

Shape parameters for automatic classification of snow particles into snowflake and graupel

Karolina Nurzyńska,^{a*} Mamoru Kubo^a and Ken-ichiro Muramoto^{a,b}

^a Faculty of Electrical and Computer Engineering, Institute of Science and Engineering, Kanazawa University, Kakuma, Kanazawa 920-1192, Japan

^b Ishikawa National College of Technology, Tsubata, 920-0392, Japan

ABSTRACT: The meteorological radar backscatter profile depends on the shape of a falling particle. Because there are many shapes of snow particles it is difficult to estimate a precipitation rate in the case of snowfall. This study presents research which aimed to develop an automatic system for snow particle classification into snowflake and graupel. Having the information about the snowfall type during an analysis of snowfall rate and backscatter values allows improved forecasting of snowfall by better understanding of these phenomena.

Five novel shape features derived from grey-scale images and designed in order to improve an automatic snow particle classification into snowflake and graupel are introduced. Their performance is compared to statistical and shape features well known from literature. For classification purposes, threshold, k -nearest neighbours, and support vector machine are used. Different classification systems are presented. The most correct classification ratio, of 91%, was achieved for a classifier built from a pair of roughness and Hu first order features. The suggested min max centre distance feature is in second place, with 90% efficiency. Copyright © 2011 Royal Meteorological Society

KEY WORDS image processing; radar meteorology; shape features

Received 3 January 2011; Revised 3 August 2011; Accepted 24 August 2011

1. Introduction

During the winter monsoon the cold Siberian winds take part in the creation of snow clouds over the Japan Sea. Those clouds cause a huge snowfall over the Japan Sea coast, especially in the Honshu region of Japan. This is due to the specific topographic characteristic, where mountains (up to 3000 m) occur near the coast (Harimaya and Nakai, 1999). Moreover, it is known that during snowfall precipitation graupel is a predominant particle (Mizuno, 1992) and also that the snowflake size distribution varies in this area (Harimaya *et al.*, 2004).

The large amount of snow, which results from an orographic snowfall, may cause many disasters. Therefore, accurate forecasting of the precipitation rate should also permit improvement in preparation of weather alerts. There are already many applications dealing with similar problems in different regions of the world. DeGaetano and Wilks (1999) describe a system which takes under consideration the information about the snow cover in order to prevent the roof from collapsing due to too much snow. Moreover, the authors also notice that this is mostly caused by a consecutive snowfall events rather than by one snowfall. On the other hand, precipitation in low temperatures is responsible for ice creation on

railways (Shao *et al.*, 2003). In this case, in order to minimize the loss of income to railway companies due to service interruptions, the authors developed a system which uses statistical techniques to predict the probability of ice creation over the railway. Moreover, this system tries to forecast the origin of the ice. Boi (2009), on the contrary, presented a system which is responsible for global snow cover monitoring. Using visible near infrared and infrared data based on a simple statistical hypothesis this system tries to reconstruct the snow cover from satellite images, even when clouds are present.

All the systems mentioned above use forecast information about the precipitation rate (R), which is estimated on the basis of the weather radar reflectivity factor (Z). Both these parameters are joined by the so-called Z - R relation, given by the formula (Marshall and Gunn, 1952):

$$Z = BR^\beta, \quad (1)$$

where B and β are coefficients which change depending on the precipitation type. Moreover, these parameters are also strongly related to the sizes of the falling particles (Matrosov, 1992; Loffler-Man and Blahak, 2001). Therefore, there is a need to find the most accurate coefficients which would allow description of the snowflake and graupel precipitation events. It is known that the characteristics are different for snowflake and graupel, and this has been used previously by El-Magd *et al.*

*Correspondence to: K. Nurzyńska, School of Natural Science and Technology, Kanazawa University, Kanazawa, Japan.
E-mail: Karolina.Olga.Nurzynska@gmail.com

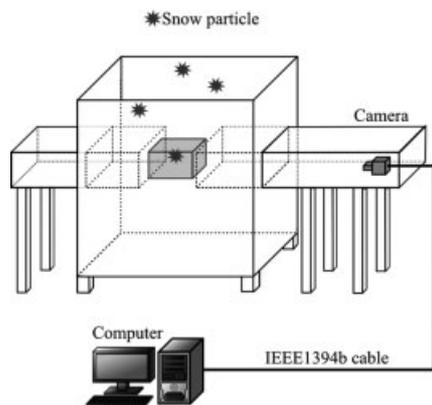


Figure 1. General measuring system setup.

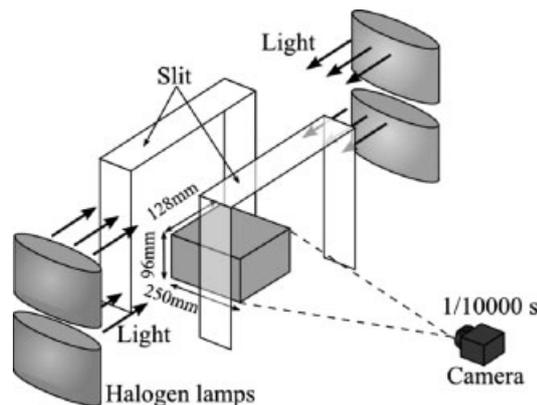


Figure 2. Measuring volume.

