

Clusters in Hyper-Cubic Multi-Channel Satellite Imagery

K.A. Hawick

Computer Science, Massey University, Albany, North Shore 102-904, Auckland, New Zealand

<http://complexity.massey.ac.nz>



Remote Sensing

Multi-spectral remotely-sensed data such as satellite imagery can yield excellent insights into complex phenomena such as weather systems. Analysing the multi-channel space to separate out different features still presents a challenge, which will increase with the availability of hyper-spectral satellites. We use component labelling and population thresholding techniques to separate out clusters in hyper-dimensional channel space and use this information to identify different cloud types in geostationary satellite imagery[1]. Three dimensional visualisation techniques are used to study the hyper-dimensional channel population data.

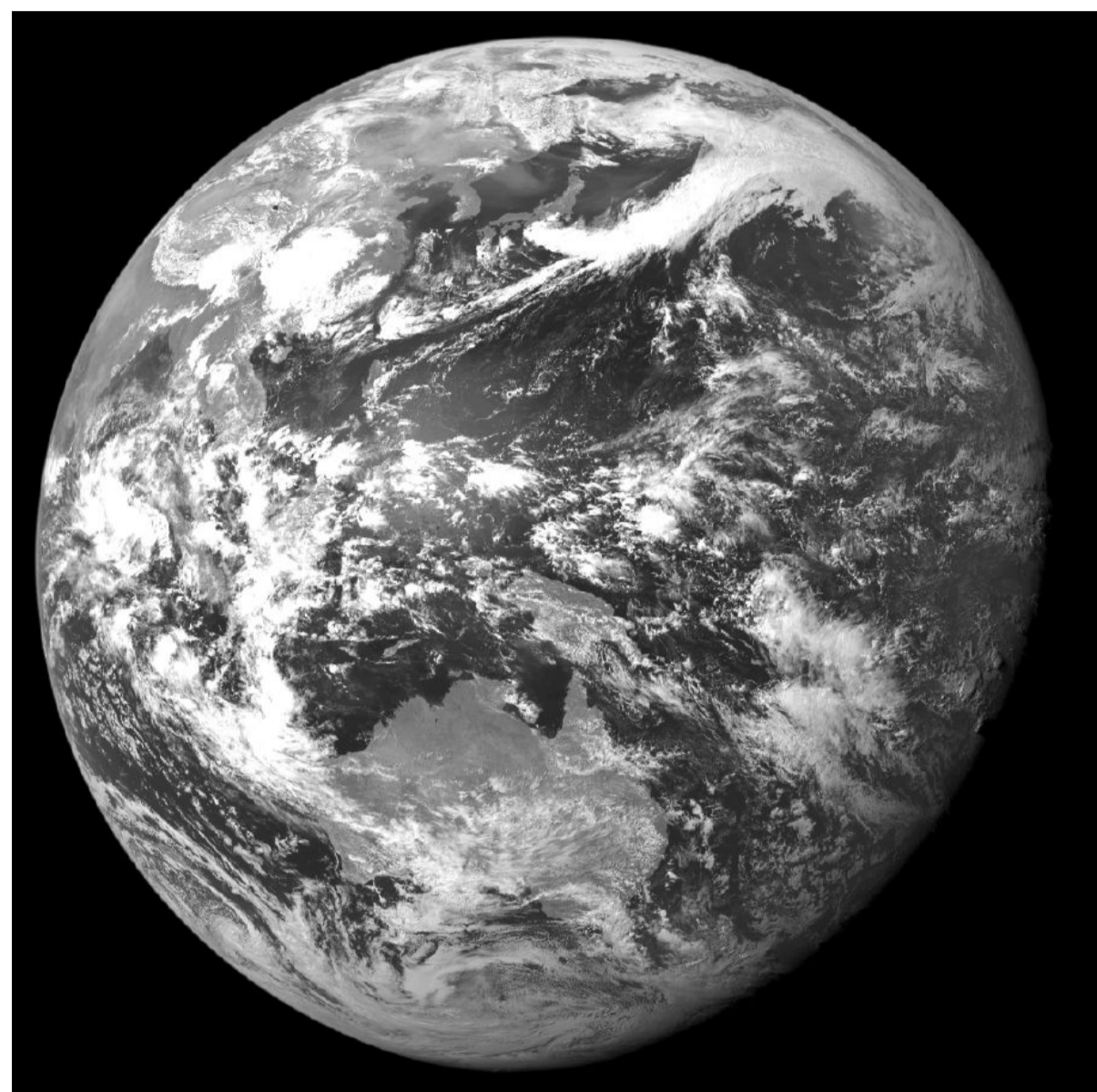


Figure 1: Full-disk Earth image obtained from GMS5.

