


METHODOLOGY

Open Access



Segmentation of dust storm areas on Mars images using principal component analysis and neural network

Ryusei Gichu¹ and Kazunori Ogohara^{2*} 

Abstract

We present a method for automated segmentation of dust storm areas on Mars images observed by an orbiter. We divide them into small patches. Normal basis vectors are obtained from the many small patches by principal component analysis. We train a classifier using coefficients of these basis vectors as feature vectors. All patches in test images are categorized into one of the dust storm, cloud, and surface classes by the classifier. Each pixel may be included in several dust storm patches. The pixel is classified as a dust storm or the other classes based on the number of dust storm patches that include the pixel. We evaluate the segmentation method by the receiver operator characteristic curve and the area under the curve (AUC). AUC for dust storm is 0.947–0.978 if dust storm areas determined by our visual inspection are assumed to be ground truth. Precision, recall, and *F*-measure for dust storm are 0.88, 0.84, and 0.86, respectively, if we remove false negative pixels efficiently and maintain the size of true positive dust storms using two different threshold values. The tuning parameters of the classifier used in this study are determined so that the accuracy for dust storm is maximized. We can also tune the classifier for cloud segmentation by changing the parameters.

Keywords: Mars, Dust storm, Segmentation, Machine learning, Principal component analysis

Introduction

The planet Mars is regularly affected by dust storms that cover contiguous areas ranging in size from local (10,000 km²) to global ($> 5 \times 10^7$ km²) and present a variety of hazards/challenges to present and future exploration of the planet. Martian dust storms are atmospheric phenomena visualized by dust suspended in the atmosphere. The atmospheric phenomena controlling dust storms are suggested by the storms' shapes and sizes, as well as the season, region, and local time of initiation. Wang et al. (2011) showed the distribution of curvilinear dust storms in the southern hemisphere observed by MGS/MOC. Guzewich et al. (2015) clearly showed the frequency, regionality, and seasonality of "textured dust storms" seen in MGS/MOC images. Guzewich et al. (2017) revealed the relationship between the frequency of textured dust storms in images from the Mars Color

Imager onboard the Mars Reconnaissance Orbiter (MRO/MARCI) and the surface albedo. Their climatological research on curvilinear or textured dust storms suggested their mechanisms statistically for the first time, which had a large impact on dust storm studies. Although they assumed that active convection inside dust storms was reflected in the storms' textures, they did not clarify what types of atmospheric phenomena are associated with the production of large surface stresses that trigger first dust lifting.

The shape, texture (mesoscale cloud top morphology, Kulowski et al. 2017), and size of dust storms and the climatology of these characteristics are important clues to understanding Martian dust storms. The relations between such characteristics and the phases of the variety of atmospheric waves as well as the frequency of textured/curvilinear dust storms and the spatial and seasonal variations of the cumulative area are also worth investigating (Wang 2007; Wang et al. 2011; Guzewich et al. 2015, 2017). However, it is time-consuming to detect all dust storms visually because of the vast number

* Correspondence: ogohara.k@e.usp.ac.jp

²School of Engineering, University of Shiga Prefecture, 2500, Hassaka, Hikone, Shiga 522-8533, Japan

Full list of author information is available at the end of the article

of images taken by just two instruments, the MGS/MOC and MRO/MARCI. It is also difficult to objectively define which features characterize dust storms. Even approximate categorization of dust events into textured dust storms, untextured yet discrete dust storms, and haze depends on the subjective experience of the human observer. For these reasons, visual detection and categorization of the three types of dust storm textures introduced by Kulowski et al. (2017) would not be exactly reproduced by another human viewing and characterizing the images. Furthermore, we perceive a need to update the criteria for detecting dust events, especially obscured ones such as untextured yet discrete dust storms and haze, via visual detection. Therefore, it would be useful to be able to automatically detect dust events, measure their shape, pattern, and size, and improve the objectivity and reproducibility of dust storm detection. If a dust storm “catalog” summarizing the shape, size, and textures of dust storms was produced objectively, it would be an important database for studies of atmospheric phenomena controlling dust storm initiation. In addition, we may be able to automatically detect and categorize dust storms using images taken by narrow-angle cameras as well as global swath images taken by MGS/MOC and MRO/MARCI.

Maeda et al. (2015) automatically evaluated the existence of dust storms in Mars images using a support vector machine (SVM), which is an algorithm for pattern recognition using supervised learning that has been applied to classification and regression. Their algorithm successfully detected 80% of human-identified dust storms, but they did not try automated segmentation of dust storm regions in their study. Ogohara et al. (2016) also successfully evaluated the existence of water ice clouds (WICs) in the Martian atmosphere using an SVM trained by a few simple statistics of images. Performance of their algorithm was about 80%, comparable with Maeda et al. (2015), but automated segmentation of WICs was beyond their work. In a general application of pattern recognition and computer vision, an image is divided into several small patches and each patch is classified into one or more classes (e.g., Coupé et al. 2011). Turk and Pentland (1991) introduced eigenfaces using principal component analysis (PCA) for face recognition. Jiang et al. (2017) also used the patch-based method and PCA for face recognition. In this study, we apply the combination of the patch-based approach and PCA to Mars images to segment Martian dust storms. This enables us not only to evaluate the existence of dust storms, but also to measure the shape and size of dust storms automatically. Furthermore, we may be able to recognize and categorize textures of dust storms because we can obtain information on which basis images contribute to a patch classified as a dust storm. In the

“Methods/Experimental” section, we provide an overview of the observation data that we use in this study and the pre-processing of input image data. We also describe details of our algorithm for automated segmentation of dust storm areas using PCA and machine learning. Results and performance evaluation of the algorithm are shown in the “Results and discussion” section.

