: Automatic cloud classification of whole sky images, Atmos. Meas. Tech., 3, 557-567, doi:10.5194/amt-3-557-2010, 2010.

Illingworth, A. J., Hogan, R. J., O'connor, E. J., et al.: Cloudnet: Continuous evaluation of cloud profiles in seven operational models using ground-based observations, B. Am. Meteorol. Soc., 88, 883-898, 2007.

Isaac, G. A., and Stuart, R. A.: Relationships between cloud type and amount, precipitation, and surface temperature in the mackenzie river valley-beaufort sea area. J. Climate, 9, 1921-1941, doi:10.1175/1520-0442(1996)<1921, 1996.

Jayasumana, S., Hartley, R., Salzmann, M., Li, H., and Harandi, M. T.: Kernel methods on Riemannian manifolds with Gaussian RBF kernels, IEEE T. Pattern Anal., 37, 2464-2477, doi:10.1109/TPAMI.2015.2414422, 2015.

Li, J., Menzel, W. P., Yang, Z., Frey, R. A., and Ackerman, S. A.: High-spatial-resolution surface and cloud-type classification from MODIS multispectral band measurements, J. Appl. Meteorol., 42, 204-226, doi:10.1175/1520-0450(2003)042<0204, 2002.

Li, Q., Zhang, Z., Lu, W., Yang, J., Ma, Y. and Yao, W.: From pixels to patches: a cloud classification method based on a bag of micro-structures, Atmos. Meas. Tech., 9, 753-764, doi:10.5194/amt-9-753-2016, 2016.

Liu, C. and Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition, IEEE T. Image Process., 11, 467-476, doi:10.1109/tip.2002.999679, 2002.

Liu, L., Sun, X., Chen, F., Zhao, S., and Gao, T.: Cloud classification based on structure features of infrared images, J. Atmos. Ocean. Techn., 28, 410-417, doi:10.1175/2010JTECHA1385.1, 2011.

Liu, L., Sun, X., Gao, T., and Zhao, S.: Comparison of cloud properties from ground-based infrared cloud measurement and visual observations, J. Atmos. Ocean. Techn., 30, 1171-1179, doi:10.1175/JTECH-D-12-00157.1, 2013.

Liu, S., Zhang, Z., and Mei, X.: Ground-based cloud classification using weighted local binary patterns, J. Appl. Remote Sens., 9, 095062, doi:10.1117/1.JRS.9.095062, 2015.

Liu, Y., Key, J. R., and Wang, X.: The influence of changes in cloud cover on recent surface temperature trends in the Arctic. J. Climate, 21, 705-715, doi:10.1175/2007JCLI1681.1, 2008.

Naud, C. M., Booth, J. F. and Del Genio, A.D.: The relationships between boundary layer stability and cloud cover in the post-cloud-frontal region. J. Climate, 29, 8129-8149, doi: 10.1175/JCLI-D-15-0700.1, 2016.

Ripley, B. D.: Pattern recognition and neural networks, Cambridge University Press, 8 edn., 2005.

Sanin, A., Sanderson, C., Harandi, M. T., and Lovell, B. C.: Spatio-temporal covariance descriptors for action and gesture recognition, In Proceedings of 2013 IEEE Workshop on Applications of Computer Vision (WACV), Tampa, FL, USA, 103-110, doi:10.1109/WACV.2013.6475006, 2013.

Shi, C., Wang, C., Wang, Y., and Xiao, B.: Deep convolutional activations-based features for ground-based cloud classification, IEEE Geosci. Remote S., PP, 1-5, doi:10.1109/LGRS.2017.2681658, 2017.

Shields, J. E., Johnson, R. W., Karr, M. E., Burden, A. R., and Baker, J. G.: Daylight visible/NIR whole-sky imagers for cloud and radiance monitoring in support of UV research programs, SPIE International Symposium on Optical Science & Technology, 155-166, doi:10.1117/12.509062, 2003.

Singh, M., and Glennen, M.: Automated ground-based cloud recognition, Pattern Anal. Appl., 8, 258-271, doi:10.1007/s10044-005-0007-5, 2005.

Souzaecher, M. P., Pereira, E. B., Bins, L. S., and Andrade, M. A. R.: A simple method for the assessment of the cloud cover state in high-latitude regions by a ground-based digital camera, J. Atmos. Ocean. Techn., 23, 437, doi:10.1175/JTECH1833.1, 2006.

Sun, X. J., Liu, L., Gao, T. C., and Zhao, S. J.: Classification of whole sky infrared cloud image based on the LBP operator, Transactions of Atmospheric Sciences, 32, 490-497, doi:10.3969/j.issn.1674-7097.2009.04.004, 2009 (in Chinese).