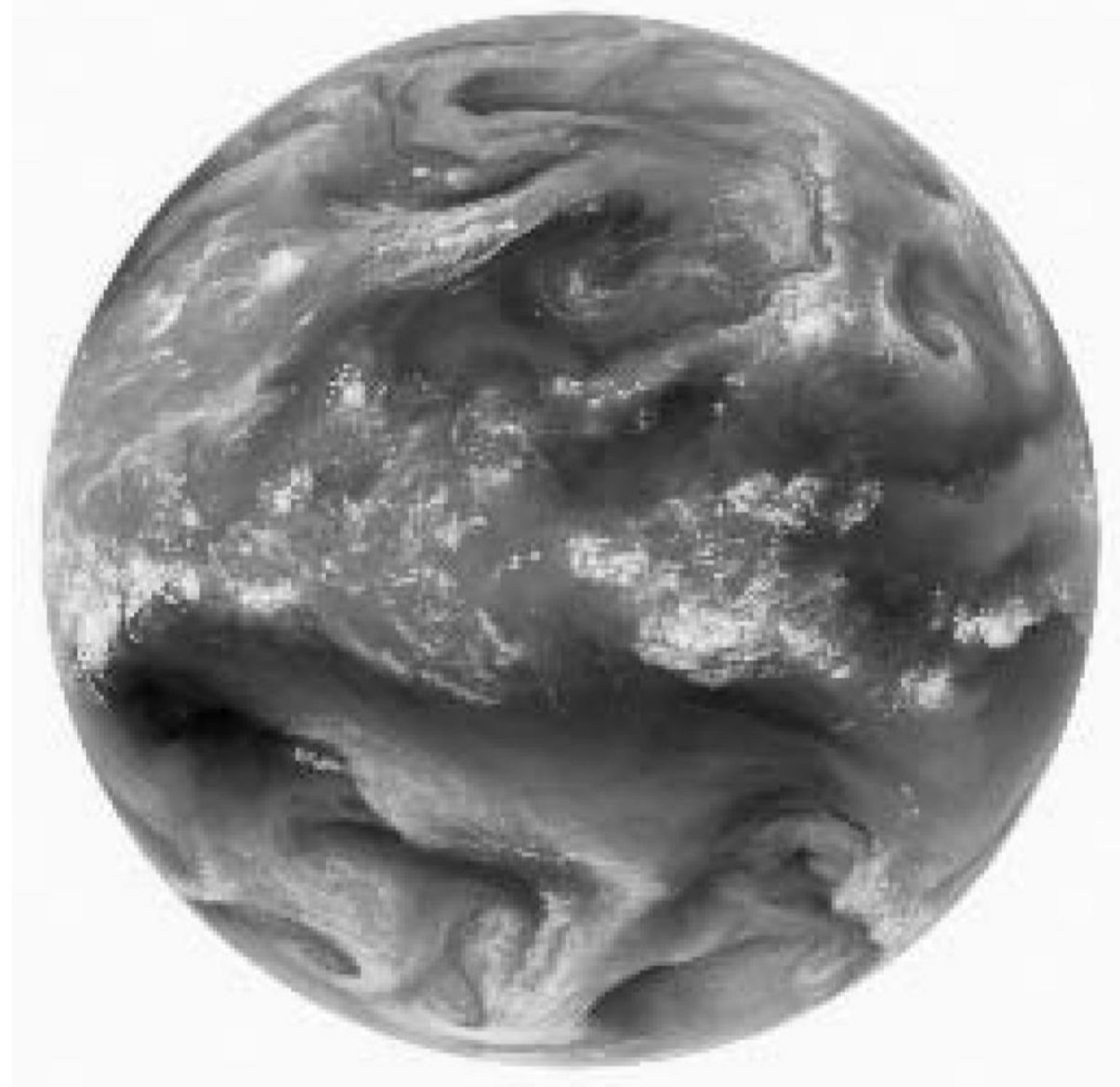


Figure 2: GMS-5 Data: Water Vapour Channel (obtained at 30 June, 1996 2332GMT)