Methods/experimental

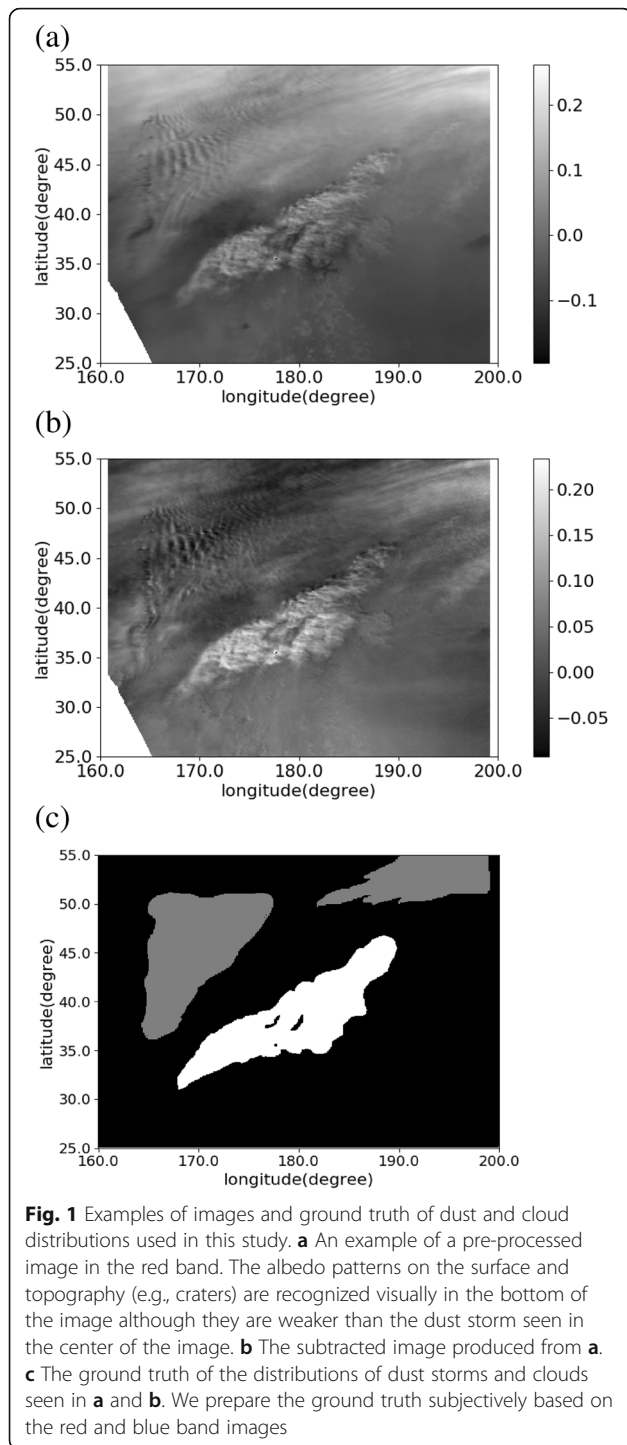
Data

We used reflectance data from red and blue bands (575–675 nm and 400–450 nm, respectively) images acquired by MGS/MOC (Malin et al., 2010). In this study, we begin our investigation of methods for segmentation of dust storm areas with only the region centered around 180° E–40° N where many textured dust storms have been observed (Guzewich et al. 2015). This region is suitable for training a classifier described in the following subsection because there are many samples of true dust storms. We examine 800 × 600 pixels (40° × 30°) subsets of the full-size MOC wide-angle camera (MOC-WA) image swathes. The subsets are centered at 180° E–40° N and extracted from MOC-WA images taken in the late northern fall of MY24 and MY26, one of the active seasons of dust storms in this latitude band (Guzewich et al. 2017). Red and blue band images of MOC-WA were obtained through the U.S. Geological Survey (USGS, https://ida.wr.usgs.gov/d_access.htm). Noise reduction, radiometric calibration, and coordinate transformation to longitude-latitude coordinates were done using ISIS3 (<https://isis.astrogeology.usgs.gov/index.html>). Opposition surge and low-frequency patterns of reflectance were removed from the images using methods presented by Wang and Ingersoll (2002) and Ogohara et al. (2016). Coherent longitudinally inhomogeneous patterns shown by Ogohara et al. (2016) were filtered out using the longitudinal running mean of a full 31-pixel width.

Figure 1a shows a typical dust storm red band image in the region of 160° E–200° E and 25° N–55° N obtained after the pre-processing noted above. The resolution is 0.05° pixel⁻¹. Textures at the top of the dust storm are clearly seen. However, surface albedo patterns clearly visible in the bottom of the image might potentially be falsely classified as a dust storm, and therefore, we attenuate the surface albedo patterns by subtracting a regionally defined, dust-free background image from each investigated image of interest. Each pixel value of the background image, $I_b(i, j)$, is expressed as follows:

$$I_b(i, j) = \min_{1 \leq n \leq 9} I_n(i, j), \quad (1)$$

where I_n means the n th image. The nine images used for



producing I_b are listed in the (Additional file 1: Figure S1). They were chosen arbitrarily from images where few or no dust storms were seen. After some trials using less than nine images to create a background subtraction image, we determined that we were best able to attenuate the surface patterns relative to dust storms and clouds using at least nine images. Figure 1b shows an

example of a dust storm image obtained by subtracting a background image from Fig. 1a. The surface patterns seen in Fig. 1a have been attenuated in Fig. 1b, and the dust storm has been emphasized. Note that we do not necessarily remove the surface patterns completely although the surface patterns might be much attenuated in the subtracted image if more than nine images were used.

Basis determination

We extract many patches (a sub-sampled region of size $N \times N$ pixels within the original image) from six images showing dust storms in the red band observed at different times and those in the blue band observed at the same times as the red band. These twelve images of dust storms (six observations in two bands) were chosen because they were already known to have dust storms in them and were chosen arbitrarily from among those examined that had dust storms. A patch of $N = 10$, at the resolution of the MOC-WA images, is comparable in size to the smallest dust event summarized by Cantor et al. (2001) in the northern part of the images used for this study. Extreme values of N (e.g., $N = 2$ or 100) are inappropriate because the method for dust storm segmentation proposed in the following sections implicitly assumes that the patch size is comparable to or smaller than dust storms and larger than the patterns characteristic of dust storms. Therefore, the cases of $N = 10, 20$, and 30 are investigated in this study. All patches are extracted from the six images allowing the superposition of two adjacent patches, which share $(N-2)N$ pixels. A list of the patches is created by starting at one corner of the image and stepping by two pixels in one direction, until the next corner is reached, then returning to the start, at the same place in the first direction, but then shifting by two pixels in the second direction.

Patches classified as “surface” are the dominant classification in the list because the majority of the pixels seen in these images shows the surface, rather than dust or ice clouds. Imbalance in the number of patches between the three classes may cause a reduction of classification accuracy of the two minor classes. This means that the classifier can distinguish the surface from the other regions but cannot separate dust storm and cloud. In addition, we should use the same number of patches in the three values of N to compare the performances of the proposed method between $N = 10, 20$, and 30 . Therefore, we actually extract 140,000 patches of each class randomly from the list of the patches in each of the values of N after assigning each patch to one of the three classes in the manner described in the “Classifier and ground truth” subsection. This number of patches for each class is limited by the memory constraints of our computer system that are required during

processing, especially in the case of $N = 30$. Letting M be the number of patches generated from the six subtracted images, M is 420,000 ($140,000 \times 3$) in each of the three values of N used.

The i th patch ($1 \leq i \leq M$) can be expressed as an N^2 -dimensional vector, $\mathbf{t}_i \in \mathbb{R}^{N^2}$, whose elements are identical to the pixel values of the patch. We derive basis vectors of the N^2 -dimensional space using PCA. Note that the average patch, the mean over all the M patches, $\bar{\mathbf{t}} \in \mathbb{R}^{N^2}$, is subtracted from each of the M patches before the PCA is done. The normalized basis vectors derived by PCA are perpendicular to each other, and an arbitrary vector can be uniquely expressed as a linear combination of them. The number of the basis vectors is generally N^2 but N^2 is so large that it takes a long time to determine the best combination of parameters for classification and to train a classifier using it. Therefore, we choose the K ($< N^2$) basis vectors with the largest eigenvalues, $\mathbf{e}_{PCA_j} \in \mathbb{R}^{N^2}$ ($1 \leq j \leq K$), from the N^2 basis vectors derived by PCA. K is determined so that the cumulative proportion of variance, i.e., the variability in the pixel reflectances across the M patches, represented by a linear sum of the K basis vectors, becomes larger than 99%. Note that each of the K basis vectors does not necessarily correspond to some kind of variability in reflectance that can be interpreted physically.

