New Journal of Physics
An Institute of Physics and Deutsche Physikalische Gesellschaft Journal

Identification of clouds and aurorae in optical data images

R Seviour[1,2], M Kosch[1] and F Honary[1]
[1] Department of Communication Systems, Lancaster University, Lancaster LA1 4YR, UK
[2] Department of Imaging Science and Biomedical Engineering, Manchester University, Manchester M13 9PT, UK
E-mail: Rebecca.Seviour@isbe.man.ac.uk

New Journal of Physics **5** (2003) 6.1–6.7 (http://www.njp.org/)
Received 25 September 2002
Published 17 January 2003

Abstract. In this paper we present an automatic image recognition technique used to identify clouds and aurorae in digital images, taken with a CCD all-sky imager. The image recognition algorithm uses image segmentation to generate a binary block object image. Object analysis is then performed on the binary block image, the results of which are used to assess whether clouds, aurorae and stars are present in the original image.

The need for such an algorithm arises because the optical study of particle precipitation into the Earth's atmosphere by the Ionosphere and Radio Propagation Group at Lancaster generates vast data-sets, over 25 000 images/year, making manual classification of all the images impractical.

Knowledge of the energy spectrum for particles precipitating into Earth's high-latitude atmosphere is fundamental for understanding processes that occur in the magnetosphere. Riometer absorption [1] yields information about the high end (>25 keV) [2] of the energy spectrum, whereas optical observations of auroral emissions yield information on the low end (<15 keV) [3] of the energy spectrum (see [4] for information on the energy spectrum). To study the optical auroral emissions the Ionosphere and Radio Propagation Group at Lancaster uses optical data from a digital all-sky imager (DASI), at 557.7 nm, located at Skibotn in Norway (69.35°N, 20.36°E). DASI operates only during the winter for all dark and moon-free periods, with an integration time of 10 s and a spatial resolution of 10×10 km. The images give a coverage for an area of 520×520 km, when projected onto a geographic grid at 100 km altitude [5].

To ensure that DASI measures the optical intensity correctly, photometrically clear viewing is required, i.e. without any scattering or absorption from any form of cloud. As DASI uses a

PII: S1367-2630(03)54104-0
1367-2630/03/000006+7$30.00

narrow bandpass filter (557.7 nm) most stars do not appear in the DASI images, making DASI ineffective for cloud detection. To identify clouds, a second camera at Skibotn, a slow-scan CCD all-sky imager (SCASI), is used.

SCASI consists of a commercial amateur astronomy CCD camera, the ST6 camera from the Santa Barbara Instrument group with a 50 mm F1.3 C-mount focusing lens which gives a field of view of $\sim 7^{\circ}$. This field of view is widened to 180° using a combination of two mirrors. Scientifically, the optical system is limited, as the secondary mirror blocks the local zenith, blinding the imager to this region. The camera uses a front-illuminated TC241 CCD chip, with 375×242 pixels, with a pixel size of 23×27 μm. The camera is fitted with an on-board microprocessor with memory, a mechanical shutter, temperature-regulated Peltier cooling and a serial port. The large pixel size makes the camera ideal for low-light-level applications, although the use of serial communications limits the time resolution for data acquisition.

Typically, a full spatial resolution image, with a 30 s exposure, is taken once every 3 min. The 16-bit dynamic range of the CCD ensures that the majority of aurorae are properly recorded along with all stars visible to the naked eye. When the exposure is completed, the image is digitized into the ST6 memory using double correlation. This means that each pixel is read twice while the image is slow-scanned out of the CCD, ensuring a much better image quality. The camera automatically subtracts a dark image (thermionic noise) from the scanned image. The dark image, an image taken with the shutter closed, is taken at the start of recordings and stored. In this way the data images are compensated for thermal non-uniformities of the CCD. The resulting data image is downloaded, stored and then assessed for cloud cover. Although, as previously stated, photometrically clear viewing is required for detailed analysis of auroral activity, in practice if the cloud is not perceptible in the CCD image then the data are acceptable for use.