Assimilating data from multiple remotely-sensed sources[2] is an important aspect of modern numerical weather prediction codes[3]. Data come from a variety of sources but one that has become dominant in recent years is the multi-spectral imagery obtained from geostationary satellites. Weather prediction itself remains a challenging problem, not least due to the chaotic and stability effects inherent in weather systems. Nevertheless there has been considerable success in recent years in monitoring weather systems, analysing cloud patterns[4] and tracking features, which can then be incorporated as near live corrections to weather simulation codes. Analysing cloud cover is an important aspect of monitoring albedo effects and other phenomena which feed into longer term climate change studies.

There have been good studies reported in the literature on deriving cloud-cover patterns and statistics from various satellite sources[5] including work using data from the now defunct, but previously very successful GMS5 Japanese weather satellite[6]. Regularly sampled full-disk Earth image data from GMS5 was made widely available during its lifetime and we use some of that data collection[7] to explore some multi-spectral channel space analysis techniques including cluster component labelling, 3D visualisation and population thresholding to separate out clusters linked to image features.

Figure 1 shows a typical visible spectrum Earth full-disk image obtained from the GMS5 satellite. GMS5 provided four different spectral channels but modern successors such as the latest in the GOES series of satellites[8] already provide several more, and new generation satellites promise even more. This multi and hyper spectral data sources will require sophisticated analysis techniques to yield maximal value from the combined channel sources for land, sea, and atmospheric feature detection and statistical counting.

The work in this paper is based on multi-spectral imagery obtained from the Japanese GMS5 geostationary satellite. During its lifetime this satellite was an excellent source of regularly sampled full-disk Earth images of 2291×2291 pixel resolution. Other modern satellite sources such as the GOES series of satellites also yield excellent multi-spectral data and would be amenable to similar analyses as we describe in this paper. We give a brief description of the characteristics of GMS5 data as indicative of the data features and handling issues.

The Geostationary Meteorological Satellite (GMS) typically provided more than twenty-four full-disk hemisphere multi-channel images per day. This required approximately 204MBytes of storage capacity per day, or 75GBytes per year. Although this quantity of data was originally a challenge to manage[9, 10] it is no longer perceived as a great deal as modern satellites produce several times this amount and storage technology itself has advanced considerably since the satellite's start-of-lifecycle.

See: <http://www.massey.ac.nz/~kahawick/cstn/071/cstn-071.html> for more information.

GMS5 Satellite Imagery

The GMS-5 satellite was launched in June 1995 and provided visible and infra-red spectral data in various wavelength channels, measured using a Visible and Infra-Red Spin Scan Radiometer (VISSR). The satellite operated from an altitude of approximately 35,800km and a position almost directly over the equator at 140 degrees east. The pixel resolution is approximately 4km and the four channel images are 2291×2291 pixels in size. The raw pixel data, sampled at hourly intervals was made available at a NASA ftp site in Hierarchical Data Format (HDF) [11] and was subsequently archived by the DHPC research group at the University of Adelaide[12].

Figure 1 shows an image of the visible channel, which is quite different in character to the water vapour channel image in Figure 2. Two of the infra-red channels were chosen to be close to one another to allow for differential spectral analysis, but the third infra-red channel was chosen to coincide with spectral properties of water vapour and is therefore an excellent source of study for cloud pattern analysis.

The Visible and Infra-Red Spin Scan Radiometer (VISSR) recorded visible radiation through a photo-multiplier tube and infra-red radiation through a HgCdTe detector, using a scanning mirror system. Signals were quantized into 64 bits (visual) and 256 bits (infra-red) prior to transmission to an earth based receiving station. The satellite took approximately 27.5 minutes to record visual and IR data in a 20×20 degree area which includes the Earth disk image with approximately 2500 mirror scan steps. Imaging swathes are 4 km by 1 scan line for infra-red and 1.25 km by 4 scan lines for visual data. The satellite is spin-stabilized and scans are synchronised with the spin rate. The observed schedule is full earth disk images hourly. Variations in the observation schedule occurred during periods when the satellite was eclipsed, for periods of solar interference, for typhoon special observation periods and for occasional satellite maintenance periods.