Feature vector

$\bar{\mathbf{t}}$ is subtracted from each patch vector. Subtracted patch vectors are defined as follows:

$$\mathbf{s}_i = \mathbf{t}_i - \bar{\mathbf{t}} \quad (1 \leq i \leq M) \quad (2)$$

\mathbf{s}_i can be expressed uniquely as a linear combination of N^2 PCA basis vectors. But, because $K < N^2$, we cannot exactly express it using \mathbf{e}_{PCA_j} ($1 \leq j \leq K$). Thus, we approximate it by an orthogonal projection, \mathbf{s}'_i , on a plane mapped by \mathbf{e}_{PCA_j} :

$$\mathbf{s}'_i = U \mathbf{c}_i, \quad (3)$$

$$\mathbf{c}_i = U^T \mathbf{s}_i, \quad (4)$$

where U is an orthogonal projection matrix, $U = (\mathbf{e}_{PCA_1}, \mathbf{e}_{PCA_2}, \dots, \mathbf{e}_{PCA_K})$. $\mathbf{c}_i = (c_{i1}, c_{i2}, \dots, c_{iK})^T \in \mathbb{R}^K$ is a coefficient vector and means the position of \mathbf{s}'_i in the subspace mapped by U .

We use reflectance images in the two bands, red and blue, as described in the previous section. We extract K basis vectors using PCA from M patches of each band. K depends on the band. A feature vector used for the classification is a combination of the two feature vectors derived from one red patch and one blue patch. Letting \bar{t}_{Ri} and \bar{t}_{Bi} be the averages of the i th patch (the mean value over all elements of \mathbf{t}_i) in the red and blue bands,

respectively, the feature vector of the i th patch, \mathbf{f}_i , is defined as follows:

$$\mathbf{f}_i = \begin{pmatrix} \mathbf{c}_{Ri} \\ \mathbf{c}_{Bi} \\ \frac{\bar{t}_{Ri}}{\bar{t}_{Bi}} \end{pmatrix} \quad (5)$$

where $\mathbf{c}_{Ri} = (c_{Ri1}, c_{Ri2}, \dots, c_{RiK_R})^T$ and $\mathbf{c}_{Bi} = (c_{Bi1}, c_{Bi2}, \dots, c_{BiK_B})^T$ are coefficient vectors of the i th patch in the red and blue bands, respectively. K_R and K_B are the numbers of the basis vectors that contribute most to encompassing the image variability in their respective bands as defined by the PCA analysis.

Classifier and ground truth

We adopt a multilayer perceptron (MLP) classifier, which is included in scikit-learn0.18.1 (Python 3.5.1). MLP is a feedforward artificial neural network (NN) that imitates biological neural networks. NNs have been used for several fields in planetary science (e.g., classification of asteroid spectra, Howell et al. 1994) and engineering (e.g., identification of environment models around space robots, Venkataraman et al. 1993). An MLP at least consists of three layers of neurons, the input layer, output layer, and one or more hidden layers between them. The numbers of hidden layers and neurons in each layer are typical tuning parameters controlling the complexity and ability of the network. There is only one hidden layer in this study. We provide an overview of the MLP in the “Appendix” section for readers who have no experience in machine learning. Please see the “Appendix” section for the meanings of technical terms used later in this section. The number of neurons in the hidden layer $N_{\text{neuron}} = \{10, 20, 30, 70, 100\}$, the learning rate (const.) $\gamma = \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, and the maximum number of iterations $T_{\text{max}} = \{100, 300, 500, 700, 1000, 5000\}$ are determined by grid search for each patch size so that accuracy for the dust storm class is maximized. The activation function for the hidden layer is the so-called “ReLU” function, $f(x) = \max(0, x)$, for all patch sizes. The solver for weight optimization is a stochastic gradient-based optimizer, Adam, proposed by Kingma and Ba (2015). The parameters for the optimization are listed in Table 1. We did not arbitrarily choose the ReLU function, the Adam optimizer, and the parameters shown in Table 1, but adopted them in this study because they are popular techniques with popular parameter choices and produced adequate results, as discussed in the “Results and discussion” section.

The ground truth of the segmentation was prepared by the authors in advance of training the classifier. We first assign each pixel to a dust storm, cloud, or surface class based on our subjective experience. Thus, some

Table 1 Parameters for the MLP classifier included in scikit-learn0.18.1. These are the same as commonly used values. See Kingma and Ba (2015) for the meanings of the parameters for the Adam optimizer

Parameter	Value
Penalty for regularization	0.0001
Batch size	200
Tolerance of optimization	10^{-4}
Whether to shuffle samples in each iteration	True
Whether to use early stopping to terminate training when the validation score is not improving	False
β_1 for adam	0.9
β_2 for adam	0.999
for adam	10^{-8}

patches may contain two or more classes representing different labeled regions of pixels with them. Next, if both the dust storm and cloud areas inside each patch are smaller than 20% of the patch area (N^2 pixels) (Ishii et al., 2016), the patch is labeled “0” indicating the surface class. If not, the patch is assigned the label of the larger of the two other classes (“1” for dust storm and “2” for cloud). The MLP is trained using the feature vectors of the patches (f_i in Eq. (5) and Eqs. (6) and (10) in the “Appendix” section) with the true labels (g_i in Eq. (10) in the “Appendix” section). The six subtracted images and corresponding ground truth images of the segmentation that we used for training are shown in the (Additional file 2: Figure S2). Dust, cloud, and surface patches used for training are the same as those from which the basis patch vectors are determined and the number of patches of each class is 140,000 in the cases of $N = 10, 20,$ and 30 . We could use more surface patches because the surface area is much larger than areas of dust storm and cloud. However, we chose 140,000 patches from all surface patches randomly to avoid an imbalance in the number of patches between the three labels as described in the “Basis determination” subsection.

Observations that can physically decide whether features seen in images are dust storm, clouds, or something else already exist. The Thermal Emission Spectrometer onboard MGS (MGS/TES) can measure optical thickness of dust and water ice cloud separately. Such types of observation data with physical information on dust and cloud seem to be suitable for the ground truth data instead of visual inspection by the authors. However, the footprint of TES is extremely narrow compared to the field of view of MOC and thus rarely crosses local dust storms. TES observations are not enough to train the classifier. Therefore, we do not prepare the ground truth images based on TES observations but based on visual inspection by the authors to maintain the number of patches available for training.

Probability image

What is classified into the dust storm class or the others is not each image but each patch. If patches into which an image is divided as shown in the “Basis determination” subsection are targets of classification, the shape of a dust storm cannot be sufficiently resolved. Then, we shift a target patch, which is classified into one of the three classes in a test image, pixel by pixel, longitudinally, and latitudinally. An arbitrary pixel is basically included in N^2 target patches. Letting $n_D(i, j)$ be the number of patches which include a current pixel (i, j) ($i, j \in \mathbb{Z}$) and are categorized as a dust storm, we hereafter call $n_D(i, j)/N^2$ the “dust storm probability” of the pixel (i, j) . An image whose pixel values are a dust storm probability between 0 and 1 is obtained, which is called a “probability image” in this study. The shape and size of a dust storm can be measured accurately by regarding pixels whose dust storm probability is larger than a threshold value as the dust storm. Letting $n_C(i, j)$ be the number of patches which include a current pixel (i, j) and are categorized as a cloud, we can also define the “cloud probability” of the pixel (i, j) , $n_C(i, j)/N^2$.

Results and discussion

The classifier, in practice, must be able to separate surface, dust storm, and cloud areas in images different from those used in its training. In the following subsections, we test five chosen subtracted images (Additional file 3: Figure S3) and refer to them as the “test images” hereafter. Textured dust storms seen in four of the five test images were most distinctive (note that one of the five test images shows no dust storm), and therefore, we could prepare reliable ground truth images.

