**Atmospheric
Measurement
Techniques**

Automatic cloud classification of whole sky images

A. Heinle[1], **A. Macke**[2], **and A. Srivastav**[1]

[1]Excellence Cluster "The Future Ocean", Department of Computer Science, Kiel University, Kiel, Germany
[2]Leibniz Institute of Marine Sciences at Kiel University (IFM-GEOMAR), Kiel, Germany

Abstract. The recently increasing development of whole sky imagers enables temporal and spatial high-resolution sky observations. One application already performed in most cases is the estimation of fractional sky cover. A distinction between different cloud types, however, is still in progress. Here, an automatic cloud classification algorithm is presented, based on a set of mainly statistical features describing the color as well as the texture of an image. The k-nearest-neighbour classifier is used due to its high performance in solving complex issues, simplicity of implementation and low computational complexity. Seven different sky conditions are distinguished: high thin clouds (cirrus and cirrostratus), high patched cumuliform clouds (cirrocumulus and altocumulus), stratocumulus clouds, low cumuliform clouds, thick clouds (cumulonimbus and nimbostratus), stratiform clouds and clear sky. Based on the Leave-One-Out Cross-Validation the algorithm achieves an accuracy of about 97%. In addition, a test run of random images is presented, still outperforming previous algorithms by yielding a success rate of about 75%, or up to 88% if only "serious" errors with respect to radiation impact are considered. Reasons for the decrement in accuracy are discussed, and ideas to further improve the classification results, especially in problematic cases, are investigated.

1 Introduction

Clouds are one of the most important forces of Earth's heat balance and hydrological cycle, and at the same time one of the least understood. It is well known that low clouds provide a negative feedback and high, thin clouds a positive feedback on the radiation budget. The net effect of clouds, however, is still unknown and they cause large uncertainties in climate models and climate predictions (Solomon et al., 2007).

The effect of clouds on solar and terrestrial radiation is due to reflection and absorption by cloud particles and depends strongly on the volume, shape, thickness and composition of the clouds. Large-scale cloud information is available from several satellites, but such data is provided in a low resolution and may contain errors. For example, small clouds are often overlooked due to the limited radiometer field of view. Low or thin clouds and surface are frequently confused because of their similar brightness and temperature (Ricciardelli et al., 2008; Dybbroe et al., 2005). Additionally, the solar radiation reaching the ground with respect to the cloud type cannot be determined, even though this is essential for cloud-radiation studies.

Nowadays ground-based imaging devices are commonly used to support satellite studies (Cazorla et al., 2008; Feister and Shields, 2005; Sakellariou et al., 1995). One of the best known commercial manufacturer of such instruments is the Scripps Institute of Oceanography at the University of California San Diego. Their Whole Sky Imagers are constructed to measure sky radiance at diverse wavelength bands (visible spectrum and near infrared) across the whole hemisphere (Shields et al., 1998, 2003). Due to the high-quality components involved, these imagers are often too expensive for small research groups. Therefore, as a cost-effective alternative, a few research institutions in several countries have developed non-commercial sky cameras for their own requirements (Pagès et al., 2002; Seiz et al., 2002; Pfister et al., 2003; Souza-Echer et al., 2006; Kalisch and Macke, 2008). In most cases an upward looking fisheye-objective is used to image the whole sky with a field of view (FOV) of about 180°. Individual algorithms to automatically estimate cloud cover already exist for many of them (Pfister et al., 2003; Long et al., 2006; Kalisch and Macke, 2008). Automatic cloud type recognition, however, is still under development and few papers have been published on that subject.
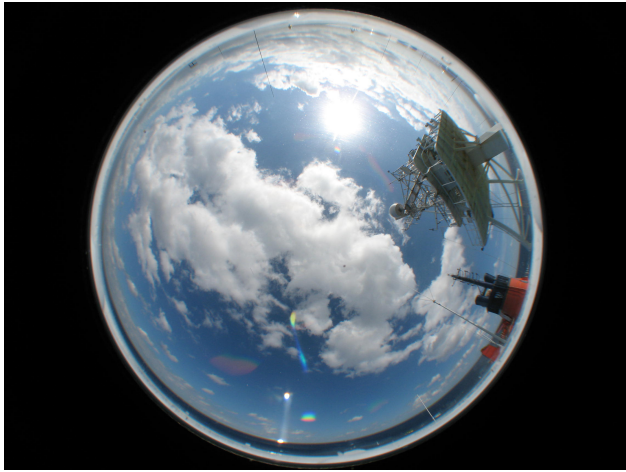
Fig. 1. All-sky image example used for training (2 November 2007, 12:48 UTC).

In one prior study Singh and Glennen (2005) present an approach of cloud classification for common digital images (without 180° FOV) to be utilized in air traffic control. Numerous features have been extracted and used to distinguish five different sky conditions, but the authors acknowledge their results as modest. Another recent paper (Calbó and Sabburg, 2008) introduces some possible criteria for sky-images to classify eight predefinded sky conditions, which include statistical features, features based on the Fourier transform, and features that need the prior distinction between clear and cloudy pixels. However, the classifier is based on a very simple classification method and achieves an accuracy of only 62%. Other publications handle simpler issues such as the estimation of cloud base height or the identification of thin and thick cloudiness (e.g. Seiz et al., 2002; Kassianov et al., 2005; Long et al., 2006; Cazorla et al., 2008; Parisi et al., 2008). Parisi et al. (2008) in particular report that they were not able to classify cloud type.

The objective of this study is the development of a fully automated algorithm classifying all-sky images in real-time with high accuracy. The cloud camera and the associated imager data are introduced in the following section. In Sect. 3 the features used to classify cloud types as well as the algorithm, a k-nearest-neighbour (kNN) classifier assigning the pre-processed images due to their feature vector to one of seven different sky conditions, are presented. The performance and results of the algorithm are discussed in Sect. 4, and Sect. 5 contains the summary and proposals for future research.