(2000) in research concerning graupel and hail recognition using multiparameter radar. However, before it is possible to estimate the B and β coefficients the type of the precipitation must be identified. Hence, the automatic classification of snow particles into snowflake and graupel (which are the most common snow particle types in the study region) is needed. It is also believed that being able to distinguish between those particle types could improve the understanding of snow particle growth and snow cloud creation.

The remainder of this paper is structured as follows. Firstly, the system created for snow particle image acquisition is presented. Secondly, the snow particle parameter description is given. Statistical and shape features are well known from the literature, but five novel features are introduced. Next, experiments whose aim is to narrow the number of considered features are presented, followed by experiments addressing the efficiency of chosen methods. Finally, the conclusions are drawn.

2. System overview

The weather monitoring system which is used in research consists of many elements, e.g. micro-rain radar, precipitation occurrence sensor system, electronic balances, multiple imaging systems. All this equipment is located in a small area. This permits the assumption that the recorded data describe the same precipitation event. However, in the research addressed in this paper only one imaging system was used. This imaging system collects grey-scale images of falling snow particles. It uses a camera with shutter speed of $1/10\,000$ s and resolution of 1280×960 pixels. The camera is connected to an internal PC by an IEEE 1394b cable. The camera is placed 2 m from a measuring volume in a special vertical tunnel, which protects it from the influence of wind and sunlight. Additionally, wind breaks are installed around the measured area to improve the protection. The system setup is depicted in Figure 1. The imaging space is illuminated by four halogen lamps, two on each side of the measuring space (Figure 2). The volume of the measuring space is $W128 \times H96 \times D250$ mm.

Having images from the camera, some image pre-processing is necessary in order to store only images of snow particles, not all the recorded data. Therefore, firstly the binary representation of an image is calculated in order to distinguish particles from background. Each of the particles is then given a label to distinguish between them. Finally, the image of each particle, with a 30 pixel margin around it, is cropped from the image and stored in the database. Examples of achieved images are presented in Figure 3(a).

3. Object description

The methods designed for image processing aim to teach the computer to distinguish between images. They try to describe in a mathematical way the processes which take place when a human brain recognizes images, which is not a trivial task. As there are many applications of image processing, many techniques for understanding images have been suggested. However, most of them are designed to deal with general problems and prove to be insufficient for a specialized case.

3.1. Statistical image features

From one point of view, an image represents a discrete two dimensional function $I(x, y)$, where grey level intensities correspond to the values of this function. In the domain of this function many statistical parameters might be derived. One of the most common statistical features is the weighted average of the function values, called moments (Jahne, 2002). For a grey-scale image I the moment M is defined as follows:

$$M_{ij} = \sum_x \sum_y x^i y^j I(x, y), \quad (2)$$

where x, y are pixel coefficients in the image and $(i + j)$ define moment order, where $i, j = 0, 1, 2, \dots$. Generally, depending on how the weights (i and j) are chosen, different spatial moments are achieved (usually for the pair ij the following moment values are considered: 00, 01, 10, 11, 20, 02, 21, 12, 30, 03). Moreover, the

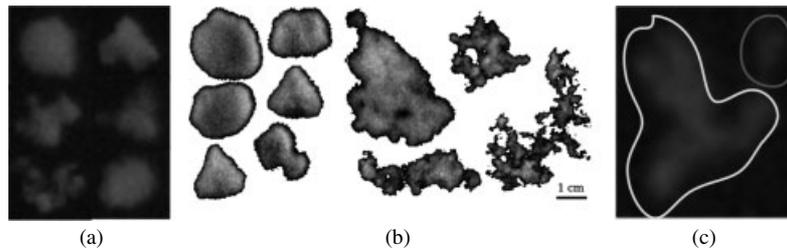


Figure 3. Examples of snow particles gathered by the system. (a) Original data (b) Extracted graupel and snowflake particles (c) Enlarged image with two particles.

moment can be additionally normalized according to an object centroid in order to remove the influence of object transition in the image on the parameter value. The central moment, μ , for known object centroid (\bar{x}, \bar{y}) is formulated as follows:

$$\mu_{ij} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j I(x, y). \quad (3)$$

Furthermore, to diminish the influence of object scaling the normalized central moment η was introduced:

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^{(1 + \frac{i+j}{2})}}. \quad (4)$$

However, only Hu moments (Hu, 1962) achieve such a parameter which returns the same value regardless of the object's translation, rotation and scaling within the image. There are seven Hu invariant moments, which base their definitions on the normalized central moments. They describe the image in such a way that by knowing all of them it is possible to reconstruct the image. The formulae for Hu moments are:

$$\begin{aligned} Hu_1 &= \eta_{20} + \eta_{02} \\ Hu_2 &= (\eta_{20} - \eta_{02})^2 + (2\eta_{11})^2 \\ Hu_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{02})^2 \\ Hu_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\ Hu_5 &= \frac{(\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{02})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]}{(\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]} \\ Hu_6 &= \frac{(\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{03} + \eta_{12})(\eta_{21} + \eta_{03})}{(3\eta_{21} - \eta_{02})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]} \\ Hu_7 &= \frac{-(\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]}{-(\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]} \end{aligned} \quad (5)$$

In preliminary research (the experiment described in Section 4.3) the classification efficiency of most of the features did not overcome the threshold of 80% for correct classification. From those which present better results (some normalized central moments and Hu

moments) the best performing one is the Hu moment of the first order (**Hu1**), and only this one has been chosen to represent this group.