Taravat, A., Frate, F. D., Cornaro, C., and Vergari, S.: Neural networks and support vector machine algorithms for automatic cloud classification of whole-sky ground-based images, IEEE Geosci. Remote S., 12, 666-670, doi:10.1109/LGRS.2014.2356616, 2014.

Tuzel, O., Porikli, F., and Meer, P.: Pedestrian detection via classification on Riemannian manifolds, IEEE T. Pattern Anal., 30,1713-1727, 2008.

Tzoumanikas, P., Kazantzidis, A., Bais, A. F., Fotopoulos, S., and Economou, G.: Cloud detection and classification with the use of whole-sky ground-based images, Atmos. Res., 113, 80-88, doi:10.1016/j.atmosres.2012.05.005, 2012.

Wang, Z., and Sassen, K.: Cloud type and macrophysical property retrieval using multiple remote sensors, J. Appl. Meteorol., 40, 1665-1683, doi:10.1175/1520-0450(2001)040<1665, 2001.

Zhuo, W., Cao, Z., and Xiao, Y.: Cloud classification of ground-based images using texture–structure features, J. Atmos. Ocean. Techn., 31, 79-92, doi:10.1175/JTECH-D-13-00048.1, 2014.

Table 1. The sky condition classes and corresponding description.

Sky condition classes	Description	Cloud types
Stratiform clouds	Horizontal, layered clouds that stretch out across the sky like a blanket	St、As、Cs (Sc、Ac、Cb、Ns)
Cumuliform clouds	Thick clouds that are puffy in appearance, like large cotton balls	Cu、Cb
Waveform clouds	Thin or thick clouds occurring in sheets or patches with wavy, rounded masses or rolls	Sc、Ac、Cc
Cirriform clouds	Thin clouds; very wispy and feathery looking	Ci
Clear sky	Clear	No clouds

Table 2. The numbers of each class on two datasets.

Sky condition classes	Zenithal	Whole-sky
Stratiform clouds	100	246
Cumuliform clouds	100	240
Waveform clouds	100	239
Cirriform clouds	100	46
Clear sky	100	88
Total	500	859

Table 3. The 10-fold cross validated classification accuracy (%) on two datasets.

	Zenithal	Whole-sky
Texture features	83.49	78.01
Manifold features	96.46	82.38
Combined features	**96.50**	**85.12**

Table 4. The overall classification accuracy (%) on the zenithal dataset. 1/10, 1/2 and 9/10 are the certain proportions of the training set selected randomly from each category, and the rest part forms the testing set correspondingly.

	1/10	1/2	9/10
Liu's method	81.64	92.24	93.48
Cheng's method	81.30	81.92	81.32
Proposed method	**90.85**	**95.98**	**96.36**

Table 5. The overall classification accuracy (%) on the whole-sky dataset. 1/10, 1/2 and 9/10 are the certain proportions of the training set selected randomly from each category, and the rest part forms the testing set correspondingly.

	1/10	1/2	9/10
Liu's method	73.58	80.55	81.31
Cheng's method	66.99	67.36	68.18
Proposed method	**78.27**	**83.54**	**85.01**