A number of techniques were employed by the Adelaide research group in managing the data storage system for processing the data archive[13, 14]. Some tiling technologies for carrying out these operations are described in [15] which builds on earlier work on parallel data transforms [16]. These ideas have been further developed and are described in [17]. Generally, over a decade later, the data can be readily stored on a desktop workstation and a reasonable sequence of data can be manipulated in memory. This makes the multi-spectral analysis and the cluster component labelling techniques we discuss below feasible in interactive time, whereas these techniques would barely have been possible as overnight batch processing runs on technology available around 1995.

Multiple Channels

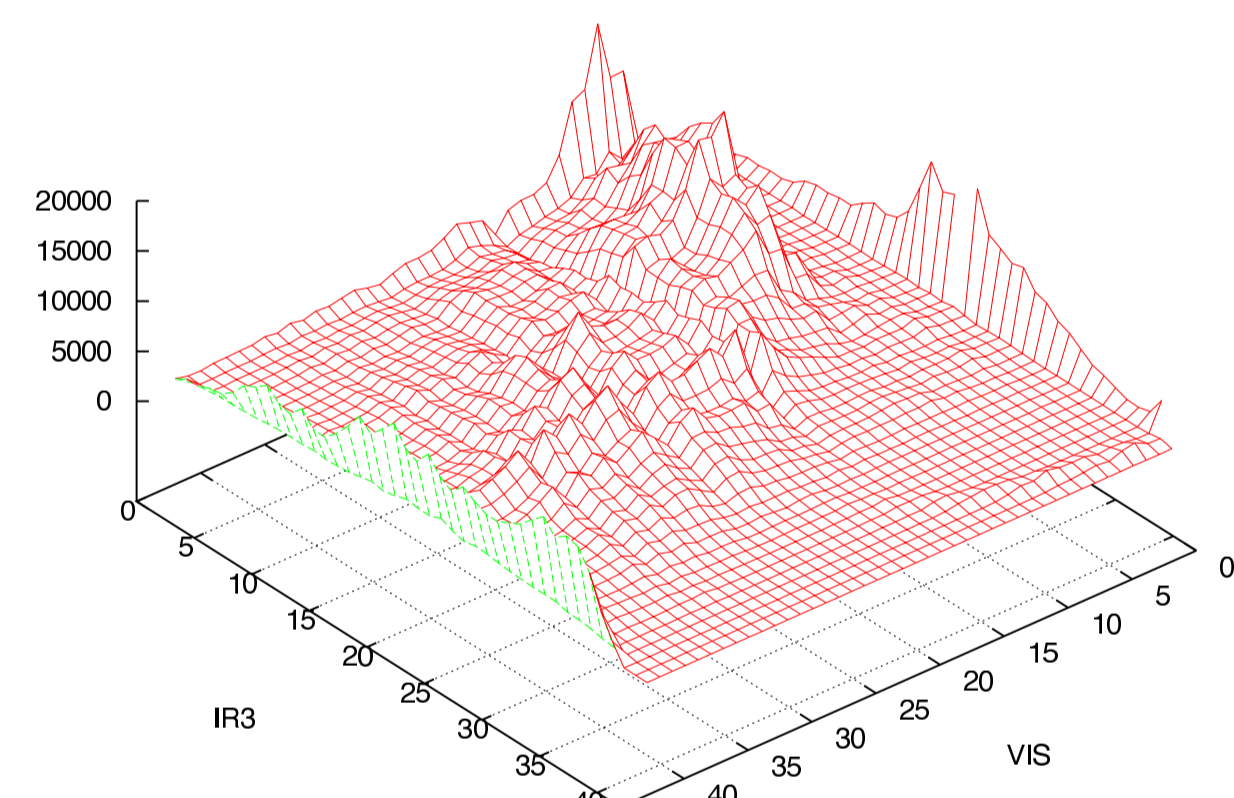


Figure 3: Frequency (pixel) Histogram of VIS vs IR3 for GMS5 0305090425 Dataset

A more interesting surface plot histogram arises from a combination of the visible and water-vapour channels, as shown in Figure 3. Some features we can identify simply. The narrow crest at $\text{vis}=0$ corresponds to the dark portion of the image that is space - outside the earth disk. Likewise, the crest at maximal $\text{vis}=41$ corresponds to where albedo is maximal and the visible sensor is saturated. Otherwise however, this is an uninteresting surface showing a number of features that are not easily separated by the eye. Cold, bright areas correspond to regions of cloud, with high albedo, where warmer regions of varying brightness correspond to different sorts of land or vegetative coverage.