Generally, CCD images are assessed for cloud cover and the presence of aurorae by manual inspection; this is a both slow and laborious method, prone to error. Recently, various research groups have started to investigate the use of imaging processing techniques to automate cloud detection. The fractal dimensionality of objects in CCD images was considered in [6] to identify clouds and aurorae using self-organized neural networks. Although the initial results look promising, they are not conclusive, and further work in this area is required. In [7], skeletal shape analysis using batch-mode minimum-spanning tree self-organizing maps, was used, successfully, to identify aurorae. However, this technique does not identify cloud and only identifies east–west auroral arcs.

In this paper we present a PC-based image recognition algorithm which automatically assesses CCD image data for clouds, stars and aurorae. We considered several techniques for providing identification algorithms. One such was sequential temporal image analysis, which was rejected due to the poor temporal resolution of the CCD data. Using certain stars as landmarks was rejected due to non-linearities in the plastic dome covering SCASI; even when this was compensated to second order, the positions of stars were still imprecise. Also, as SCASI is located on permafrost, its position can change. The image recognition algorithm that we present in this paper is a variation of the shape analysis commonly used in soil image recognition [8].

The algorithm presented here can be split into two distinct parts: binary image creation and object analysis. The binary image creation algorithm creates a binary block image from the SCASI CCD image; this binary image is then used by the object analysis algorithm to decide whether the SCASI CCD image contained cloud or aurora.

Binary image creation consists of two distinct subsections: the first section identifies the area of the SCASI CCD image to be used; and the second section segments the CCD image to

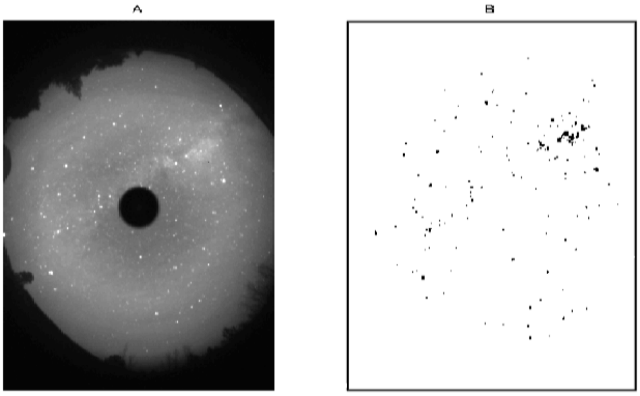

Figure 1. (A) A SCASI CCD image, showing a clear sky. (B) The segmented version of image (A). The greyscale of the CCD image has been non-linearly contrast stretched purely for printing purposes.

build a binary image used for object analysis.

As the SCASI CCD image contains the secondary mirror and the horizon, as can be seen in the centre and at the edge of figure 1(A), it is necessary to identify the *sky* region of the image. The relative position of the horizon and secondary mirror in the CCD image can change, during maintenance of the camera. The image area to be used is found by determining the maximum intensity value (L_c) of the first row of each image, which does not contain real all-sky image data (the CCD image is square and the all-sky image is round). Any pixel in the image whose value is less than L_c is discarded.

The image segmentation is achieved using a single-level threshold [8] to generate a binary block object image. However, due to the large luminance intensity variation from image to image, the threshold level used must be different for each image. By trial and error, the most appropriate threshold level was found to be given by

$$L = \langle I \rangle + \sigma \tag{1}$$

where L is the image threshold level, $\langle I \rangle$ is the mean intensity of the image and σ the standard deviation of the image intensity. This threshold level is then used to create a binary image from the CCD image. CCD image pixel values greater than or equal to L are set to one, and pixel values less than L are set to zero. This creates a binary block image where the top 5% of intensity values from each SCASI image form the discrete objects in the binary image. Figures 1–3 show examples of SCASI CCD images and the associated binary image. The greyscales of the SCASI CCD images have been non-linearly contrast stretched purely for printing purposes.

Figure 1(A) shows a CCD SCASI image from a clear sky period; figure 1(B) shows the binary image created from figure 1(A) using the above technique. Immediately, from figure 1(B) we can see that there are a number of small distinct objects, which relate to the stars seen in figure 1(A).