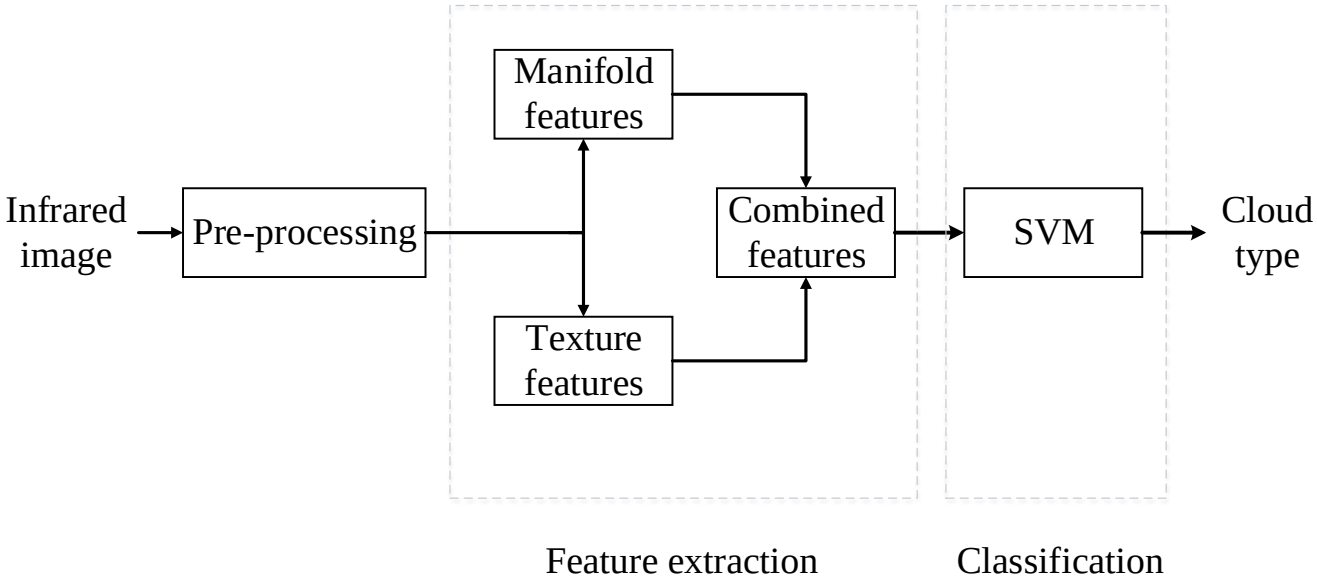

Figure 1: System framework.

Figure 2: Cloud samples from the zenithal dataset. (a) stratiform clouds, (b) cumuliform clouds, (c) waveform clouds, (d) cirriform clouds and (e) clear sky.

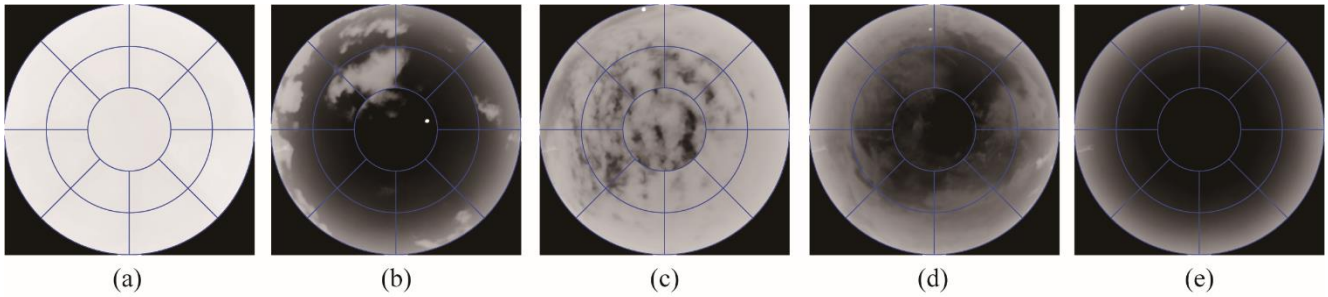

Figure 3: Cloud samples from the whole-sky dataset. (a) stratiform clouds, (b) cumuliform clouds, (c) waveform clouds, (d) cirriform clouds and (e) clear sky.

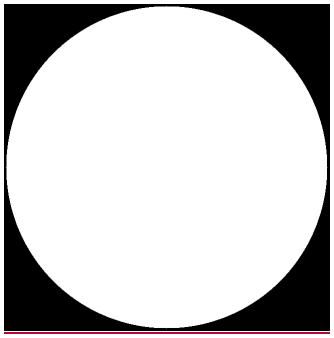

Figure 4: The mask of the whole-sky images. The area within the circle is the ROI, and the area outside the circle is not the ROI.