2   Data

2.1   Camera

The images used to develop the algorithm have been obtained by one of two cloud cameras constructed to enable cost-effective continuous sky observations for research associated with radiative transfer at the Leibniz Institute of Marine Sciences at the University of Kiel (IFM-GEOMAR). These all-sky imagers are based on commercially available components and are designed to be location-independent and run during adverse weather conditions, as one of them is primarily operating onboard a research vessel. The basic component is a digital camera equipped with a fisheye lens to provide a field of view larger than 180°, enclosed by a water and weather resistant box. In order to obtain a high temporal resolution, the cameras are programmed to acquire one image every 15 s, stored in 30-bit color JPEG format with a maximal resolution of 3648×2736 pixel. As such, the images are rectangular in shape, but the whole sky mapped is circular where the center is the zenith and the horizon is along the border (spherical projection, see Fig. 1). More details about the cameras and their usage can be found in Kalisch and Macke (2008).

2.2   Images

For the development of the cloud type classification algorithm, images with a resolution of 2272×1704 pixel, captured during a transit of the german research vessel "Polarstern" from Germany to South Africa in autumn 2007 (ANT XXIV/1), are used (Schiel, 2009). In the course of this expedition, different climate zones in several seasons were crossed and therefore the acquired data covers a wide range of possible sky conditions and solar zenith angles.

To create an image set required for feature search and later training of the cloud type classifier, we screened the complete data set and selected approximately 1500 all-sky images from the 75 000 obtained onboard in total. The selection procedure focused on temporal independence and uniqueness with respect to our pre-defined cloud classes (see next section). Furthermore, we insured that the final image set includes a large variety of different cloud forms as well as images of different daytimes and consequently different states of solar zenith angle.

The training set generated in this fashion, called TRAIN, contains about 200 independent images per cloud class.

3   Algorithm

In this section, the individual cloud classes are presented, followed by an introduction to the methodology of the applied classifier, the kNN classifier. We then describe the preprocessing of the imager data and explain the features integrated, as well as the feature selection method.

Table 1. Classes to be distinguished.

Label	Cloud genera according to WMO	Description
1	*Cumulus*	Low, puffy clouds with clearly defined edges, white or light-grey
2	*Cirrus & Cirrostratus*	High, thin clouds, wisplike or sky covering, whitish
3	*Cirrocumulus & Altocumulus*	High patched clouds of small cloudlets, mosaic-like, white
4	*Clear sky*	No clouds and cloudiness below 10%
5	*Stratocumulus*	Low or mid-level, lumpy layer of clouds, broken to almost overcast, white or grey
6	*Stratus & Altostratus*	Low or mid-level layer of clouds, uniform, usually overcast, grey
7	*Cumulonimbus & Nimbostratus*	Dark, thick clouds, mostly overcast, grey

3.1 Cloud classes

In contrast to other publications handling automated cloud classification, we used phenomenological classes to be separated according to the International Cloud Classification System (ICCS) published in WMO (1987). Therein, ten genera are defined which represent the basis of our classification. Based on visual similarity we combined some genera (altostratus and stratus, cirrocumulus and altocumulus, cumulonimbus and nimbostratus) to avoid systematical misclassifications. Aditionally, we merged the genera cirrus and cirrostratus due to lack of available data showing the latter, as well as the difficulty in detecting very thin clouds, such as some kinds of cirrostratus. Besides, it must be noted that the class of clear sky includes not only images without clouds, but also images with cloudiness below 10%.

Despite these generalizations, the resulting classes (see Table 1) represent a suitable partitioning of possible sky conditions and are especially useful for radiation studies. In order to simplify the application of the cloud classes, each is labeled by an individual identification number (see also Table 1).

3.2 Classifier

To classify the images described in Sect. 2, the k-nearest-neighbour (kNN) method is chosen, which is part of the supervised, non-parametric classifiers (Duda and Hart, 2001). "Supervised" means that the separating classes are known and a training sample is used to train the classifier. Non-parametric classifiers in general do not assume an a-priori probability distribution. Compared with other classifiers, the kNN method is very simple (and therefore associated with only low computatitional costs) and simultaneously quite powerful (Serpico et al., 1996; Vincent and Bengio, 2001; Duda and Hart, 2001). Even in the specific field of cloud type recognition, some results for comparison with linear classifiers and neural networks exist, underlining the high performance of kNN classifiers (Singh and Glennen, 2005; Christodoulou et al., 2003).

kNN classifier. The assignment of an image to a specific class using kNN classifiers is performed by majority vote. After pre-processing, several spectral and textural features are extracted from an image. In the next step, the computed and normalized feature vector $\boldsymbol{x}$ is compared with the known feature vectors $\boldsymbol{x}^i$ of each element in the training data by means of a distance measure, in our case the Manhattan Distance

$$d(\boldsymbol{x}, \boldsymbol{x}^i) := \sum_{j=1}^{\dim} |x_j - x_j^i|. \qquad (1)$$

The class associated with the majority of the k closest matches determines the unknown class. In the case that this majority is not unique, the training date with the absolute smallest distance to the unknown image specifies the target class. Therefore, the composition of the training sample and a meticulous selection of suitable images is of great importance.

Complexity. The kNN classifier is often critizised for slow runtime performance and large memory requirements (in other words high time and space complexity, respectively). The time complexity of an algorithm is a measure of how much computation time is needed to run the algorithm and is thus dependent on the number of calculation steps. In the case of image classification, this measure refers to the computational expense in classifying an unknown image. Using the kNN classifier, all distances between the feature vector of this image and each of the n members of the training sample are required for the calculation. These distances depend on the dimension d of the feature vector and we get a total complexity of $O(nd)$ (here $n = 1497$ and $d = 12$).