Probability image

Figure 1b already mentioned in the “Data” subsection is one of the subtracted images used for evaluation of the method developed in this study and is excluded in the phase of training the classifier. Figure 1c indicates the true distributions of a dust storm and clouds included in

Fig. 1b, which were prepared based on the authors' visual inspection. Figure 2a, b, and c are probability images of dust for $N=10$, 20, and 30, respectively, derived from Fig. 1b. Figure 2d, e, and f are those of WIC. The shape of the dust storm is resolved more precisely in the case of $N=10$ (Fig. 2a). But, the wavy features in the north-west area of the image should not be recognized as dust storms, because they seem to be typical WICs associated with topographically generated gravity waves, the so-called lee waves. In the cases of $N=20$ and 30 (Fig. 2b, c), a high probability of dust storm associated with WICs still remains in the northwest of the probability images. However, dust storm probability near the eastern and western edges of the images is smaller than that in the case of $N=10$. Nevertheless, the size of the dust storm in the case of $N=20$ (Fig. 2b) is comparable with that in the case of $N=10$ shown in Fig. 2a. Cloud probability is high in the northern part of the images in Fig. 2d, e, and f and seems to vary inversely proportional to N . Cloud probability in the wavy area does not decrease largely with the size of patches. However, the area of high probability around the wavy WICs itself is clearly smaller than that of the true wavy WIC area shown in Fig. 1c. The wavy patterns of WICs seem to be confused with dust storms, especially in the case of $N=10$.

Receiver operator characteristic curves

In this subsection, we evaluate how accurately dust storm areas are determined in the test images that were

not used for training. Note that we here have to divide an image into two classes (dust storm or not dust storm) and do not need to distinguish between cloud and surface areas. Therefore, we binarize probability images of dust by thresholding of the images, that is, we produce an image of pixel values of either 0 or 1 (black or white) by determining whether the dust probability in that pixel is below or above some chosen threshold value. However, the binary images highly depend on the threshold value used to binarize the probability images. The best threshold values may be different between $N=10$, 20, and 30. Therefore, we cannot compare the results of $N=10$, 20, and 30 derived using a single threshold value straightforwardly. Thus, evaluation methods independent of threshold values as well as those that depend on threshold values have been proposed in the discipline of pattern recognition. Those pertinent to this study are described briefly in the "Appendix" section. Figure 3 shows one method, receiver operator characteristic (ROC) curves calculated by sweeping the threshold values in probability (Fawcett 2006), indicating the relationship of the true and false positive rates. ROC curves have been used in the fields of medical imaging and are independent of the ratio of the true positive area (dust storm) and the true negative area (the others). Area under the curve (AUC) is used for evaluating the performance of the algorithm independently of the threshold value (see Appendix and Fawcett 2006 for details). The maximum and best values in AUC are 1. AUC in

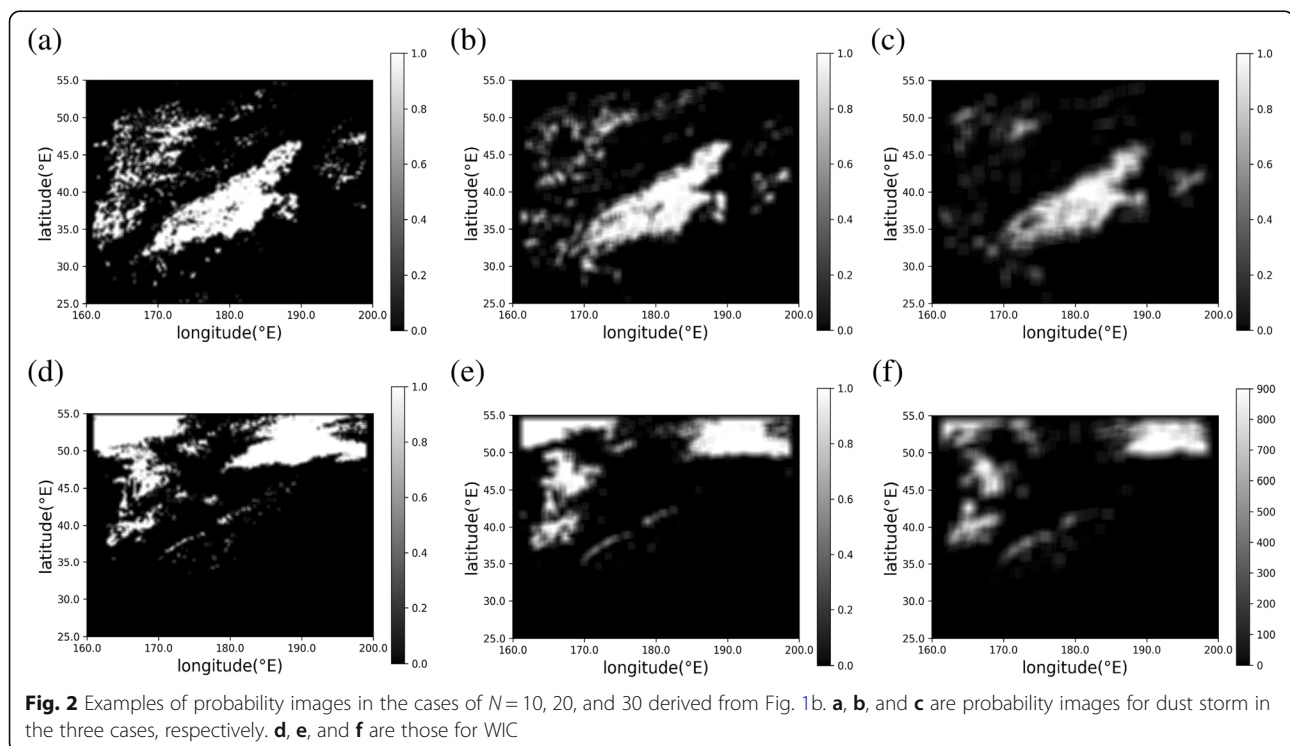
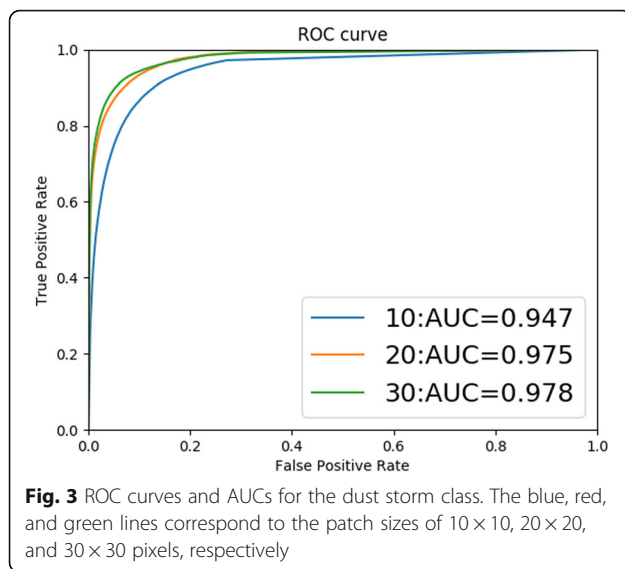