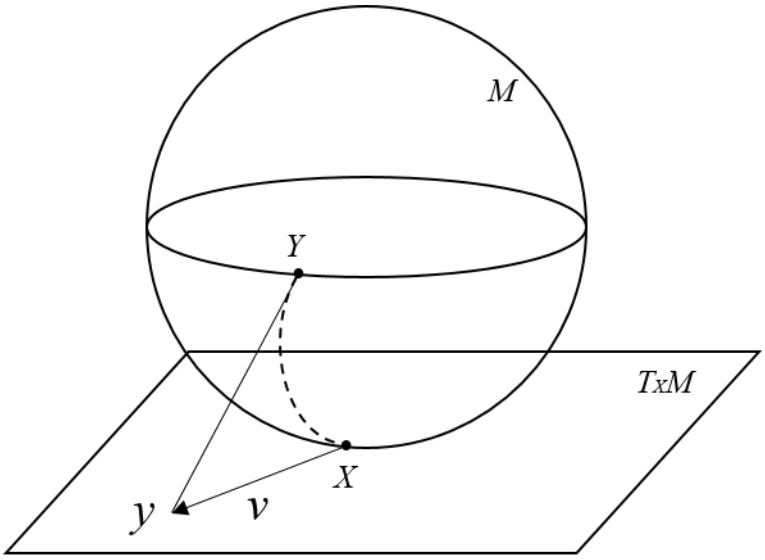

Figure 5: Illustration of the tangent space $T_X M$ at point X on a Riemannian manifold. A SPD matrix can be interpreted as point X in the space of SPD matrices. The tangent vector v can be obtained through the logarithm map, ie. $v = \log_X(Y)$. Every tangent vector in $T_X M$ can be mapped to the manifold through the exponential map, ie. $\exp_X(v) = Y$. The dotted line shows the geodesic starting at X and ending at Y.

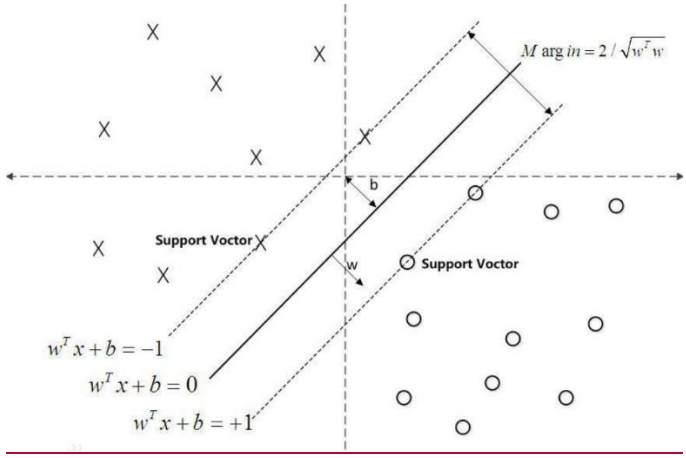

Figure 6: The decision boundary of support vector machine with the largest margin. × and o denote two-class training examples, respectively. $w^T x + b = 0$ is the optimal hyperplane to do the separation, where w is a weight vector and b is a bias, and an SVM training model with the largest margin $2/\sqrt{w^T w}$ is built. The support vectors are the samples on the dotted lines. The optimization classification hyperplane is determined by the solid line.

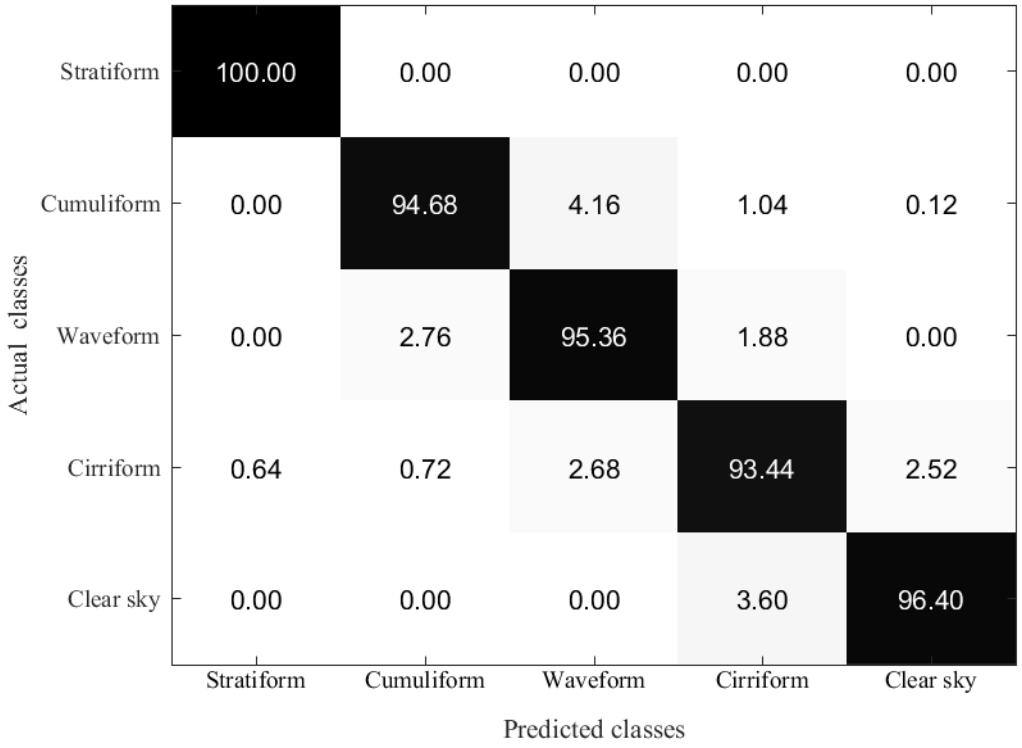

Figure 47: Confusion matrix (%) on the zenithal dataset. (1/2 for training and the overall accuracy is 95.98%)