Since kNN methods store a set of prototypes in memory once, the space complexity of such an algorithm is $O(nd)$ as well.

3.3 Pre-processing

To obtain suitable features for separating the defined classes, it is necessary to eliminate some areas of the analysed raw images, as they are rectangular in shape but the interesting part, the mapped sky, is circular. Due to varied camera locations, disruptive factors like ship superstructures or towering edifices may also be mapped on the image and should be excluded as well from further calculations. Additionally,
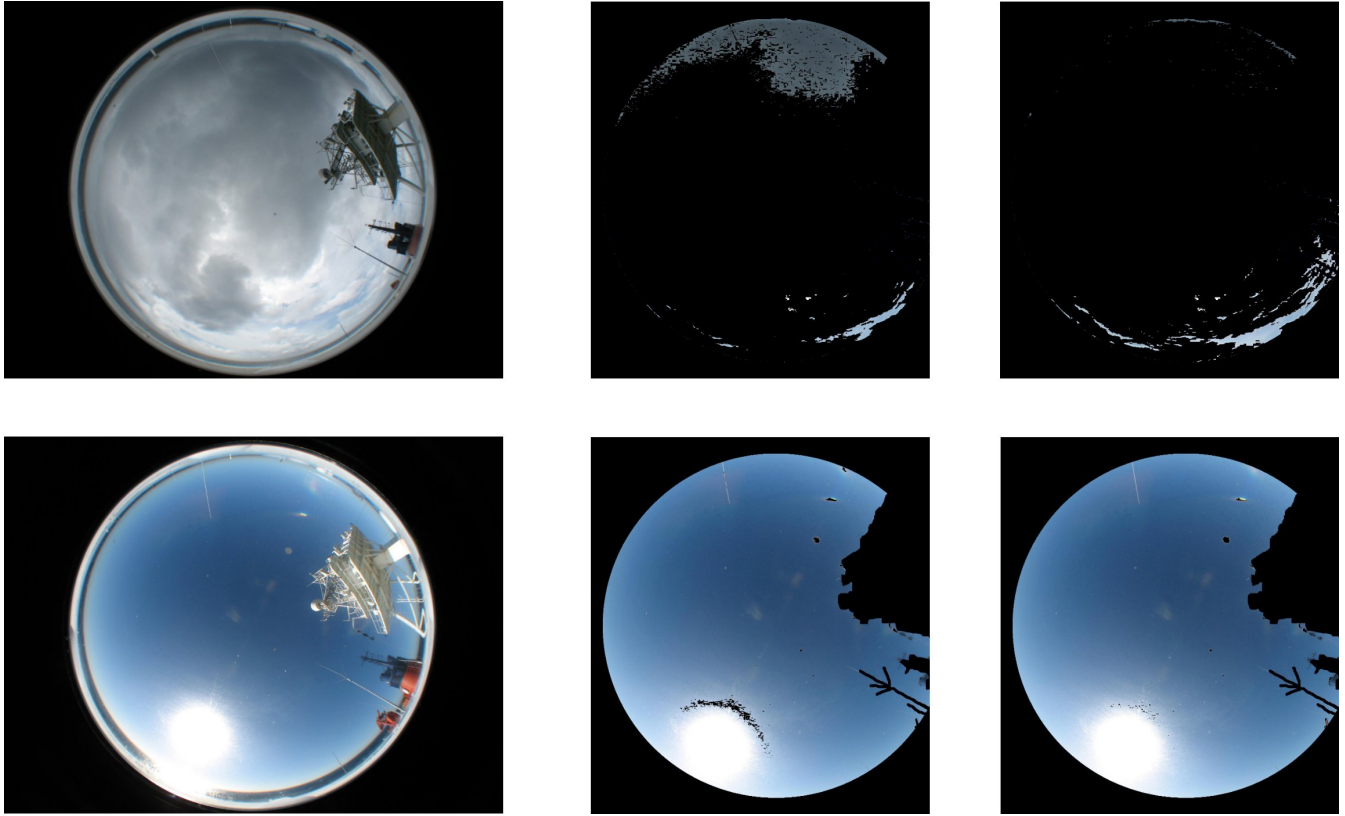
Fig. 2. Segmentation for optical thick clouds (top) and clear sky (bottom) using a treshold of $R/B = 0.8$ (middle) and a treshold of $R - B = 30$ (right).

analyses showed that it is useful to segment the images into clear and cloudy areas before calculating features.

Therefore, an image-mask, constructed by visually identifying image regions containing confounding information, is used first. The mask adapts the detected sections as well as completely white pixels (such as the ones displaying the sun) to the background by setting all corresponding pixel values to zero. Afterwards the remaining area is divided pixel by pixel into clear and cloudy regions, utilizing their red and blue pixel values.

In a clear atmosphere (without aerosols), more blue than red light is scattered by gas molecules, which is why clear sky appears blue to our eyes. In contrast, clouds (containing particles like aerosols, water droplets and/or ice crystals) scatter blue and red light to a similar extent, causing them to appear white or grey (Petty, 2006). Therefore, image regions with clear sky show relatively lower red pixel values compared to regions showing clouds, and the ratio R/B may be used to differentiate these areas. A separating threshold, whose exact value depends on both the camera used and prevailing atmospheric conditions, has to be determined. Suitable values are discussed in several papers handling cloud cover estimation (e.g. Pfister et al., 2003; Long et al., 2006).

However, during the testing phase we noticed problems in detecting thick clouds and classifying circumsolar pixels at the same time. Therefore we modified the criterion and considered the difference $R - B$ instead of the ratio R/B. Comparisons showed that segmentation using such a difference threshold still results in minor errors, but outperforms the ratio method. For our application the value $R - B = 30$ is optimal (see Fig. 2).