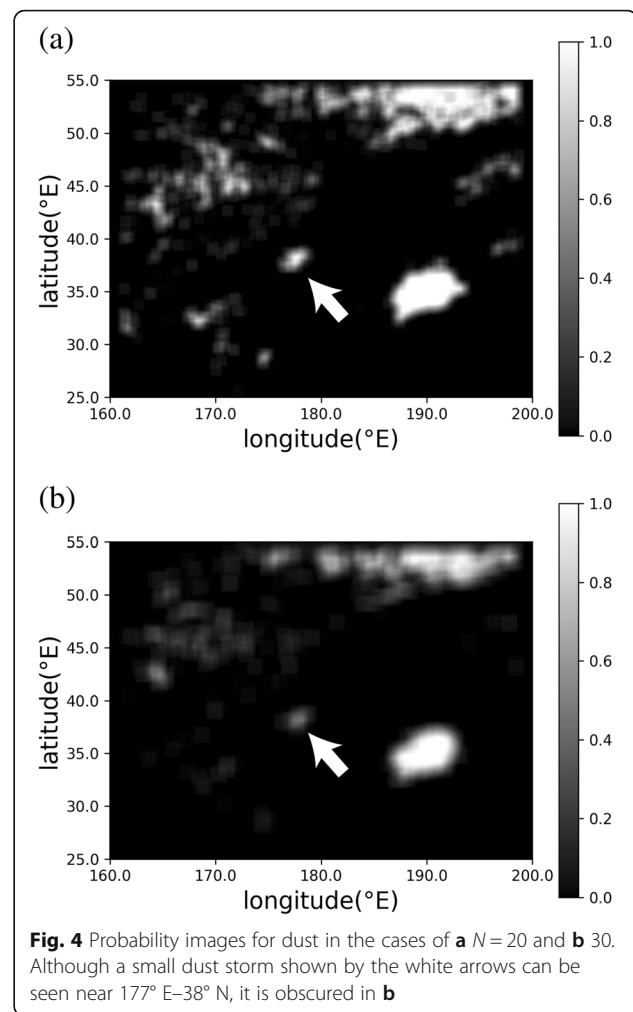
Fig. 2 Examples of probability images in the cases of $N=10$, 20, and 30 derived from Fig. 1b. **a**, **b**, and **c** are probability images for dust storm in the three cases, respectively. **d**, **e**, and **f** are those for WIC



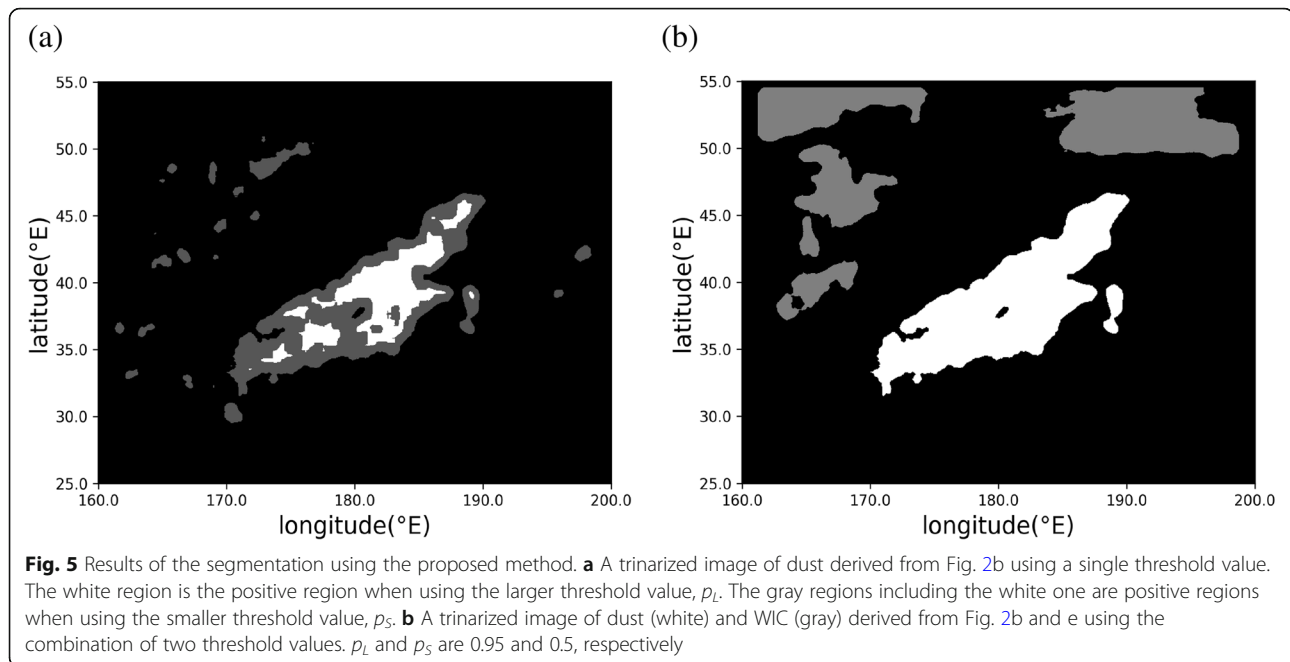
the case of $N = 30$ is the highest among the three cases for dust storm as shown in Fig. 3. The case with $N = 30$ is the best in that true positive pixels can drastically increase only by tolerating just a small number of false positive pixels. However, 30 pixels (1.5°) in latitude roughly correspond to 90 km. The smallest dust storms among local dust storms summarized by Cantor et al. (2001) were about $17 \times 17 \text{ km}^2$. They may be attenuated in dust probability images in the case of $N = 30$ due to their small size compared to the patch size. Figure 4 shows the dust probability images in the cases of $N = 20$ and 30 produced from a subtracted image in which a small local dust storm can be seen. The small dust storm has a horizontal scale of 100 km. Dust probability around it is obscured and clearly lower in the case of $N = 30$ than in the case of $N = 20$. This does not only lead to underestimation of the area of the small dust storm but also makes it difficult to detect the existence of the small dust storm. Nevertheless, the difference in AUC between $N = 20$ and 30 is just 0.003. Thus, we focus on results of $N = 20$ hereafter rather than $N = 30$ to suppress false negatives of dust storms.

Binarized image

For segmentation of dust storm regions, we need to binarize the probability images using a threshold value as indicated in the above subsection and evaluate the generalization ability of the method using evaluation indices such as precision, recall, and F -measure (e.g., Fawcett 2006). A threshold value close to 1 drastically reduces positive pixels in binary images in general. Precision gets close to 1 but recall is extremely reduced. As the threshold value decreases, the positive areas grow and other tiny positive areas appear in the binary images. As a result, recall becomes greater than precision.



Although priorities of precision and recall are up to individual research interest, improving precision (i.e., reducing false positive pixels) without reducing recall (i.e., maintaining true positive pixels) is generally encouraged. Thus, we binarize the probability images using two threshold values in this study. At first, we generate a binary image, B_L , from a probability image using the larger threshold, p_L , as shown by the white region in Fig. 5a. Positive areas seen in B_L tend to be small and then recall is low. However, the positive areas in B_L can be regarded as being in dust storms with high reliability. At the next step, we generate another binary image, B_S , from the same probability image using the smaller threshold, p_S , as shown by the gray regions in Fig. 5a. The positive areas seen in B_S tend to be large, and then, recall is high although precision is low. This means that the positive areas in B_S include the surface or WIC regions as well as dust storm regions. It should be noted that the all positive regions in B_L (white in Fig. 5a) are inside the positive regions in B_S (gray in Fig. 5a). However, not all the positive regions in B_S include positive regions in B_L .