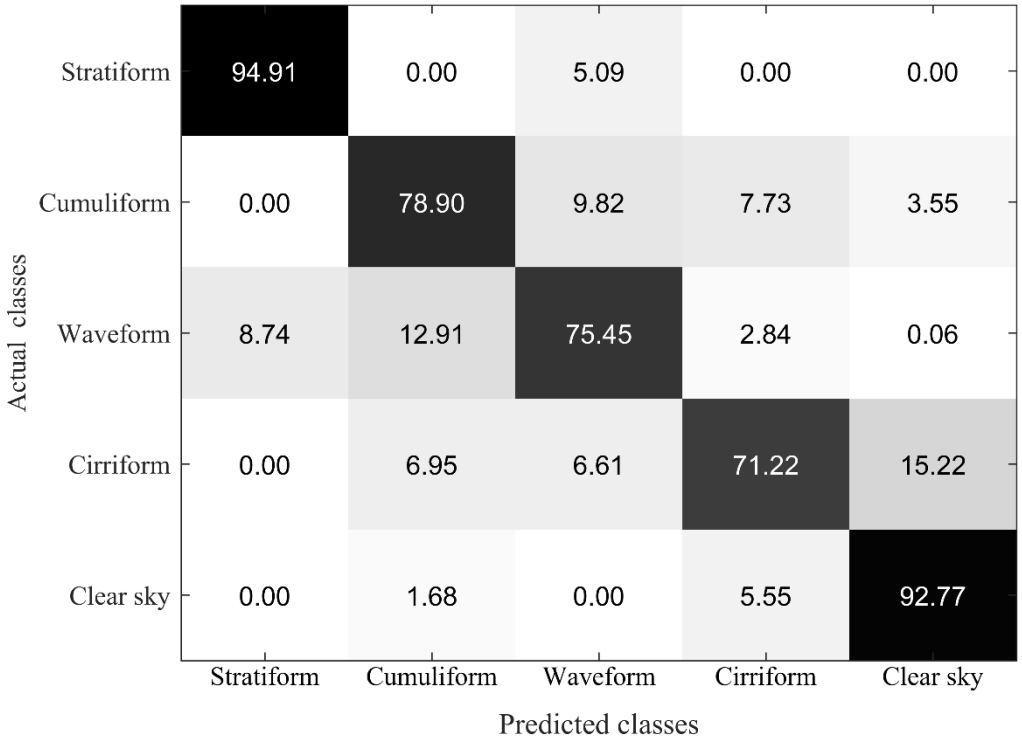

Figure ~~5~~8: Confusion matrix (%) on the whole-sky dataset. (1/2 for training and the overall accuracy is 83.54%)

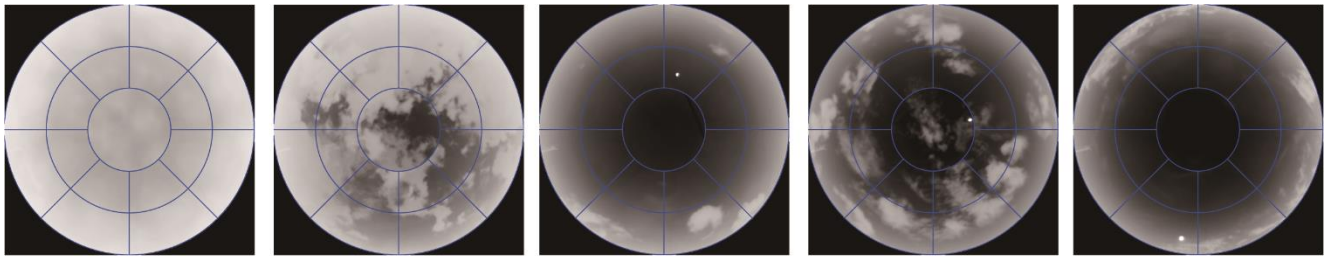

Figure ~~6~~9: Selected misclassified whole-sky images. (a) stratiform clouds to waveform clouds, (b) cumuliform clouds to waveform clouds, (c) cumuliform clouds to cirriform clouds, (d) waveform clouds to cumuliform clouds and (e) cirriform clouds to cumuliform clouds.