One approach to visualising the multi-spectral data is to make a surface plot frequency histogram of two of the channels.

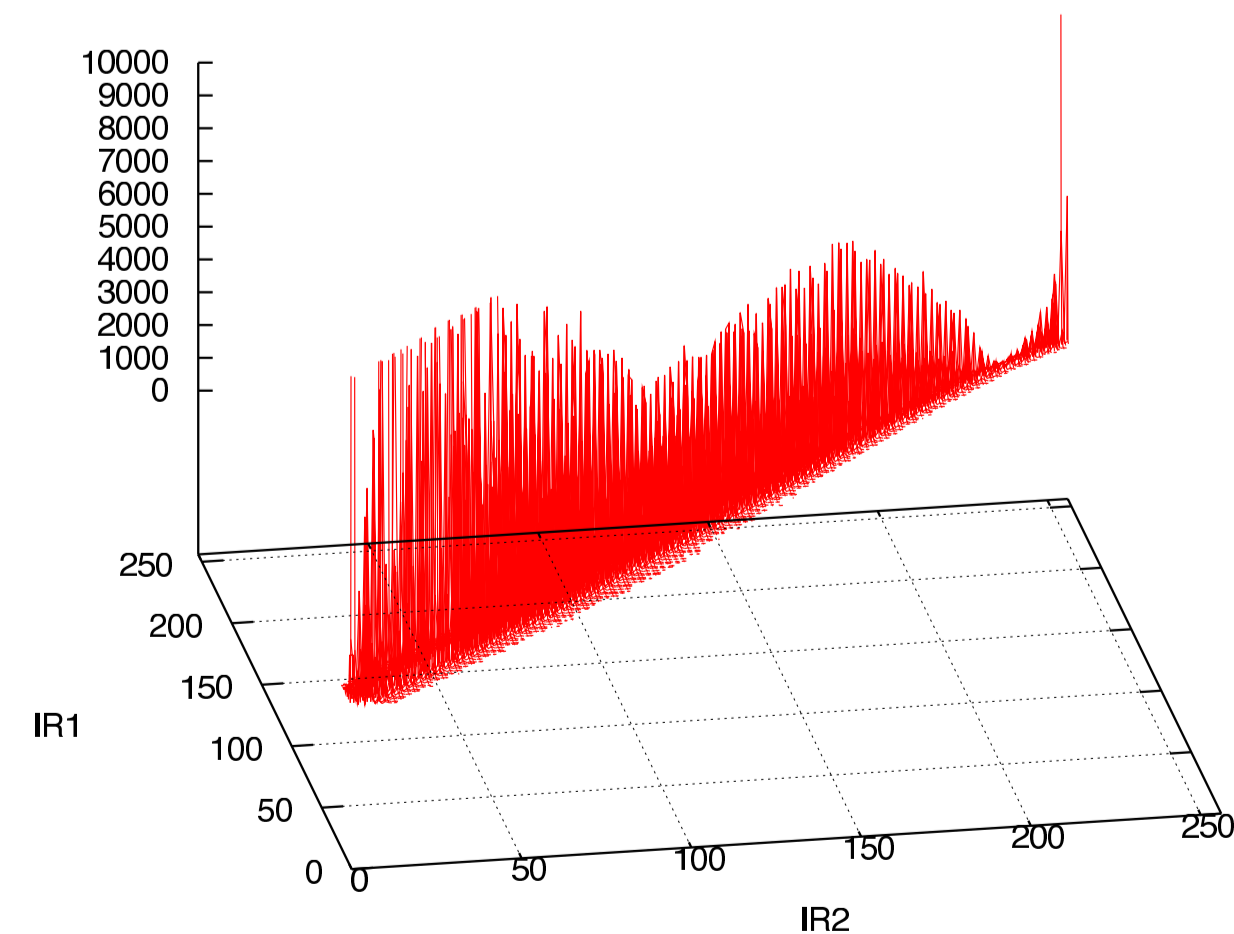


Figure 4: Frequency (pixel) Histogram of IR1 vs IR2 for GMS5 0305090425 Dataset

Figure 4 shows a frequency histogram of the pixel values that occur in the two differential IR1 and IR2 channels. These are very close to one another and the surface plot of the histogram shows they are very closely correlated. For our purposes we focus on combinations of Visible, IR1 and IR3, since IR1 and IR2 are so close.