Then, we regard the positive regions in B_S which do not include positive regions in B_L as false positives and remove them. We can detect the true dust storm regions with high recall, suppressing the increase of false positive pixels due to decreasing the threshold value. Figure 5b shows a trinarized image generated from the probability images of dust and WIC shown in Fig. 2b and e using $(p_L, p_S) = (0.95, 0.5)$. The size of the dust storm is maintained, and the tiny false positive regions seen in Fig. 5a are removed. We binarize the probability images used for the ROC curves in the “receiver operator characteristics curves” subsection and calculate precision, recall, and F -measure from the binarized images and the ground truth images. Table 2 shows the evaluation indices directly calculated from all of the pixels of the five binary images derived from the test images via the probability images. Note that each evaluation index is not an average over the five index values calculated from the five binary images. For both dust storm and cloud, recall in the case of the two threshold values (0.84) is comparable with or higher than those in the cases of the single threshold value. On the other hand, precision (0.88) is between the two cases of the single threshold value. Therefore, F -measure improves in the two-threshold values cases.

Table 2 Precision, recall, and F -measure for dust storm calculated from all of the five test images. The two threshold values, p_L and p_S , are 0.95 and 0.5, respectively

Threshold value	Precision	Recall	F -measure
p_L	0.97	0.51	0.67
p_S	0.83	0.84	0.83
p_L and p_S	0.88	0.84	0.86

$(p_L, p_S) = (0.95, 0.5)$ used for binarization is just an example and actually depends on the individual motivation of the segmentation of dust storms. If one is investigating the size distributions of dust storms including small dust storms, one has to decrease p_L . If the shape of relatively large dust storms is a major target, both p_L and p_S should be large. The sensitivity of these results to location and season is discussed in the next section.

Generally speaking, Martian dust storms are bright in the red band and as dark as the surface in the blue band. On the other hand, Martian clouds are bright in the red and blue bands and especially much brighter than the surface in the blue band. Cantor et al. (2001) visually detected dust storms based on the difference in the surface albedo from adjacent areas in both the bands. As they did, we have prepared the ground truth of the segmentation based on differences in appearance between dust storms, clouds, and the surface due to such differences in their optical properties. The ground truth data for optically thick phenomena with textures different from the surrounding surface are reliable (e.g., Kulowski et al. 2017). However, those for optically thin dust haze and clouds are not always as reliable as those for optically thick dust events. We have experienced difficulty in visually deciding whether the pattern seen in the northern part of the target region is a dust haze or cloud. Therefore, what have been reported in this study are conservatively an algorithm for the segmentation of optically thick dust events and clouds, and its performance.

Nonetheless, our algorithm for the automated segmentation achieves high AUC if we evaluate the algorithm based on the subjective ground truth. The results of the

segmentation shown in this paper can depend on who prepares the ground truth used for training the classifier and evaluating the algorithm. For example, someone may regard the area around 187° E– 38° N as a part of a dust storm. Figure 2a and b show that the area around 187° E– 38° N is partially recognized as a dust storm, although it is not included in the true dust storm area shown in Fig. 1c. This means that the classifier is not so sensitive to such minor variations in the ground truth data depending on the inspectors' experience. Therefore, we expect that we could achieve physics-based segmentation with as high AUC as that presented in this study if objective, adequate ground truth data based on other observations (e.g., radars and spectrometers) became available.

The segmentation results presented in the current work are not sensitive to possible variations in the ground truth images depending on the inspectors, but it is also true that the segmentation results are not necessarily independent of the authors' experience because the ground truth images used for training the classifier and evaluating its performance are prepared subjectively. The proposed method, however, enables fast detection and segmentation of dust storm areas in Mars images with less work based on rigid (but subjective) detection criteria. This helps us intercompare the results of segmentation or detection of dust storms done by several researchers and enables us to reproduce their results. Therefore, the proposed method still makes a large contribution to the improvement of objectivity and reproducibility of dust storm detection.

Precision and recall in the case of the two threshold values are 0.88 and 0.84 if the two threshold values are 0.95 and 0.5. Twelve percent of dust storm areas detected by our method are not dust storm and 16% of true dust storm areas are not detectable. However, the overestimation in the area of dust storms is just 4% as a result of our tuning the two threshold values so that precision and recall become comparable (i.e., F -measure is as high as possible). Our method for dust storm segmentation has a high reliability on measuring area of dust storms. Even if one is intent to measure the shape and size of each dust storm, the method has high enough accuracy because the size distribution of dust storms is not modified over several orders by the error in the area of $\sim 15\%$.

We have tuned the parameters for an MLP so that the performance of dust storm segmentation reaches the peak. Therefore, Figs. 2 and 5 imply that AUC and F -measure for ice clouds are both lower than those for dust storms. But, this method is not necessarily unsuitable for the segmentation of ice clouds. The evaluation indices for clouds can improve if we adopt parameters that maximize the performance of cloud segmentation. This method is also useful for segmentation of clouds.

This study has focused on the region with a high frequency of dust storms revealed by Guzewich et al. (2015, 2017) and Kulowski et al. (2017). The tuning parameters of the MLP mentioned above may be valid in some regions but may be invalid in other regions because patterns of dust storm and haze may depend on locations (i.e., latitude, altitude, and slope angle). They may be also invalid for other images taken at another time if patterns of dust storms vary with season and Mars year, because training and test images we have used are much less in number than images of the region taken during the lifetime of MOC. The region we have focused on is favorable for "plume-like" dust storms defined by Kulowski et al. (2017). There still remains the possibility that the classifier in this study is not applicable to the segmentation of the other two types of dust storms ("pebbled" and "puffy" dust storms). However, we can do the segmentation with a performance comparable with the presented method if the classifier is trained for every type of dust storm depending on the region, season, and local time. Textured dust storms as shown in Fig. 1b should be successfully separated at least.

Conclusions

We have investigated an algorithm for automated segmentation of dust storms in Mars images observed by MGS/MOC. We have divided an image into small patches and have expressed them as linear combinations of the basis patch images extracted by PCA. Coefficients of the basis images have been used for training the classifier, a neural network. All patches included in the test images are categorized into one of the three classes, dust storm, cloud, and something else (usually the ground surface) by the classifier. Pixels will appear in multiple patches, for which some patches containing that pixel will be categorized as a dust storm while some other patches containing that pixel will be categorized as a different class. Using the ratio of patches that classify a pixel as a dust storm to the total number of patches that include that pixel, we evaluate whether the pixel is a part of a dust storm or not to sufficiently resolve the shape and size of dust storms. The ratio of patches in which a pixel is classified as a dust storm to all patches that include the pixel is called the dust storm probability (of that pixel) in this study. Area under the ROC curves can be calculated from the dust storm probability images by sweeping the threshold value for binarization. AUC for the patch sizes $N = 10, 20,$ and 30 are 0.947, 0.975, and 0.978, respectively. AUC is the highest in the case of $N = 30$ but relatively small dust storms cannot be detected in that case due to the large patch size. The AUC of the algorithm presented in this study cannot be compared with that of other algorithms, because the AUC of algorithms for dust storm segmentation has never been

reported by previous research so far. However, the AUC of our algorithm is comparable with those of algorithms reported in other research fields (e.g., medical imaging, Hatanaka et al. 2018). Our algorithm has as high performance as other segmentation algorithms in the fields of general image processing.

We can automatically obtain a catalog summarizing the shape, size, and texture of the variety of dust storms if we apply this algorithm to the vast number of Mars images taken by orbiters for about 20 years. The existence of such a comprehensive catalog would be an important resource for understanding the climatology and processes of Martian dust storms.