3.4 Features used

Out of numerous features tested (for example, features describing edge or color, features considering the run-length of primitives, their quantity or frequency, or features describing the texture of an image), we selected 12 features for application (see below). The choice of these features is based on their Fisher Distances F_{ij}^x, a selection criterion used in cloud classification work relating to satellite imagery (Pankiewicz, 1995). It is defined as

$$F_{ij}^x := \frac{|\mu_i^x - \mu_j^x|}{\sigma_i^x + \sigma_j^x}, \tag{2}$$

where μ_i^x and μ_j^x are the means of feature x with respect to class i and j, σ_i^x and σ_j^x the corresponding standard deviations. The features best suited to separate the defined classes are those which have the largest Fisher distances F_{ij}^x. It should be noted, however, that the feature set chosen in this way has to guarantee the separation of all classes. That means that features with smaller Fisher distances have to be included in the final set as well, if they discriminate classes which are not separated by other features with higher distances.

Most of the features are based on grey-scale images. Since the original data is provided in color, a partition into the three components R, G and B has to be performed before the features can be calculated. A simple transformation provides the grey-scale images, containing only the color information of one channel (R, G or B).

Spectral features. Spectral features describe the average color and tonal variation of an image. In cloud classification they are useful to distinguish between thick dark clouds, such as cumulonimbus, and brighter clouds, such as high cumuliform clouds, and to separate high and transparent cirrus clouds from others.

The spectral features implemented in the algorithm are the following:

– **Mean (R and B)**

$$\mathrm{ME}_i = \frac{1}{N} \sum_{l=0}^{N-1} a_l, \tag{3}$$

– **Standard deviation (B)**

$$\mathrm{SD}_i = \sqrt{\frac{1}{N-1} \sum_{l=0}^{N-1} (a_l - \mathrm{ME}_i)^2}, \tag{4}$$

– **Skewness (B)**

$$\mathrm{SK}_i = \frac{1}{N} \sum_{l=0}^{N-1} \left(\frac{a_l - \mathrm{ME}_i}{\mathrm{SD}} \right)^3, \tag{5}$$

– **Difference (R-G, R-B and G-B)**

$$\mathrm{D}_{ij} = \mathrm{ME}_i - \mathrm{ME}_j, \tag{6}$$

where $i, j \in \{\mathrm{R, G, B}\}$ and a_l denotes the corresponding grey value of pixel $l \in \{0, ..., N-1\}$.

In the brackets, R, G and B specify the color for which the individual feature is calculated. Due to the color of the sky and the different translucency of clouds, the color component B has the highest separation power. Thus most features are calculated for the grey-scale image containing the B color information.

Spectral features like the ones above support a division of cloud classes, but considering only those is not sufficient.

They do not provide information about the spatial distribution of color in an image. In most issues of pattern recognition and particularly in cloud type recognition, however, this distribution is equally significant. For example, images showing cumulus clouds and others showing altocumulus or stratocumulus clouds have similar mean color values and cannot be separated with those features. On the other hand, their spatial distribution of color is quite different, and other kinds of features can be added to separate those cases.

Textural features. To describe the texture of an image, statistical measures computed from Grey Level Co-occurrence Matrices (GLCM) may be used. A GLCM is a square matrix for which the number of rows equals the number of grey levels in the considered image. Every matrix element represents the relative frequency $P^\Delta(a, b)$ that two pixels occur, separated in a defined direction by a pixel distance $\Delta = (\triangle x, \triangle y)$, one with grey value a and the other with grey value b. To avoid dependency on image orientation, often an average matrix is calculated out of two or four matrices, expressing mutually perpendicular directions. Furthermore, because the computation of GLCMs strongly increases with increasing number of intensity levels G, it is advantageous to reduce the original number ($G = 256$) of grey levels.

The textural features used in this study are the following four of 14 statistical measures proposed by Haralick et al. (1973), computed from an average GLCM with pixel distance $\Delta = (1, 1)$:

– **Energy (B)**

$$\mathrm{EN}_i = \sum_{a=0}^{G-1} \sum_{b=0}^{G-1} P^\Delta(a, b)^2 \tag{7}$$

The energy shows the homogenity of grey level differences.

– **Entropy (B)**

$$\mathrm{ENT}_i = \sum_{a=0}^{G-1} \sum_{b=0}^{G-1} P^\Delta(a, b) \, \log_2 \, P^\Delta(a, b) \tag{8}$$

The entropy is a measure of randomness of grey level differences.

– **Contrast (B)**

$$\mathrm{CON}_i = \sum_{a=0}^{G-1} \sum_{b=0}^{G-1} (a - b)^2 \, P^\Delta(a, b) \tag{9}$$

Contrast is a measure of local variation of grey level differences.

– **Homogenity (B)**

$$\mathrm{HOM}_i = \sum_{a=0}^{G-1} \sum_{b=0}^{G-1} \frac{P^\Delta(a, b)}{1 + |a - b|}. \tag{10}$$

The homogeneity reflects the similarity of adjacent grey levels.

Table 2. Confusion matrix of CV for equally involved features in %.

True class	Classified as							CV
	1	2	3	4	5	6	7	
1	**93.73**	1.57	4.31	0.39	0.00	0.00	0.00	
2	0.72	**97.13**	1.07	1.07	0.00	0.00	0.00	
3	2.49	1.93	**95.30**	0.00	0.28	0.00	0.00	
4	0.00	1.20	0.00	**98.80**	0.00	0.00	0.00	
5	0.00	0.00	0.00	0.00	**96.43**	3.57	0.00	
6	0.00	0.00	0.94	0.00	0.00	**96.23**	2.83	
7	0.00	0.00	1.84	0.00	1.38	0.00	**96.77**	96.13

Cloud cover. In addition to the features described above, we computed the cloud cover (CC):

– **Cloud cover**

$$CC := \frac{N_{\text{bew}}}{N},\qquad(11)$$

where N_{bew} denotes the number of cloudy pixels.