3.2. Shape features

On the other hand, it is possible to derive from an image information about object perimeter, area, horizontal or vertical length, which allows description of the shape in a straightforward manner. These features find specific relations between the chosen object descriptors to describe by only one value the object shape characteristics (Russ, 1998; Jahne, 2002). For instance, the Feret and Malinowska features describe object elongation. On the contrary, shapeless, roughness and circular features depict object roundness. There are also others, such as the Danielsson, Haralick and Blair-Bliss features, which aim to concentrate also on the complexity of shape.

For further consideration only the roughness (**Rough**) and Danielsson (**Dan**) parameters were chosen, as the rest of the named shape features gave a very low correct classification ratio (below 80%). Roughness shape feature is given with the following formula:

$$Roughness = 2\sqrt{\frac{S}{\pi}}, \quad (6)$$

where S is the object's area. This parameter calculates the diameter of a circle whose perimeter has the same length as the object's one (Figure 4). Therefore, the more the shape resembles the circular shape the closer the value becomes to unity. On the other hand, the Danielsson parameter relates an object's area, S , with the minimal

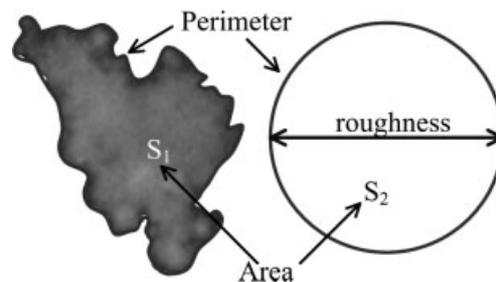


Figure 4. The idea of the roughness feature.

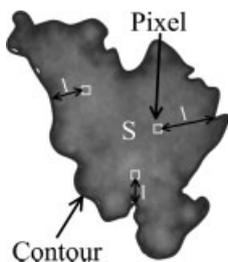


Figure 5. The idea of the Danielsson feature.

distance between each object pixel and contour one l_i :

$$Danielsson = \frac{S^2}{\left(\sum_i l_i\right)^2}. \quad (7)$$

Generally, this relation characterizes elongated objects with big values, whereas for round objects the parameter values are small. Contrary to the other parameters describing the object elongation, this parameter is calculated locally (for each pixel). Therefore its value does not depend on whether the object is straight or not (Figure 5).

3.3. Novel shape features

This section introduces novel features which have been developed in order to improve the classification between snowflake and graupel. Distinguishing between those two particle types in images is a very demanding task, especially given that the images not always are sharp (see Figure 3 for examples).

3.3.1. Flake number (FN)

While the snow is falling the snowflake may split into smaller flakes, some may aggregate and others may just fall together. It was also noticed that sometimes in one snowflake image there are visible parts of other snow particles (Figure 3(c)). This is the result of the particle image cropping algorithm used in the system (see Section 2 for details), although this situation is very rare in the case of graupel. Hence, it is assumed that when other snow particles are visible in the image that it is a snowflake.

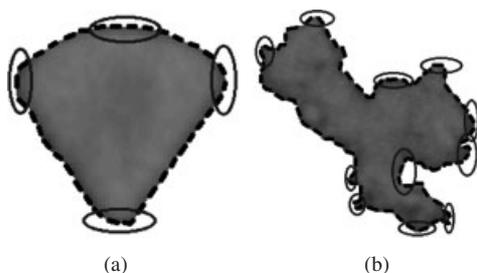


Figure 6. The idea of the corner number feature. Dashed line represents contour, the ellipses mark the place of a counted corner, but in case of snowflake only some representatives were marked. (a) graupel (b) snowflake.

3.3.2. Corner number (CorN)

The shape of a graupel particle resembles a circle, whereas the snowflake may take different shapes. Generally, it can be said that the snowflake contour line is more complex due to its angular outline. Therefore, some features which exploit this characteristic are suggested. First of all, an object contour can be described by a two dimensional discrete function $f(x, y)$. The object contour usually is frayed, therefore in order to remove minor contour line shape changes, which do not contain any important information, the function is smoothed by applying a discrete mean filter:

$$f_S(x, y) = \frac{1}{2k + 1} \sum_{i=x-k, j=y-k}^{i=x+k, j=y+k} f(i, j), \quad (8)$$

where $k = 2$. Each angle of the contour then corresponds to change of signs of consecutive values in the first derivative of the function ($df_S/dx, df_S/dy$) calculated separately for x and y . Because the contour is a closed line, the sign of the derivative in x and y changes at least twice. The sum of all bends is stored in the corner number feature value (Figure 6).

3.3.3. Concave number (ConN)

Keeping in mind the differences in the shapes of graupel and snowflakes mentioned in previous subsection, the problem of snow particle description is approached differently. Here the convex hull for an object is calculated. In the case of a convex object the convex hull overlaps the contour, otherwise, when the object is concave, there are places where the contour line and convex hull line split. Every time this occurs it means that a cavity has been found. The concave number parameter finds the number of such cavities, which are marked with different shades in Figure 7. In comparison with *CorN* this parameter is more general, as it concentrates on the biggest changes in contour line shape (compare the results presented for the examples given in Figures 6 and 7).