Voxel Histograms

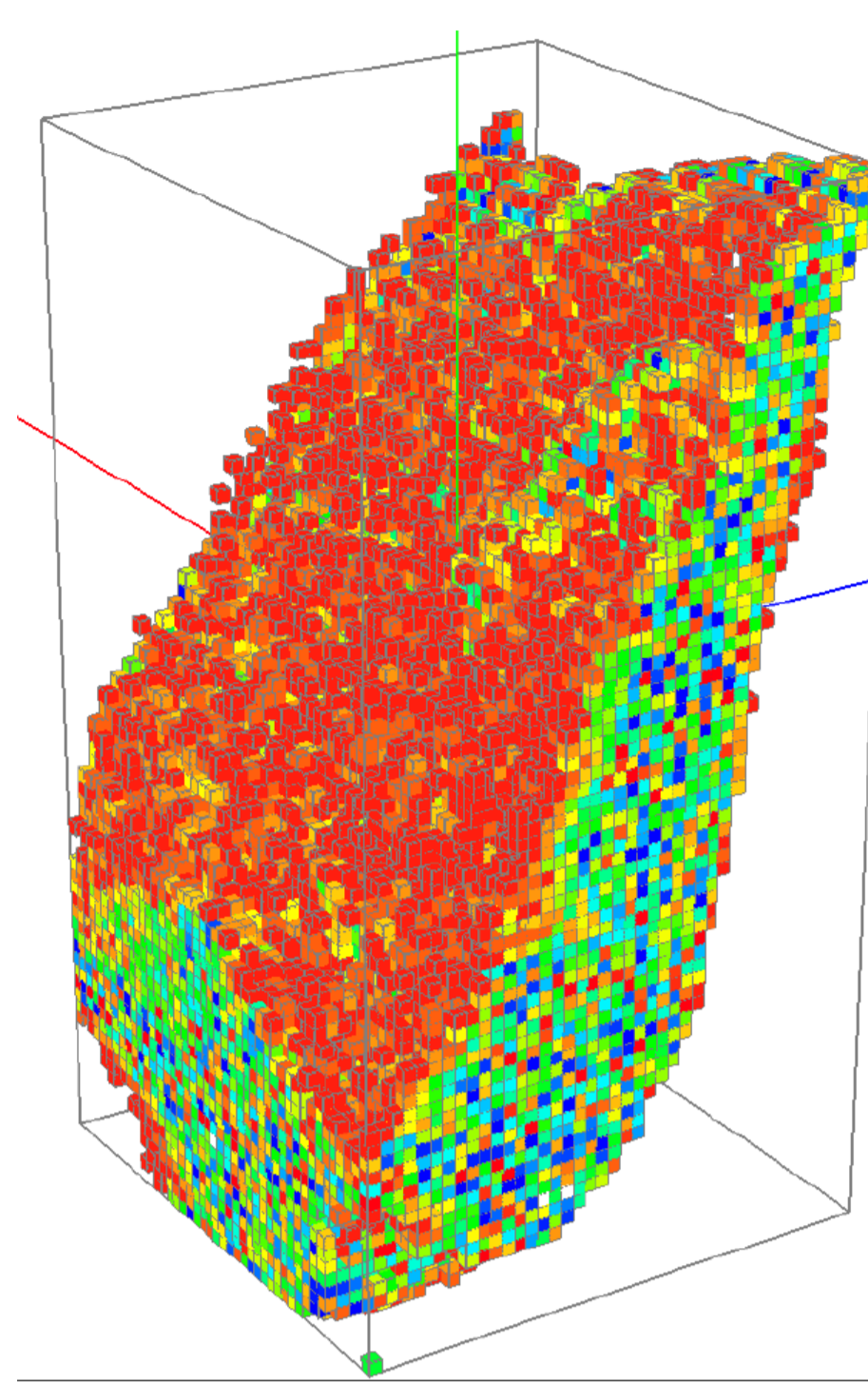


Figure 5: Rainbow log₂-coloured Frequency (pixel) Histogram Projection of Visible, IR1 and IR3 for GMS5 0305090425 Dataset. Red-Green-Blue axes correspond to increasing directions for Vis, IR1, IR3.