CC is a measure of the average cloudiness, and for example stratiform clouds could be well distinguished from other sky conditions using this feature.

For each pre-classified image in the training sample TRAIN we computed the features presented and stored them with their assigned cloud class. Since the kNN classifer chooses the target class of an unknown image based on its distance in the feature space to the training images and the features differ in their value ranges, we normalized the features to the interval [0,100]. This ensures that all features are equally weighted in the decision process.

4   Results and discussion

In this section we describe the methodology used to estimate the performance of the created algorithm as well as to optimize the included parameters and the respective results are discussed. Afterwards, an additional test sample of random images is presented to assess the performance of the algorithm in classifying more ambiguous images.

The algorithm was implemented in IDL and tested on an Intel Celeron 530 with 1.73 GHz and 512 MB RAM. For one image it took about 1.3 s to return the classification result.

4.1   Methodology of performance estimation

To estimate the performance of the selected features and the created algorithm we applied the Leave-One-Out Cross-Validation (LOOCV). Cross validation methods in general have the advantage that they reuse the known training sample to estimate the capability of an algorithm, nevertheless being unbiased, instead of needing an additional test sample

(Ripley, 2005). Therefore, they are often used for validation or feature selection in the area of pattern recognition. In cloud type recognition the LOOCV has been applied by e.g. Tag et al. (2000) or Bankert and Wade (2007).

LOOCV. From the training sample T, one single element t is removed and the algorithm is trained with the remaining data $(T - t)$. Then the element excluded, which is independent from the data used for training, is classified. This is repeated n times, where n is the number of elements in T, such that each element in the training sample is used for validation exactly once. The average number of correctly classified elements

$$CV = \frac{|\{t \in T \,|\, t \text{ is classified correctly }\}|}{n}\qquad(12)$$

is finally used as measure of performance.

First results. The results of the first LOOCV performed are given in Table 2. All features were equally involved in the classification process and the parameter k, the number of considered neighbours (see Sect. 3.2), was set to 3 as a first guess. We see an overall accuracy of about 96%, where the class clear shows the best classification results with 98.8%. Confusions of these class primarily exist with cirrus clouds and also rarely with cumulus clouds in case of low cloudiness. This is caused by thin and transparent parts of cirrus clouds which cannot be detected by the algorithm. Consequently, such images are classified as clear sky. Moreover, the so-called "whitning effect" provides a missclassification of cloud free pixels near the solar disk. Such pixels are often whiter and brighter than the rest of the hemisphere due to forward scattering by aerosols and haze (see Fig. 3, left) and therefore interpreted as thin clouds by the algorithm (see also Long et al., 2006).

Most of the remaining cloud classes show accuracies of about 96% or 97%, except for the cumulus class and the class of high cumulus. Both have slightly lower hit ratios due to confusions among themselves which is caused by the difficulty in distinguishing those two classes. They differ only in the size of the individual cloudlets for which no clear boundary exists, so that a discrimination can be extremely difficult.

Table 3. Overall accuracy for $k = 3$ (top) and $k = 5$ (bottom) for the LOOCV using weighted features according to $\boldsymbol{w}_y = (1, 1, ..., x, .., 1)$ in %. That means, feature y is weighted x times as often as the others.

$k=3$	EN	ENT	CON	HOM	ME$_R$	ME$_B$	D$_{RG}$	D$_{RB}$	D$_{GB}$	SD	SK	CC
x												
2	96.1	95.9	96.1	96.2	96.1	96.1	96.0	96.3	96.6	96.0	95.5	96.1
10	94.6	93.3	92.3	94.9	93.3	93.5	95.5	94.8	95.5	94.5	93.9	94.8
20	92.7	91.3	89.0	93.5	90.8	91.9	93.3	92.7	93.9	93.0	91.9	92.8
1						96.1						

$k=5$	EN	ENT	CON	HOM	ME$_R$	ME$_B$	D$_{RG}$	D$_{RB}$	D$_{GB}$	SD	SK	CC
x												
2	95.1	95.0	95.0	95.3	95.1	95.1	95.4	95.4	95.7	95.1	94.7	95.3
10	93.1	92.1	92.3	94.3	91.9	92.9	94.3	93.9	94.6	92.9	93.0	93.4
20	90.6	89.5	88.7	91.5	89.7	91.2	91.8	91.2	92.1	91.2	91.2	92.1
1						95.3						
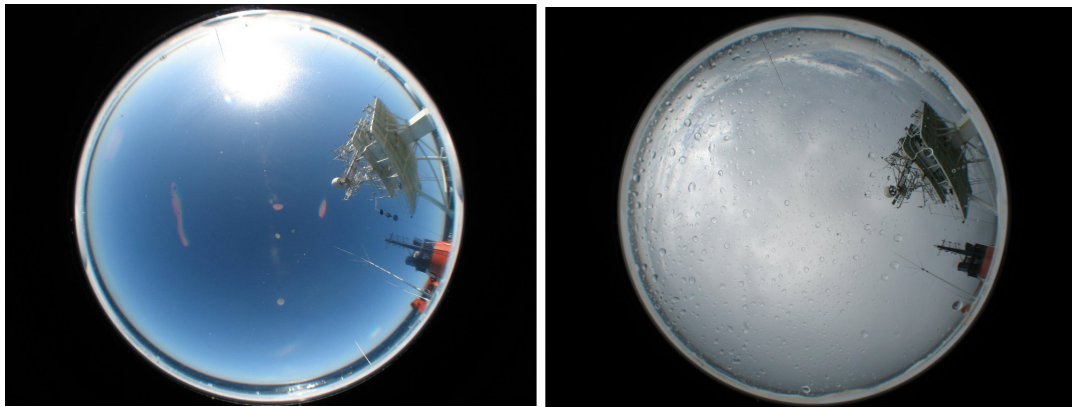
Fig. 3. Misclassification of clear sky caused by the "whitening effect" (left) and missclassification of stratus due to raindrops (right).