3.3.4. Max min distance (MMD)

The max min distance is a different feature, which describes the relation between the longest and shortest

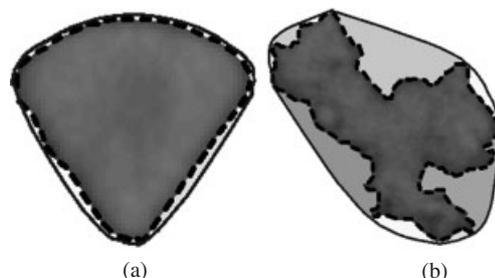


Figure 7. The idea of the concave number feature. Dashed line represents contour; solid line the convex hull; different shades mark the concave parts. (a) graupel (b) snowflake.

length of line connecting the contour of an object. The line is defined between two points belonging to a contour. The longest line is for that pair of points with the biggest Euclidean distance. On the contrary, many short lines are found. Therefore, the selection is narrowed to those which start in places where the contour bends. Choosing the place where the contour is convex (bends) assures that the chosen lines are the shortest. Next, the median value of all lengths becomes a representative length (see Figure 8 as an example).

3.3.5. Max min centre distance (MMCD)

The max min centre distance feature defines the relationship between the points belonging to a contour which is the longest distance from an object's centre of mass and that which is the nearest. Figure 9 depicts an example of this idea.

4. Results and discussion

4.1. Snow particle database description

The snow particle database contains 8480 images. The database has been divided into training and testing

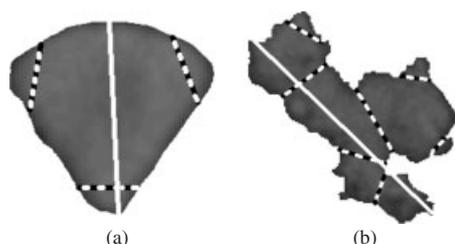


Figure 8. The idea of the max min diameter feature. The longest line is marked as the solid line, the examples of the short with a dashed line. (a) Graupel, (b) snowflake.

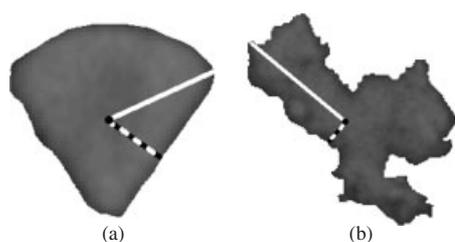


Figure 9. The idea of the max min centre distance feature. The dot represents the mass centre. The longest distance is marked with the solid line, the shortest with a dashed line. (a) graupel (b) snowflake.

datasets. The training dataset (TD1) consisted of 460 snowflake images and 461 graupel images. The testing dataset is bigger and contains 3924 snowflake and 3635 graupel images. However, in case of SVM (support vector machine) classification the training dataset consisted of 85 images, which is representative of 1% of snowflakes and graupel with similar proportions as in TD1, whereas the rest of the images (8395) were in the testing dataset.

4.2. Classification methods

In this research three different classification methods have been applied. All of them are supervised classification methods: this means that they need a labelled training dataset. The preparation of the labelled dataset was based on the visual inspection of chosen images. When classified data were described by only one value the threshold classifier was used: in other cases both the *kNN* (*k*-nearest neighbour) and *SVM* (support vector machine) classifiers were used.

The aim of this research is to evaluate the descriptive properties of the presented features for snow particle classification into snowflake and graupel. This is a two-class classification problem. The easiest solution, especially when each class is described only by one feature, is to find the threshold. If the feature value is then below that threshold it indicates that the particle belongs to one class, otherwise it belongs to the other (see Figure 10(a)). The drawback of this method is the fixed threshold value, which should be determined on the basis of the training dataset. Choosing a proper threshold value is a difficult task, especially for new data whose characteristics are not familiar to the researcher. Therefore, sometimes it is useful to apply the *kNN* or *SVM* classifier in this case, because those methods decide the threshold automatically.

The *kNN* and *SVM* classifiers do not assume any data distribution pattern, however they deal with the classification problem differently. In the case of *kNN*, for each testing object a virtual hyper-sphere is created which encloses *k* training objects only (in a hyper-sphere each feature describing the object corresponds to one dimension of this sphere; e.g. for two features it is a circle, for three a sphere). The testing object is classified to this class which is represented by the majority of training objects enclosed in the hyper-sphere. Therefore, it is important to choose the *k* parameter to be an odd number in the case of a two-class problem, otherwise

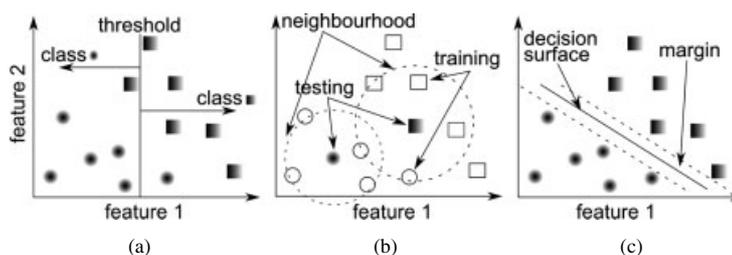


Figure 10. Classification methods idea presentation. (a) threshold classification (b) *kNN* classification (c) *SVM* classification.

there might be the same number of neighbours belonging to each from two classes: this makes the classification impossible. An example of *kNN* classification (for $k = 3$) in a two-dimensional case is presented in Figure 10(b).