Figure 2(A) shows a CCD SCASI image of a clear sky with an auroral arc going from east to west; figure 2(B) is the binary image created from figure 2(A). We see in figure 2(B) a number of small objects, which relate to the stars, and a well defined strong region, which relates to the auroral arc seen in figure 2(A). The presence of the secondary mirror in the SCASI image splits the auroral arc in two.

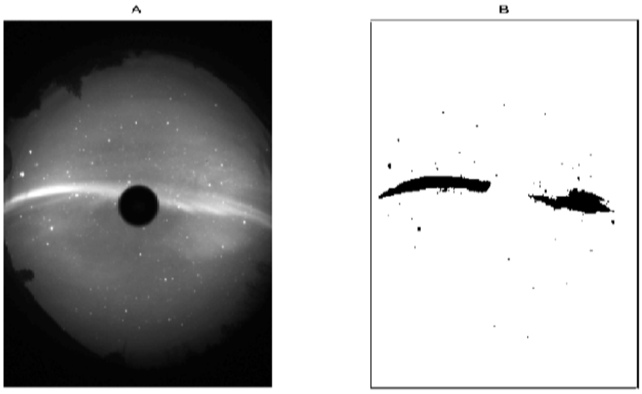

Figure 2. (A) A SCASI CCD image, showing an auroral arc in a cloud-free sky. (B) The segmented version of image (A). The greyscale of the CCD image has been non-linearly contrast stretched purely for printing purposes.

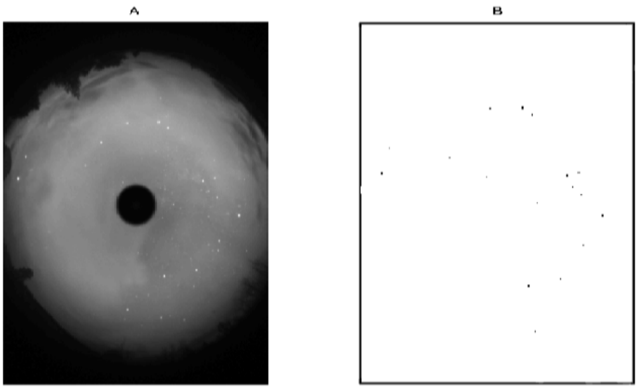

Figure 3. (A) A SCASI CCD image showing cloud. (B) The segmented version of image (A). The greyscale of the CCD image has been non-linearly contrast stretched purely for printing purposes.

Figure 3(A) shows a CCD SCASI image of a cloudy sky; figure 3(B) shows the binary image created from figure 3(A) using the thresholding technique. In figure 3(B) we can see that the number of distinct objects is less than that shown in figure 1(B). This is due to cloud in the image obscuring the stars.

The second part of the algorithm performs object analysis on the binary block image, counting the number of objects and the size of each discrete object. These data are then compared to a statistical, empirically derived, table to decide which category the image falls in. Images that are borderline or do not conform to the empirical data are flagged as unidentified and are left for manual identification.

The empirical look-up table was derived by the manual identification of just over 3000 SCASI CCD images, taken during March, October, November and December 1998. Each image was classified as: clear sky (no cloud or auroral activity present); aurora with no cloud; or cloudy. For each CCD image in the above data-set the associated binary image was created and analysed according to the manual classification assigned to each image. For each binary image the number of distinct objects present and the size of each object were recorded.

Figure 4 shows the binary object statistics for 4000 randomly chosen SCASI images from January 2001; each point on each graph relates to the properties of a single SCASI image. The first row of graphs in figure 4 relates to images identified by the algorithm as containing aurorae and no cloud. The second row relates to images that the algorithm identified as clear sky. The third row relates to images that the algorithm identified as containing cloud. The first graph in each row shows the number of distinct binary objects in an individual image, the second graph in each row shows the mean size of distinct binary objects in each image and the third graph in each row shows the maximum object size in each binary image. From figure 4 we see clearly how the distinction between clear, cloudy and images containing aurorae can be made. For CCD images that have clear sky in associated binary images (the second row of figure 4) there are a relatively large number of small objects per image; these of course relate to the stars in the CCD image.