Appendix

Multilayer perceptron

Readers can easily find tutorials on MLPs or NNs in standard textbooks (e.g., Bishop 2006) and on the internet. Here, we briefly overview the concept, structure, and parameters of the MLP used in this study for readers who have no experience in machine learning based on Bishop (2006). Figure 6 shows a schematic view of the MLP used in this study, which consists of an input, output, and one hidden layer. In this study, the input layer is a feature vector of each image patch expressed by f_i shown in Eq. (5), the output layer is the series of possible classifications for a patch (i.e., dust storm, ice cloud, or surface), and the hidden layer is the mathematical process transforming the input to the output. The numbers of neurons in the input layer represent the number of features (the dimension of f_i), $K' = K_R + K_B + 2$, and the number of neurons in the output layer represents the number of classes, 3. The number of

neurons in the hidden layer, N_{neuron} , is a tuning parameter and roughly controls the number of adjustments made to the inputs to accurately transform them into one of the output classes (neurons). This network, shown in Fig. 6, is expressed as the following function:

$$y_{ih}(\mathbf{f}_i, \mathbf{W}^{(1)}, \mathbf{W}^{(2)}) = \sigma\left(\sum_n^{N_{\text{neuron}}} z_n W_{hn}^{(2)}\right) \quad (6)$$

$$z_n = \phi\left(\sum_l^{K'} f_{il} W_{nl}^{(1)}\right). \quad (7)$$

f_{il} indicates the l th components of the i th feature vector defined by Eq. (5). $W_{nl}^{(1)}$ is a weighting factor between the l th neuron of the input layer and the n th neuron of the hidden layer, and $W_{hn}^{(2)}$ is a weight between the n th neuron of the hidden layer and the h th neuron of the output layer. $\phi(\cdot): \mathbb{R} \rightarrow \mathbb{R}$ and $\sigma(\cdot): \mathbb{R} \rightarrow \mathbb{R}$ are arbitrary nonlinear activation functions (i.e., mathematical transformations) of the hidden and output layers, respectively, that enable the classifier to be applicable to nonlinear classification problems. For multiclass classification problems so far, the softmax function expressed by the following equation has been the conventional expression used for σ :

$$\sigma(a_h) = \frac{e^{a_h}}{\sum_j e^{a_j}}, \quad (8)$$

where $a_h = \sum_n^{N_{\text{neuron}}} z_n W_{hn}^{(2)}$.

The ground truth vector of the i th patch, \mathbf{g}_i is as follows:

$$\mathbf{g}_i = \begin{cases} (1 \ 0 \ 0)^T & (\text{The true label is "0"}) \\ (0 \ 1 \ 0)^T & (\text{The true label is "1"}) \\ (0 \ 0 \ 1)^T & (\text{The true label is "2"}) \end{cases} \in \mathbb{R}^3 \quad (9)$$

Letting \mathbf{y}_i be $(y_{i1}, y_{i2}, y_{i3})^T$, the error function of this network is expressed as follows:

$$E(\mathbf{W}^{(1)}, \mathbf{W}^{(2)}) = \frac{1}{2} \sum_i^M \left\| \mathbf{y}_i(\mathbf{f}_i, \mathbf{W}^{(1)}, \mathbf{W}^{(2)}) - \mathbf{g}_i \right\|^2 \quad (10)$$

Training this classifier means solving the optimization problem of the weights by minimizing E (i.e., searching for optimal weights so that E is minimized). Once the optimal weights are determined, we can predict the classes of test patches not used for training by replacing \mathbf{f}_i with the test patch vectors.

Optimization of Eq. (10) is usually a multidimensional problem and has multiple local minimums. Therefore, several algorithms for getting the global solution within

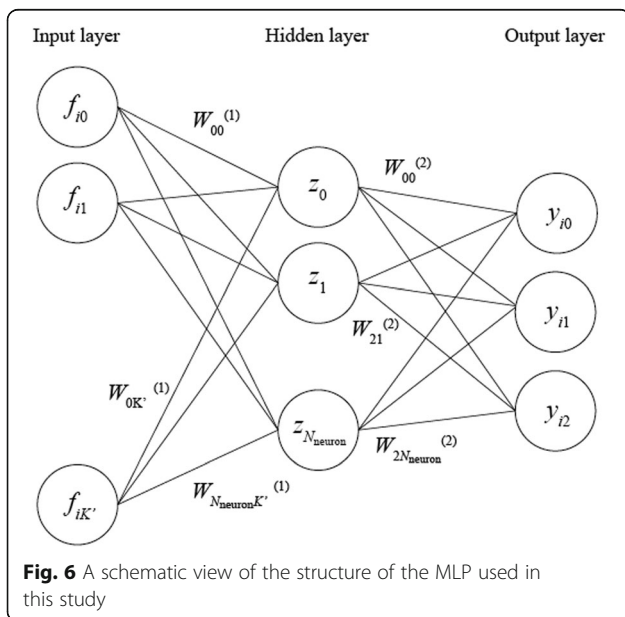


Fig. 6 A schematic view of the structure of the MLP used in this study

a realistic time, called “optimizers” or “solvers,” have been reported. An optimizer called stochastic gradient descent (SGD) has been widely used, and several optimizers that diverged slightly from SGD have been developed recently. Although the details of each optimizer are beyond the scope of this paper, it should be noted that most optimizers update the weights iteratively based on their unique updating rules at the learning rate. Large learning rates increase the risk of oscillation of E , and small learning rates increase the risk that E will not converge within realistic iteration times. Increasing K' , N_{neuron} and the number of hidden layers also make the optimization problem difficult due to the increased number of local minimums although it enables segmentation of Mars images into several classes based on slight differences in local patterns by extremely fine parameter tuning.

Evaluation indices

Here, we briefly introduce the five indices frequently used for evaluating the performance of a classifier for the understanding of readers (see Fawcett 2006 for details). The confusion matrix shown in Fig. 7 defines true positive, true negative, false positive, and false negative. For example, pixels recognized as dust storm after the binarization but that are outside the true dust storm areas defined by a human operator are called false positive pixels. The numbers of true positive, true negative, false positive, and false negative pixels are referred to as TP , TN , FP , and FN , respectively. In addition, the true positive rate, false positive rate, precision, and recall are defined as follows:

$$\text{True positive rate} = \frac{TP}{TP + FN}, \quad (11)$$

		predicted class	
		dust	non-dust
True class	dust	True Positive: TP	False Negative: FN
	non-dust	False Positive: FP	True Negative: TN

Fig. 7 The confusion matrix defining true positive, true negative, false positive, and false negative. True class means the ground truth based on the human operator’s inspection. Predicted class means a label based on the classifier trained in this study

$$\text{False positive rate} = \frac{FP}{FP + TN}, \quad (12)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (13)$$

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (14)$$

A harmonic mean of precision and recall is referred to as an F -measure. It should be noted that improving either precision or recall results in a decrease in the other. The easiest way to improve recall is to develop an algorithm that regards the entire image as dust storm and minimizes precision. It clearly makes no sense to use such an algorithm. Therefore, the F -measure, which is kept low if either of precision or recall is significantly low, has been another index to measure the performance of the classifier. The F -measure, however, still depends on a threshold value for binarization and we have had a difficulty using the F -measure to compare different classifiers trained using different parameters and features. Here, the trade-off relationship between FP and FN should be noted. Considering that an increase in TP is equivalent to decrease in FN , TP can be increased by decreasing the threshold value and increasing FP . In Eqs. (11) and (12), $TP + FN$ and $FP + TN$ are constants indicating the numbers of pixels assigned the labels “dust storm” and any others by a human operator, respectively. Therefore, an ROC curve shows the relation between true positive rate and false positive rate equivalent to that between TP and FP when continuously changing the threshold value for binarization. Assuming that a drastic improvement in the true positive rate that incurs a slight increase in the false positive rate is desirable (the left part of the ROC curves shown in Fig. 3), AUC is a reasonable evaluation index independent of the threshold value.