In contrast, *SVM* aims to find a hyper-surface which divides the data into two separate classes. Moreover, this hyper-surface is chosen in such a way that the margin separating the classes from this surface is as big as possible (see Figure 10(c)) (Burgess, 1998). The hyper-surface can divide the data linearly: thus it forces the input data to be linear, which is impossible to assure. Therefore, in general cases a feature space, defined by the data, of finite-dimensions (corresponding to the length of feature vector describing each object) is mapped by some unlinear function to another higher-dimensional space, where the data should be separated linearly. The function used for mapping is called a kernel. There are many kernels suggested, such as polynomial or RBF (Gaussian radial basis function), whose name corresponds to the function type used for the value transformation.

4.3. Initial experiment

The aim of the initial experiment was to evaluate the classification performance of each from the presented statistical and shape features known from the literature. In total, 39 features (10 spatial moment features, 10 central moment features, 10 normalized central moment features, 7 Hu moments features, and 2 shape features) have been considered. The results were then compared with the classification performance of the five novel features. The threshold classifier was applied in all cases.

The training set TD1 was used to determine the threshold value. In the case of each feature for both graupel and snowflake classes the minimal, maximal and average feature value was calculated. The threshold was then decided. In cases where the maximal value of one class was smaller than the average of the other, this value was set as a threshold (e.g. *FN*, *CorN*). Otherwise, in the cases of *ConN* and *MMD* it has been chosen empirically from values between the average class value. For *MMCD*, applying the first rule resulted in the wrong classification, therefore the second rule was used. In the second case of threshold decision only the threshold which returned the best classification result is presented in Table I. The normalized histogram distribution of training data for snowflake and graupel classes for the chosen features with marked threshold value are depicted in Figure 11. In the case of statistical features, only the feature with highest classification ratio is presented due to lack of space.

Once the threshold value for each of the features was calculated the classification performance was investigated for the training dataset. A correct classification ratio, not only for whole dataset but also for each of the classes, was calculated. The results are presented in Table II. It is worth noticing that all of the mentioned statistical features whose classification performance is not described in this table were below 80%.

Table I. Values achieved for training dataset.

Method ^a	Class	Minimum	Average	Maximum	Threshold
FN	G	1	1	1	<1
	S	1	1.22	5	≥1
CorN	G	2	3.35	10	≤10
	S	2	12.27	216	>10
ConN	G	4	13.51	28	≥10
	S	3	8.15	17	<10
MMD	G	1.16	6.65	128.22	≤13
	S	1.36	15.86	180.80	>13
MMCD	G	1.45	2.23	7.75	≤3
	S	1.73	32.52	1421	>3
Hu1	G	1.8 E-03	2.2 E-03	3.0 E-03	≤3E-03
	S	2.7 E-03	3.4 E-03	4.4 E-03	>3E-03
Dan	G	17.55	34.97	44.69	≤45
	S	2.23	54.07	157.86	>45
Rough	G	1.00	1.08	1.19	≤1.19
	S	1.00	1.33	2.31	>1.19

^a ConN, concave number; CorN, corner number; Dan, Danielson parameter; FN, flake number; Hu1, Hu moment of the first order; MMCD, Max min centre distance; MMD, Max min distance; Rough, roughness.

This experiment shows that the *MMCD* feature is the best snow particle classifier, with correct classification ratio equal 88.07%. It overcomes the classical shape parameters of about 4–8%. On the other hand, it is worth pointing out that in some cases (*CorN*, *FN*) the correct classification ratio for graupel was very high, whereas the performance for snowflake was very low. This fact could be used to create a snowflake discriminator classifier, which allows correct classification of snow particles, whilst assuring simultaneously that the graupel particles are not misclassified.

4.4. Classifier application

The threshold classifier works well when it is possible to define the threshold between two classes easily, although in cases when two classes overlap each other it is difficult to decide the correct threshold value. Moreover, sometimes it is more convenient to leave this problem to be decided by the classifier. Therefore, the next experimental goal was to check the system performance with more sophisticated classifiers. Because the distribution of the snow particle features is unknown, classifiers which do not need this information were used, such as *kNN* and *SVM*. In this experiment, for the *kNN* classifier the k parameter was set to values in a range from 5 to 95, with a step 10. In the case of *SVM* the following kernels were applied: linear (*L*), quadratic (*Q*), polynomial (*P*) and radial basis function (*RBF*). Actually, linear and quadratic kernels represent the polynomial kernel of first or second order, respectively.

Table III gathers the classification efficiency results for the *kNN* and *SVM* classifiers. Regarding all possible classifier parameters, in this table only the best results