For CCD images that contain aurorae the associated binary images (the first row of figure 4) have both a large number of small objects (the stars) and a small number of large objects. These large objects relate to the aurora in the CCD image (see figure 1(A)). For the binary images derived from CCD images containing cloud (the third row of figure 4) the situation becomes a little more complicated: if there is little or no illumination, then the binary image contains very few objects and what objects are present are very small (see figure 3(A)). If a source of illumination is present, then the clouds may become illuminated; for example, this can occur if we have aurorae occurring behind the clouds. In this case we can have a high number of large objects, but as clouds are relatively large in the night sky compared to stars, the mean and maximum object sizes in binary images containing cloud are generally much larger than for a clear sky or an image containing aurorae.

Using data such as those presented in figure 4, it is possible to define a set of criteria from which a series of classifiers can be built. These classifiers can then be used to categorize CCD images as: clear; cloudy; or containing aurorae and no cloud. For example a binary image containing 80 objects with a mean object size of three and a maximum object size of 10 relates to a clear sky CCD image.

One can see from figure 4 that there are a small number of images (approximately 10%) whose binary image objects seem to span across the classifiers given above of clear, cloudy and with aurorae, i.e. images where there appear to be a large number of relatively large objects. For these images, predominantly taken during auroral substorms, the classification by the analysis presented above is unclear as to whether the image contains a large amount of dispersed auroral activity or contains dispersed auroral activity occurring behind cloud. For the small number of images that overlap the three classifiers—predominantly images containing diffuse cloud back-illuminated by aurorae—it is necessary to perform manual categorization. These images are easily identified when more than one classifier positively identifies them, i.e. where a binary image contains 100 objects, with a mean object size of 50 and a maximum object size of 90.

To assess the reliability of the object analysis criteria we used the criteria to automatically categorize a further 7000 SCASI images also taken from 1998. All 7000 images were then identified manually and compared with the classification criteria for each image. Table 1 shows the results of this check: of the 7000 images, 90.1% were identified correctly, and not one image was incorrectly identified; the main cause of uncertainty for the remaining 9.9% of images was aurorae illuminating clouds.

An algorithm which automatically identifies cloud, aurorae and clear sky in all-sky monochrome images has been developed. The algorithm that we presented here can clearly