Additional files

Additional file 1: Figure S1. Images used for preparing the background image. (PDF 2866 kb)

Additional file 2: Figure S2. Subtracted images in the red band and ground truth data of dust storms and clouds used for training the classifier. Each row shows a subtracted image (1st column), the dust storm area (2nd column), and the cloud area (3rd column). The ground truth data are based on the authors’ visual inspection. (PDF 2600 kb)

Additional file 3: Figure S3. Subtracted images and ground truth data of dust storms and clouds used for the evaluation of the algorithm. Each row shows a subtracted image (1st column), the dust storm area (2nd column), and the cloud area (3rd column). The ground truth data are based on the authors’ visual inspection. (PDF 2485 kb)

Abbreviations

AUC: Area under the curve; ISIS: Integrated Software for Imagers and Spectrometers; MARCI: Mars Color Imager; MGS: Mars Global Surveyor; MLP: Multilayer perceptron; MOC: Mars Orbiter Camera; MRO: Mars

Reconnaissance Orbiter; NN: Neural network; PCA: Principal component analysis; ROC: Receiver operator characteristics; SGD: Stochastic gradient descent; SVM: Support vector machine; TES: Thermal Emission Spectrometer; USGS: U.S. Geological Survey; WA: Wide angle; WIC: Water ice cloud

Acknowledgements

We thank Dr. Wataru Sunayama and Dr. Yuji Hatanaka for useful comments on this work. This research has made use of the USGS Integrated Software for Imagers and Spectrometers (ISIS3). This study was partially supported by the Sumitomo Foundation.

Funding

This study was partially supported by the Sumitomo Foundation.

Availability of data and materials

The dataset supporting the conclusions of this article is available in the USGS repository (https://ida.wr.usgs.gov/d_access.htm).

Authors' contributions

RG prepared the ground truth data, trained the classifier, and evaluated the performance of the algorithm using test images. RG also wrote all source codes used for this study and generated all figures in this article. KO downloaded all images used for this study and converted them to longitude-latitude images of reflectance and illumination angle. KO also proposed the topic and the procedure of the pre-processing and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Graduate School of Engineering, University of Shiga Prefecture, 2500, Hassaka, Hikone, Shiga 522-8533, Japan. ²School of Engineering, University of Shiga Prefecture, 2500, Hassaka, Hikone, Shiga 522-8533, Japan.

Received: 31 May 2018 Accepted: 23 January 2019

Published online: 14 February 2019

References

- Bishop CM (2006) Pattern recognition and machine learning. New York: Springer Science+Business Media
- Cantor BA, James PB, Caplinger M, Wolff MJ (2001) Martian dust storms: 1999 Mars Orbiter Camera observations. *J Geophys Res* 106:23653–23687
- Coupé P, Manjón JV, Fonov V, Pruessner J, Robles M, Collins DL (2011) Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. *Neuroimage* 54:940–954. <https://doi.org/10.1016/j.neuroimage.2010.09.018>
- Fawcett T (2006) An introduction to ROC analysis. *Pattern Recogn Lett* 27:861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Guzewich SD, Toigo AD, Kulowski L, Wang H (2015) Mars Orbiter Camera climatology of textured dust storms. *Icarus* 258:1–13. <https://doi.org/10.1016/j.icarus.2015.06.023>
- Guzewich SD, Toigo AD, Wang H (2017) An investigation of dust storms observed with the Mars Color Imager. *Icarus* 289:199–213. <https://doi.org/10.1016/j.icarus.2017.02.020>
- Hatanaka Y, Samo K, Ogohara K, Sunayama W, Muramatsu C, Okumura S, Fujita H (2018) Automated blood vessel extraction based on high-order local autocorrelation features on retinal images. In: Tavares JMRS, Natal Jorge RM (eds) *VipIMAGE 2017, ECCOMAS, vol 2017. Lecture Notes in Computational Vision and Biomechanics*. Springer, Cham, pp 803–810
- Howell ES, Merenyi E, Lebofsky LA (1994) Classification of asteroid spectra using a neural network. *J Geophys Res Planets* 99:10847–10865. <https://doi.org/10.1029/93JE03575>
- Ishii T, Simo-Serra E, Iizuka S, et al. (2016) Detection by classification of buildings in multispectral satellite imagery. In: 2016 23rd International Conference on Pattern Recognition (ICPR) 3344–3349.

- Jiang T-X, Huang T-Z, Zhao X-L, Ma T-H (2017) Patch-based principal component analysis for face recognition. *Comput Intell Neurosci*:1–9. <https://doi.org/10.1155/2017/5317850>
- Kingma DP, Ba J (2015) Adam: a method for stochastic optimization. <https://arxiv.org/abs/1412.6980>. the 3rd international conference for learning representations, San Diego
- Kulowski L, Wang H, Toigo AD (2017) The seasonal and spatial distribution of textured dust storms observed by Mars Global Surveyor Mars Orbiter Camera. *Adv Space Res* 59:715–721. <https://doi.org/10.1016/j.asr.2016.10.028>
- Maeda K, Ogawa T, Haseyama M (2015) Automatic Martian dust storm detection from multiple wavelength data based on decision level fusion. *IPSP Trans Comput Vis Appl* 7:79–83. <https://doi.org/10.2197/ipsjtva.7.79>
- Malin MC, Edgett KS, Cantor BA, et al. (2010) An overview of the 1985–2006 Mars Orbiter Camera science investigation. *Mars* 5:1–60. <https://doi.org/10.1555/mars.2010.0001>
- Ogohara K, Munetomo T, Hatanaka Y, Okumura S (2016) Automated detection of Martian water ice clouds using Support Vector Machine and simple feature vectors. *Planet Space Sci* 134:64–70. <https://doi.org/10.1016/j.pss.2016.10.009>
- Turk M, Pentland A (1991) Eigenfaces for recognition. *J Cogn Neurosci* 3:71–86. <https://doi.org/10.1162/jocn.1991.3.1.71>
- Venkataraman ST, Gulati S, Barhen J, Toomarian N (1993) A neural network based identification of environments models for compliant control of space robots. *IEEE Trans Robot Autom* 9:685–697. <https://doi.org/10.1109/70.258059>
- Wang H (2007) Dust storms originating in the northern hemisphere during the third mapping year of Mars Global Surveyor. *Icarus* 189:325–343. <https://doi.org/10.1016/j.icarus.2007.01.014>
- Wang H, Ingersoll AP (2002) Martian clouds observed by Mars Global Surveyor Mars Orbiter Camera. *J Geophys Res* 107:1–16. <https://doi.org/10.1029/2001JE001815>
- Wang H, Toigo AD, Richardson MI (2011) Curvilinear features in the southern hemisphere observed by Mars Global Surveyor Mars Orbiter Camera. *Icarus* 215:242–252. <https://doi.org/10.1016/j.icarus.2011.06.029>

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com