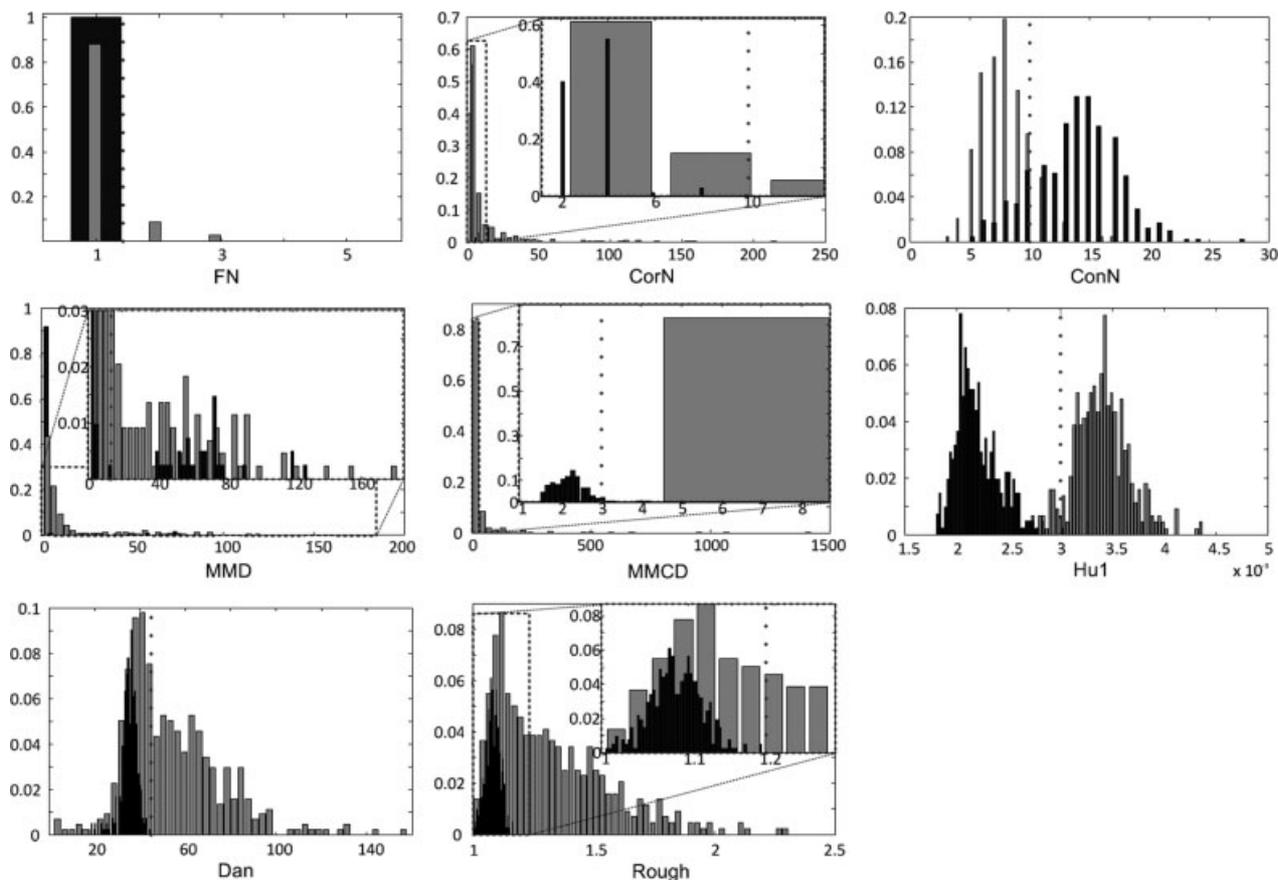


Figure 11. Data distribution for features considered in first experiment. ■, Snowflake; ▒, Graupel; |, Threshold; □, Zoom.

Table II. The results obtained for the thresholding.

Method ^a	Graupel	Snowflake	Total (%)
MMCD	87.70	88.40	88.07
Rough	99.06	70.29	84.48
Hu1	76.15	85.47	80.99
Dan	99.17	63.53	80.67
CorN	99.75	33.82	65.01
FN	99.72	17.30	56.94
ConN	46.63	66.28	56.83
MMD	69.71	19.11	43.44

^a ConN, concave number; CorN, corner number; Dan, Daniels-son parameter; FN, flake number; Hu1, Hu moment of the first order; MMCD, Max min centre distance; MMD, Max min distance; Rough, roughness.

Table III. Classification of snow particles.

Method ^a	kNN	SVM
MMCD	89.11 ₃₅	79.02 _{P8}
Rough	87.33 ₆₅	87.29 _L
Dan	84.90 ₂₅	85.17 _{P5}
MMD	83.33 ₃₅	80.95 _{RBF1.5}
Hu1	82.75 ₇₅	82.72 _L
ConN	56.77 ₃₅	59.39 _Q
CorN	51.70	73.76 _{P3}
FN	51.70	56.95

^a ConN, concave number; CorN, corner number; Dan, Daniels-son parameter; FN, flake number; Hu1, Hu moment of the first order; MMCD, Max min centre distance; MMD, Max min distance; Rough, roughness.

are presented. The subscript in the case of the *kNN* classifier describes the *k* parameter value. The subscript in the case of the *SVM* classifier describes the kernel type (*L*, *Q*, *P*, *RBF*) with the value describing the order in the case of a polynomial kernel and the parameter α in the case of *RBF*. When there is no subscript given it means that the same result was achieved for all parameters.

It is not surprising that in many cases the classification performance is higher (bold font), as the classifiers describe the class boundaries much better. The biggest

increase in performance is visible in the case of the *MMD* classifier, where the correct classification ratio rises from 43.44 to 83.33% for *kNN*, and to 80.95% for *SVM*. The others also slightly improve the correct classification ratio: *MMCD* from 88.07 to 89.11% for *kNN*, *Rough* from 84.48 to 87.33% for *kNN*, *Dan* from 80.67 to 85.17% for *SVM*, *Hu1* from 80.99 to 82.75% from *kNN*. The only problem was noticed for the *MMCD* feature used with the *SVM* classifier, where the performance decreased. That might be the result of using a training set which does not define the class distribution pattern

well and the calculated hyper-surface was created in the wrong place.