The next remarkable errors occur between stratocumulus, stratus and the class of thick clouds. Some cases of stratocumulus are classified as stratus, some images showing stratus are assigned to those showing thick clouds and, in turn, images with thick clouds are sometimes classified as stratocumulus. These confusions, however, are well understood. All three classes occur frequently as transitional forms from one in the other and the automatic classification of such images could differ from the manual preclassification.

Also, misclassifications of some images displaying stratus and thick clouds appear to be caused by raindrops on the camera protecting dome (see Fig. 3, right). The drops are naturally also mapped on the images and lead to texture feature values similar to those representing patchy altocumulus and cirrocumulus.

Apart from these errors, the first results, based on the guess of using 3 nearest neighbours, are quite good. However, we wanted to see if the performance of the algorithm could be improved by using another value of k or by weighting the individual features.

Improved results. For the LOOCV discussed above, all features were equally weighted. To assess whether improvements can be achieved by varying the impact of the individual features, we added a weight vector and ran the LOOCV for different configurations of this vector. Furthermore, because k, the number of neighbours considered, is a variable parameter, the LOOCV has also been carried out for different values of k.

Table 3 presents overall accuracies for $k = 3$ and $k = 5$ using the weight vectors $\boldsymbol{w}_y$, defined by

$$\boldsymbol{w}_y(i) := \begin{cases} x, & \text{if } i = y \\ 1, & \text{if } i \neq y \end{cases}, \tag{13}$$

Table 4. Confusion matrix of CV for optimal weighted features in %.

True class	Classified as							CV
	1	2	3	4	5	6	7	
1	**96.47**	1.18	1.96	0.39	0.00	0.00	0.00	
2	0.72	**98.57**	0.36	0.36	0.00	0.00	0.00	
3	1.93	1.66	**96.41**	0.00	0.00	0.00	0.00	
4	0.00	1.20	0.00	**98.80**	0.00	0.00	0.00	
5	0.00	0.00	0.00	0.00	**95.54**	3.57	0.89	
6	0.00	0.00	0.00	0.00	1.89	**94.34**	3.77	
7	0.00	0.00	1.84	0.00	0.46	0.00	**97.70**	97.06
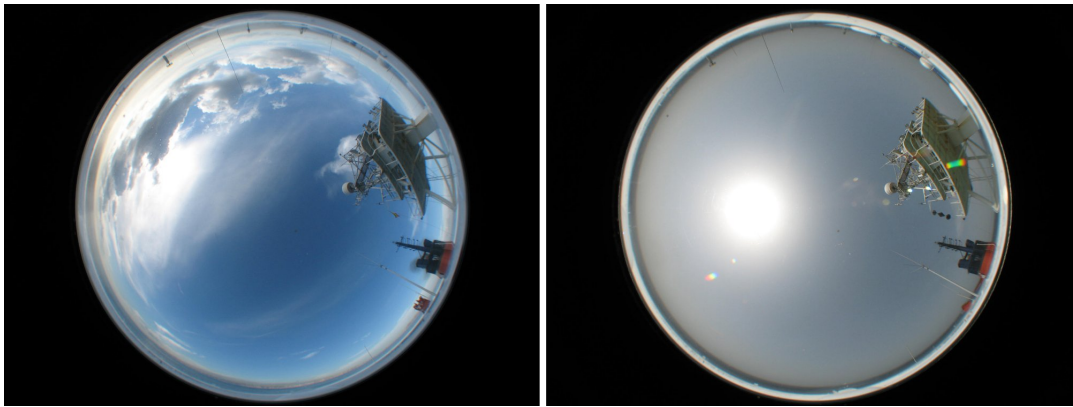
Fig. 4. Image showing cirrus and cumulus clouds (4 November 2007, 09:39 UTC) (left), image showing clear sky during a dust event (8 November 2007, 13:40 UTC) (right).

$i \in \{1, ..., 12\}$. In other words, feature y is weighted x-times, whereas the others are weighted once. As can be seen in Table 3, $k = 3$ superpasses $k = 5$ for all weights. Furthermore, weighted features appear to have certain potential to improve the classification rate. The best performance in our analyses realized the following weight vector: EN, ENT, CON, $\mathrm{ME_R}$, $\mathrm{ME_B}$, $\mathrm{D_{RG}}$, SD, SK and CC are rated once, HOM and $\mathrm{D_{RB}}$ are weighted twice and $\mathrm{D_{GB}}$ is counted three times, which indicates that a distinction between the cloud classes defined in this study is most feasible by utilizing the sky homogenity and the sky color. The analyses also showed that values of $k > 3$, in general, yield continuously decreasing performance. Thus, our first guess, the value $k = 3$, was confirmed to be the best choice.

Table 4 displays the confusion matrix of the LOOCV for $k = 3$ using the weight vector presented above. Compared to the first results (without weighted features), the overall performance raises to 97.1%. In particular, the hit rates of classes 1, 2, 3 and 7 increase, whereby the cumulus class improves by almost 3%. The rate of class 4, the class of clear sky, remains the same and the accuracy of classes 5 and 6 decrease slightly due to more confusions between the last three classes compared to the first results.

4.2 Random test sample

In addition to the evaluation using the LOOCV, we tested the algorithm with an additional random data sample, called TEST, to point out the problems occuring by classifying images not necessarily showing one unique cloud class or containing interferences like dust (see Fig. 4).

From the data obtained on board "Polarstern" during ANT XXIV/1 (see Sect. 2.2) a set of 275 images was randomly selected, covering all cloud classes as defined in Table 1. We manually classified each of them in one unique cloud category, even if the assignment was disputable. Consequently, differences in the classification results of the algorithm compared to this manual "reference classification" are unavoidable and a certain bias has to be accepted.