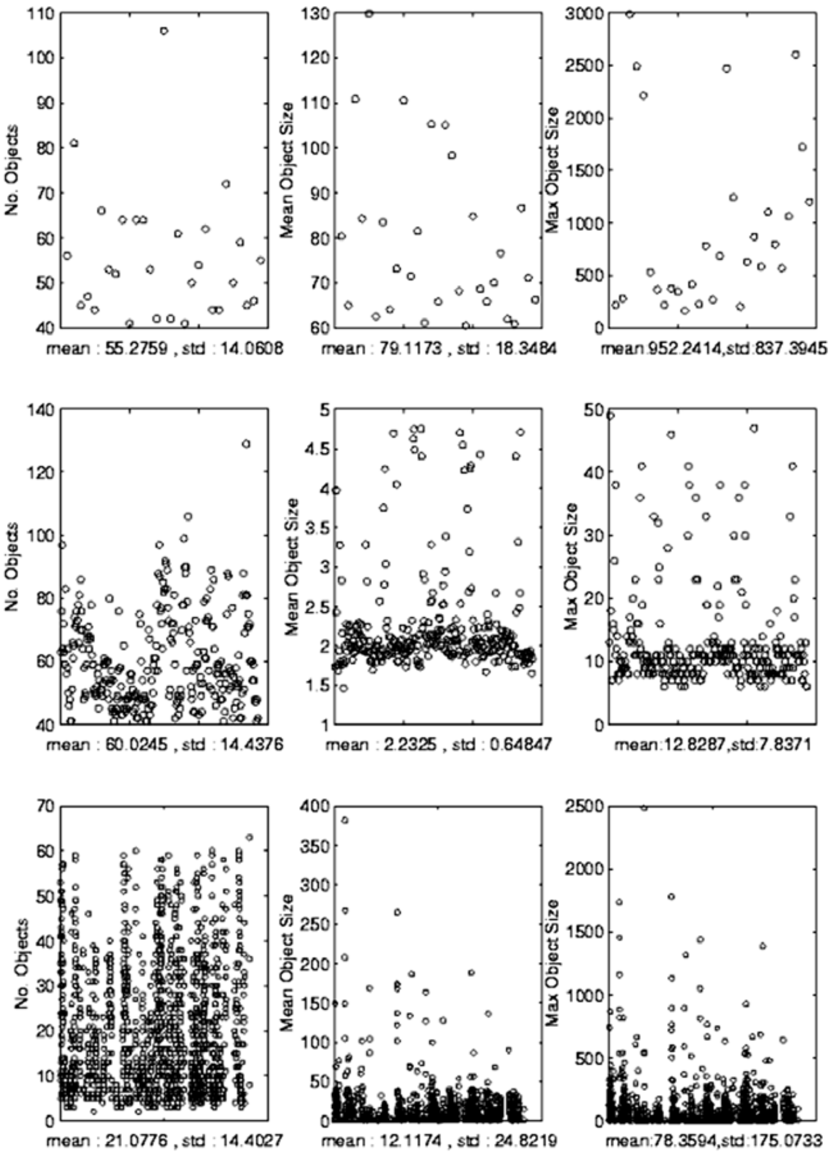

Figure 4. The binary object statistics for 4000 randomly chosen SCASI images, taken from January 2001; each point relates to the properties of a single SCASI image. The first row relates to images where the algorithm identified aurorae and no cloud. The second row relates to images that the algorithm identified as clear sky. The third row relates to images that the algorithm identified as containing cloud. The first graph in each row shows the number of distinct binary objects for each individual image, the second graph in each row shows the mean size of distinct binary objects for each image and the third graph in each row shows the maximum object size in an individual image.

Table 1. The comparison of the binary object criteria analysis used to automatically categorize 7000 SCASI images and the manual identification of each of the images. The images are taken from 1998.

Identified correctly	90.1%
Manually identified	9.9%
Misidentified	0%

distinguish between clouds and aurorae, unlike the techniques presented in [6]. Also our technique can detect all auroral activity, whereas the method presented in [7] is limited to recognizing east–west auroral arcs. The technique that we present greatly reduces the human input required when scientifically analysing auroral images, identifies periods of observations without cloud contamination and enhances the scientific value of auroral images taken with narrow-band optical filters. Due to the 90% success rate of the object analysis criteria, we have automated the process to classify over 125 000 SCASI images; it takes on average 20 s to classify an image. Further improvements are being investigated with the aim of reducing the number of images which require manual identification. As the main causes of the ambiguity of these images is aurorae illuminating cloud, it should be possible to add another stage to the object analysis routine, to analyse the edges of objects in the binary block image, as aurorae generally have well defined edges and illuminated clouds are fairly diffuse.

References

[1] Stanning P 1996 *J. Atm. Terr. Phys.* **58** 765–83

[2] Burch J L 1991 *Auroral Physics* (Cambridge: Cambridge University Press) pp 765–83

[3] Collis P N, Hargreaves J K and Korth A 1984 *J. Atm. Terr. Phys.* **46** 21–38

[4] Kosch M J, Honary F, del Pozo C F, Marple S R and Hagfors T 2001 *J. Geophys. Res.* A **106** 28925–37

[5] Kosch M J, Hagfors T and Nielsen E 1998 *Rev. Sci. Instrum.* **69** 578–84

[6] Alpatov V, Matveev A and Kosch M 2001 *Proc. ESA SP-471* (Noordwijk: European Space Agency) p 635

[7] Syrjăsuo M, Kauristie K and Pulkkinen T 2000 *Proc. Signal Image Processing Conf.* (Calgary: ACTA Press) p 381

[8] Luo Daisheng 1998 *Pattern Recognition and Image Processing* (New York: Harwood)