4.5. Two-class classification

In order to improve the classification performance, in this experiment two features were used simultaneously to describe the object's properties, since exploiting more information should return a better result. This is a two-class classification problem which is easily solved by the *kNN* and *SVM* classifiers. Table IV presents the results gathered for the *kNN* classifier and Table V for *SVM*. Similarly, as in the previous experiment, for each pair of features a set of classifiers (with different *k* parameter for *kNN* and kernel for *SVM*) is applied, but the tables present only the best results.

The results show that combining the information from two features improves the correct classification ratio. The worst correct classification ratio in this two feature case is always higher than in the case of applying only one feature (see the results for *FN* and *CorN*). Moreover, it is interesting to note that features which already have a high correct classification ratio in the previous experiment improved when joined in pairs and classified with *kNN*. On the other hand, *SVM* improved the correct classification ratio mostly for those feature pairs, which were weak (less than 80%) in the previous experiment. The best classification efficiency of 90.98%, for the *Hu1* and *Rough* pair, is achieved for the *SVM*

classification. To visualize the best correct classification performance better, the best scores are written in bold in the tables.

4.6. Snowflake discrimination

In the initial experiment, presented in Section 4.3, it was pointed out that two features (*CorN*, *FN*) might be useful to create a snowflake discrimination classifier. Its correct classification ratio of snowflake and graupel is quite low, but when inspecting those classes separately it was noticed that those features describe graupel perfectly (as in both cases the correct classification ratio is above 99%). The problem was in snowflake classification, where *CorN* correctly recognized only 33.82% of snowflakes and *FN* 17.30% (Table II). In this experiment a novel classification system was suggested, which was built from two parts. In the first step the *CorN* feature was applied in order to discriminate the snowflakes. The rest of the particles (all graupel and misclassified snowflakes) were then sent to the two-feature classifier presented in the previous section.

Table VI presents the results achieved when the *kNN* classifier was applied in the second stage, whereas Table VII gathers correct classification ratios for *SVM* classification. The application of the two step classifier improved the performance in almost all cases: compare the results from Table VI with Table IV, and Table VII with Table V. However, the biggest influence of the snowflake discrimination with the *CorN* feature was

Table IV. Snow particles described by two features and classified with *k*-nearest neighbour (Knn) classifier.

Method	CorN	ConN	Dan	FN	Hu1	MMCD	MMD
ConN	70.70 ₃₅						
Dan	85.01 ₂₅	85.97 ₃₅					
FN	51.70	58.85 ₃₅	58.18 ₂₅				
Hu1	87.58 ₃₅	82.68 ₁₅	84.90 ₂₅	84.01 ₄₅			
MMCD	88.73 ₁₅	88.25 ₁₅	88.74 ₁₅	89.16 ₃₅	89.12 ₃₅		
MMD	83.16 ₃₅	82.36 ₁₅	87.00 ₅	83.45 ₃₅	83.33 ₃₅	87.28 ₂₅	
Rough	86.63 ₂₅	86.92 ₁₅	85.05 ₂₅	87.67 ₇₅	87.33 ₆₅	89.39 ₂₅	84.96 ₁₅

ConN, concave number; CorN, corner number; Dan, Danielsson parameter; FN, flake number; Hu1, Hu moment of the first order; MMCD, Max min centre distance; MMD, Max min distance; Rough, roughness.

Table V. Snow particles described by two features and classified with support vector machine (SVM) classifier.

Method	CorN	ConN	Dan	FN	Hu1	MMCD	MMD
ConN	71.47 _{P6}						
Dan	85.22 _{P5}	86.06 _{P4}					
FN	76.25 _{P5}	62.63 _Q	86.60 _{RBF1}				
Hu1	87.16 _{P5}	84.26 _L	89.72 _{RBF1}	83.81 _{RBF2}			
MMCD	80.36 _{P9}	83.47 _{P5}	86.40 _{P7}	80.82 _{P9}	84.34 _{P9}		
MMD	81.27 _{P9}	79.17 _{P9}	84.18 _{P6}	80.24 _{P9}	85.77 _{P9}	84.35 _{P9}	
Rough	87.30 _{RBF1}	89.39 _{P3}	87.31 _{RBF1}	87.60 _{RBF1}	90.98 _L	87.49 _{P9}	87.79 _L

ConN, concave number; CorN, corner number; Dan, Danielsson parameter; FN, flake number; Hu1, Hu moment of the first order; MMCD, Max min centre distance; MMD, Max min distance; Rough, roughness.

Table VI. Two stage system results: snowflake discrimination threshold followed by k -nearest neighbour (KNN) for two features.

Method	ConN	Dan	FN	Hu1	MMCD	MMD
Dan	87.34 ₁₅					
FN	63.73 ₆₅	85.61 ₂₅				
Hu1	86.66 ₂₅	85.38 ₂₅	86.80 ₉₅			
MMCD	88.74 ₁₅	88.85 ₁₅	89.28 ₃₅	89.25 ₃₅		
MMD	83.63 ₃₅	87.14 ₁₅	83.99 ₃₅	83.87 ₆₅	87.31 ₂₅	
Rough	87.90 ₂₅	85.52 ₂₅	87.64 ₇₅	87.30 ₆₅	89.44 ₂₅	85.19 ₁₅

ConN, concave number; CorN, corner number; Dan, Danielsson parameter; FN, flake number; Hu1, Hu moment of the first order; MMCD, Max min centre distance; MMD, Max min distance; Rough, roughness.