Results. Table 5 shows the corresponding hit rates. Compared to the accuracies achieved by the LOOCV we see a decrement of the mean classification rate from 97.06% to 74.58%, where the rate of the class stratocumulus is decisive. In relation to the manual classification, only 41.30% of these images are correctly classified. The remaining images are assigned to altocumulus or stratus, types of misclassification already noted in Sect. 4.1 and in general difficult to avoid. Looking at the corresponding images reveals that they

Table 5. Confusion matrix of the random sample TEST (275 images) in % and absolute.

True class	Classified as							Mean
	1	2	3	4	5	6	7	
1	**67.44 (58)**	16.28 (14)	15.12 (13)	0.00	0.00	0.00	1.16 (1)	
2	5.88 (2)	**79.41 (27)**	0.00	14.71 (5)	0.00	0.00	0.00	
3	10.94 (7)	3.13 (2)	**81.25 (52)**	0.00	1.56 (1)	3.13 (2)	0.00	
4	5.88 (1)	17.65 (3)	0.00	**76.47 (13)**	0.00	0.00	0.00	
5	0.00	0.00	32.61 (15)	0.00	**41.30 (19)**	26.09 (12)	0.00	
6	0.00	0.00	9.52 (2)	0.00	0.00	**90.48 (19)**	0.00	
7	0.00	0.00	14.29 (1)	0.00	0.00	0.00	**85.71 (6)**	74.58

Table 6. Confusion matrix of the random sample TEST (275 images) for serious errors in % and absolute.

True class	Classified as							Mean
	1	2	3	4	5	6	7	
1	**83.72 (72)**	16.28 (14)	0.00	0.00	0.00	0.00	0.00	
2	5.88 (2)	**79.41 (27)**	0.00	14.71 (5)	0.00	0.00	0.00	
3	0.00	0.00	**96.87 (62)**	0.00	0.00	3.13 (2)	0.00	
4	5.88 (1)	17.65 (3)	0.00	**76.47 (13)**	0.00	0.00	0.00	
5	0.00	0.00	0.00	0.00	**100.00 (46)**	0.00	0.00	
6	0.00	0.00	9.52 (2)	0.00	0.00	**90.48 (19)**	0.00	
7	0.00	0.00	14.29 (1)	0.00	0.00	0.00	**85.71 (6)**	87.52

display primarily clouds in transition and the result of the algorithm is thus accurate. Moreover, for usage in radiative transfer studies such mismatches, as well as some other misclassifications like confusions between cumulus and altocumulus clouds, can be considered as "permissible" errors due to a similar impact on radiation (Rossow and Schiffer, 1991).

If we look as a next step only at "serious" errors in regards to radiation analyses, what means, the confusions furthermore considered as misclassifications are those of clouds differing significantly in their impact on radiation (e.g. cirrus and stratus clouds), we obtain another result. The corresponding hit rates are constituted in Table 6, in which misinterpretations of cumulus clouds as thick clouds and those of high cumulus as cirrus are excluded as well since the involved images are marginal cases and an assignment to both cloud classes is acceptable. Here, the random sample TEST yields an overall classification rate of 87.52%. The main part of the remaining misclassifications are confusions between cumulus, cirrus and clear sky. Checking again the corresponding images shows that without exception each of them displays less than 30% cloudiness, indicating that this may be the source of error. Another few confusions occur between images showing high cumulus, stratus or thick clouds. Once more, the respective images are marginal cases and each of these assignments is acceptable.

5 Summary and conclusions

In this study, we presented an automatic method to classify simple digital images in cloud classes similar to the genera of the ICCS. We have shown that the distinction of these genera is possible using only first and second order statistic features, at least in combination with knowledge of the actual cloudiness.

Considering obviously assignable images, the classes best recognized by the the kNN-algorithm are clear sky and cirrus. If ambiguous images are permitted to the classification process as well, some more confusions appear between these two classes due to the whitening effect presented in Sect. 4. An approach already tested in the process of our study to avoid this error caused by the misinterpretation of pixels near the sun is the determination of the position of the solar disk and its removal. This can be accomplished by use of geometrical features, the knowledge of time and location when the image is taken or the inspection of time series (if a "cloud" does not move, this "cloud" is likely to be the area around the visible sun). In case, the solar disk is displayed in an image, the pixels around could be excluded geometrically from further calculations or by use of an additional treshold.

Another remarkable error, the confusion between cirrus and cumulus, occurs primarily in the case of cloudiness amounts less than 30%. It is conceivable that here a hierarchical classification process may lead to improvements. After a first division according to cloudiness, the further

assignment to a cloud class, especially to one of these classes, might be more evident.

The remaining classes are recognized quite well. Some more confusions exist between cumulus and high cumulus due to their similarity in color and smooth transition in definition. Also confusions occur between the last three classes, stratocumulus, stratus and thick clouds. The reason is the frequently changeover from one to another class, a natural phenomenon that will always lead to some misclassifications of these classes using automatic methods.

One further problem, not visible in the results of the LOOCV, but occuring in the analysis of random data, is the incorrect class assignment due to simultaneous appearance of more than one predefined cloud class. In nature, the sky often provides a wide spectrum of different cloud types at the same time, e.g. cirrostratus and stratocumulus or cirrus and cumulus frequently occur together. In order to avoid miss-classifications due to this phenomenon, we suggest an initial partitioning of the images into smaller subimages and their separate classification. However, it is important to check if these subimages still include enough information to assign the image parts to a cloud class.

We are convinced that by use of the suggestions elucidated above and thus, an elimination of errors caused by questionable images, an improvement of the algorithm is possible. Moreover, other, not mentioned features also may lead to an increasing of the algorithm's performance. However, the algorithm here presented is already quite powerful and suitable for research purposes. For example, at the Meteorological Institute of IFM-GEOMAR in Kiel the implemented algorithm is currently in use and available for people interested.