Table VII. Two stage system: snowflake discrimination threshold followed by support vector machine (SVM) for two features.

Method	ConN	Dan	FN	Hu1	MMCD	MMD
Dan	86.69 _L					
FN	69.15 _{RBF4}	86.63 _{RBF1}				
Hu1	86.68 _L	90.07 _{P3}	86.77 _Q			
MMCD	85.41 _{P6}	86.55 _{P7}	82.66 _{P9}	86.77 _{P4}		
MMD	80.44 _{P5}	84.32 _L	81.37 _{P9}	87.49 _{P4}	84.90 _{P9}	
Rough	89.35 _{P3}	87.44 _Q	87.55 _L	91.15 _Q	87.27 _Q	88.16 _Q

ConN, concave number; CorN, corner number; Dan, Danielsson parameter; FN, flake number; Hu1, Hu moment of the first order; MMCD, Max min centre distance; MMD, Max min distance; Rough, roughness.

noticeable in the case of classifiers which presented a weak performance in the first experiment, e.g. *ConN*, *FN*, *Dan*. Generally, the results improved by around 1–2%. In particular, the highest improvement, of almost 30%, was found for the *FN-Dan* feature pair in the kNN classification. Finally, there are two pairs, *Hu1-Dan* and *Rough-Hu1*, whose score is 91.15%.

5. Conclusions

This article presents research concerning an automatic classification of snow particle images, especially for cases of snowflake and graupel. Statistical and shape features have been investigated as a method of snow particle description. Additionally, five novel shape features were introduced and tested. For classification purposes, threshold, k -nearest neighbour (kNN) and support vector machine (SVM) classifiers were used.

Firstly, by threshold classification, the *Hu1* moment was chosen as the best from the statistical features, roughness (*Rough*) and Danielsson (*Dan*) parameters were used to represent shape features, and also the performance of novel features were tested. The highest correct classification ratio, of 88%, was achieved by max min centre distance (*MMCD*). Next, for classification purposes the kNN or SVM were applied. In both cases the results improved. In the following experiment, each snow particle was described by two different features. For this two-class classification problem the kNN or SVM were also applied. The use of two features improved the classification results. The best score (90.98%) was achieved for the *Rough-Hu1* pair classified with SVM . Finally,

the two step classifier was presented. Its application improved the results slightly. In this case the best result of 91.15% was also achieved by the *Rough-Hu1* pair classified with SVM . However, it is also worth mentioning that the *MMCD-Hu1* and *MMCD-FN* pairs classified with kNN in the two last experiments score the second place. Finally, it could be stated, that considering the simplicity of creating a one dimensional classifier based on the *MMCD* parameter it works very well, as it loses only 3% to the best two-dimensional classifiers.

To conclude, the presented methods for snow particle classification proved to be a powerful tool. Considering the difficulties arising from the fact that the classified objects are the natural phenomenon, the accuracy above 90% is very high. Therefore, it is hoped that many application for this system could be found. Knowledge about snowfall gained by this system may improve the understanding of radar backscatter. Furthermore, the possible fields of application are research concerning radar meteorology, precipitation and also the physics of snow particle growth.

References

- Boi P. 2009. Snow cover retrieval over Italy and alpine regions using MSG data for climatologic purposes. *Meteorological Applications* **17**: 313–320.
- Burges CJC. 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* **2**: 121–167.
- DeGaetano AT, Wilks DS. 1999. Mitigating snow-induced roof collapses using climate data and weather forecasts. *Meteorological Applications* **6**: 301–312.

- El-Magd A, Chandrasekar V, Bringi VN, Strapp W. 2000. Multiparameter radar and in situ aircraft observation of graupel and hail. *IEEE Transactions on Geoscience and Remote Sensing* **38**: 570–578.
- Harimaya T, Kodama H, Muramoto K. 2004. Regional differences in snowflake size distributions. *Journal of Meteorological Society of Japan* **82**: 895–903.
- Harimaya T, Nakai Y. 1999. Riming growth process contributing to the formation of snowfall in orographic areas of Japan facing the Japan sea. *Journal of Meteorological Society of Japan* **77**: 101–115.
- Hu M-K. 1962. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory* **IT-8**: 179–187.
- Jahne B. 2002. Shape presentation and analysis. *Digital Image Processing*, 5th Revised and Extended Revision. Springer-Verlag: Berlin; 495–511.
- Löffler-Man M, Blahak U. 2001. Estimation of equivalent radar reflectivity factor from measured snow size spectra. *Journal of Applied Meteorology and Climatology* **40**: 843–849.
- Marshall JS, Gunn KLS. 1952. Measurement of snow parameters by radar. *Journal of Atmospheric Science* **9**: 322–327.
- Matrosov SY. 1992. Radar reflectivity in snowfall. *IEEE Transactions on Geoscience and Remote Sensing* **30**: 454–461.
- Mizuno H. 1992. Statistical characteristics of graupel precipitation over the Japan Islands. *Journal of Meteorological Society of Japan* **70**: 115–121.
- Russ JC. 1998. Image measurements. *The Image Processing Handbook*, 3rd edn. CRC Press, Springer, and IEEE Press: Boca Raton, FL; 509–574.
- Shao J, Laux SJ, Trainor BJ, Pettifer REW. 2003. Nowcasts of temperature and ice on overhead railway transmission wires. *Meteorological Applications* **10**: 123–133.