References

Bankert, R. L. and Wade, R. H.: Optimization of an instance-based GOES cloud classification algorithm, J. Appl. Meteorol. Clim., 46, 36–49, 2007.

Calbó, J. and Sabburg, J.: Feature extraction from whole-sky ground-based images for cloud-type recognition, J. Atmos. Ocean. Tech., 25, 3–14, 2008.

Cazorla, A., Olmo, F. J., and Alados-Arboledas, L.: Devlopment of a sky imager for cloud cover assessment, J. Opt. Soc. Am. A., 25, 29–39, 2008.

Christodoulou, C. I., Michaelides, S., and Pattichis, C.: Multifeature texture analysis for the classification of clouds in satellite imagery, IEEE T. Geosci. Remote, 41, 2662–2668, 2003.

Duda, R. O. and Hart, P. E.: Pattern classification, John Wiley and Sons, 2001.

Dybbroe, A., Karlsson, K.-G., and Thoss, A.: NWCSAF AVHRR Cloud Detection and Analysis Using Dynamic Tresholds and Radiative Modeling. Part II: Tuning and Validation, J. Appl. Meteorol., 44, 55–71, 2005.

Feister, U. and Shields, J.: Cloud and radiance measurements with the VIS/NIR Daylight Whole Sky Imager at Lindenberg (Germany), Meteorol. Z., 14, 627–639, 2005.

Haralick, R. M., Shanmugam, K., and Dinstein, I.: Textural features for image classification, 3, 610–621, 1973.

Kalisch, J. and Macke, A.: Estimation of the total cloud cover with high temporal resolution and parametrization of short-term fluctuations of sea surface insolation, Meteorol. Z., 17, 603–611, 2008.

Kassianov, E., Long, C. N., and Christy, J.: Cloud-base-height estimation from paired ground-based hemispherical observations, J. Appl. Meteorol., 44, 1221–1233, 2005.

Long, C. N., Sabburg, J., Calbó, J., and Pagès, D.: Retrieving cloud characteristics from ground-based daytime colorall-sky images, J. Atmos. Ocean. Tech., 23, 633–652, 2006.

Pagès, D., Calbo, J., Long, C. N., González, J.-A., and Badosa, J.: Comparison of several ground-based cloud detection techniques, in: Extended Abstracts, European Geophysical Society XXVII General Assembly, 26–28 April, Nice, France, 2002.

Pankiewicz, G. S.: Pattern recognition techniques for the identification of cloud and cloud systems, Meteorol. Appl., 2, 257–271, 1995.

Parisi, A. V., Sabburg, J., Turner, J., and Dunn, P. K.: Cloud observations for the statistical evaluation of the UV index at Toowoomba, Australia, Int. J. Biometeorol., 52, 159–166, 2008.

Petty, G. W.: A first course in atmospheric radiation, Sundog Publishing, 2006.

Pfister, G., McKenzie, R. L., Liley, J. B., Thomas, A., Forgan, B. W., and Long, C. N.: Cloud coverage based on all-sky imaging and its impact on surface solar irradiance, J. Appl. Meteorol., 42, 1421–1434, 2003.

Ricciardelli, F., Romano, F., and Cuomo, V.: Physical and statistical approaches for cloud identification using Meteosat Second Generation-Spinning Enhanced Visible and Infrared Imager Data, Remote Sens. Environ., 112, 2741–2760, 2008.

Ripley, B. D.: Pattern Recognition and Neural Networks, Cambridge University Press, 8 edn., 2005.

Rossow, W. B. and Schiffer, R. A.: ISCCP cloud data products, B. Am. Meteorol. Soc., 71, 2–20, 1991.

Sakellariou, N. K., Varotsos, C. A., and Lykoudis, S. P.: On the intercomparison of satellite and ground-based observations of prevailing cloud types over Athens, Int. J. Remote Sens., 16, 1799–1804, 1995.

Schiel, S.: The expedition of the research vessel "Polarstern" to the Antarctic in 2007 (ANT-XXIV/1), in: Berichte zur Polar- und Meeresforschung, 592, 114 pp., 2009.

Seiz, G., Baltsavias, E. P., and Gruen, A.: Cloud mapping from the ground: Use of photogrammetric methods, Photogramm. Eng. Remote Sen., 68, 941–951, 2002.

Serpico, S. B., Bruzzone, L., and Roli, F.: An experimental compar-

ison of neural and statistical non-parametric algorithms for supervised classification of remote-sensing images, Pattern Recognit. Lett., 17, 1331–1341, 1996.

Shields, J. E., Johnson, R. W., Karr, M. E., and Wertz, J. L.: Automated day/night whole sky imagers for field assessment of cloud cover distributions and radiance distributions, in: Proc. 10th Symp. on Meteorological Observations and Instrumentation, 11–16 January, Boston, MA, 1998.

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, in: Proc. SPIE, vol. 5156, pp. 155–166, 2003.

Singh, M. and Glennen, M.: Automated ground-based cloud recognition, Pattern Anal. Appl., 8, 258–271, 2005.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L.: Climate Change 2007: The Physical Science Basis – Contribution of Working group I to the Fourth Assessment Report of the IPCC, Cambridge University Press, 2007.

Souza-Echer, 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. Oceanic Technol., 23, 437–447, 2006.

Tag, P. M., Bankert, R. L., and Brody, L. R.: An AVHRR multiple cloud-type classification package, J. Appl. Meteorol., 39, 125–134, 2000.

Vincent, P. and Bengio, Y.: K-Local Hyperplane and Convex Distance Nearest Neighbor Algorithms, Tech. rep., Dept. IRO, Université de Montréal, 2001.

WMO: International Cloud Atlas, Vol. 2, Am. Meteorol. Soc., 1987.