We can extend this surface histogram analysis to make use of three out of the four channels available. Figure 5 shows the visible, IR1 and IR3 channels combined in a 3D hyper-brick where the colours correspond to different histogram population values. We have used a rainbow palette to try to make the colour scheme intuitive. The histogram colour levels are however on a logarithmic scale. So red voxels correspond to rare combinations of the three channels that occur in the image set, blue correspond to very common combinations.

As can be seen the multi-channel voxels form a complex shaped clump. Once again not all combinations of the $42 \times 160 \times 43 = 288,960$ element space actually occur - and in fact less than one third of the space is populated. However it is still very difficult to identify regions or features as they all appear to interconnect into one super-clump of 87,105 voxels.

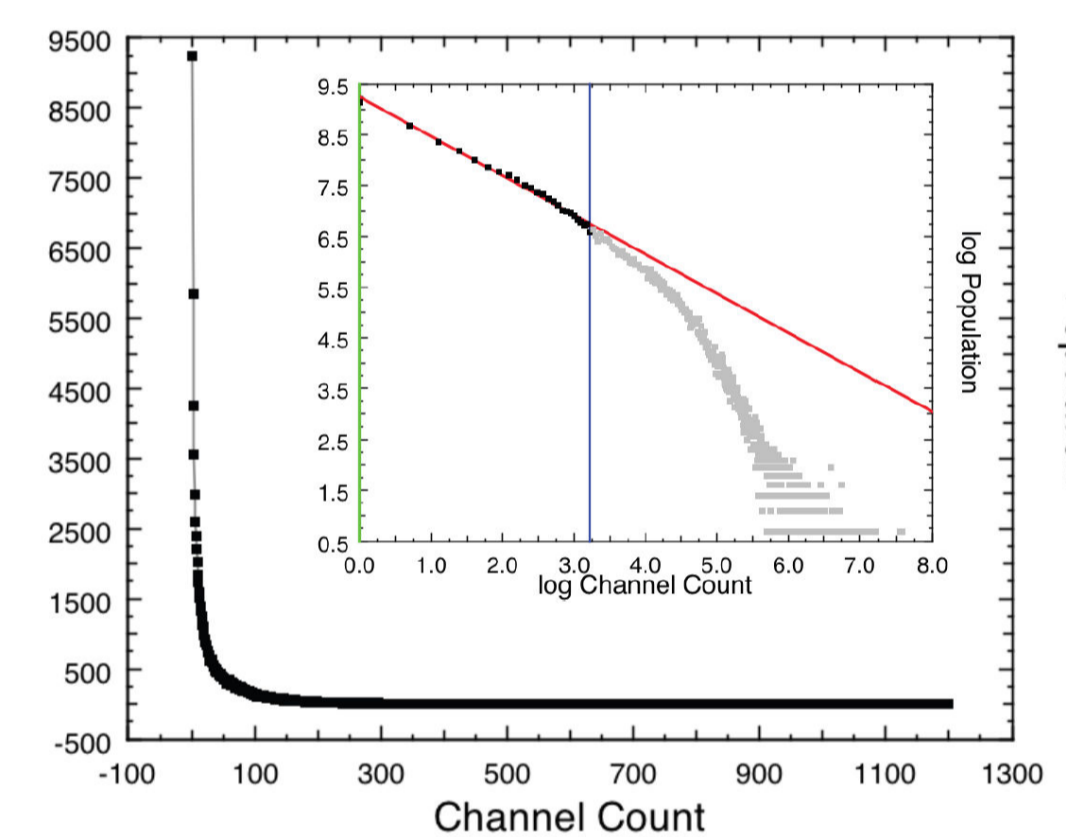


Figure 6: Population Frequency of the Vis; IR1;IR3 hyper-brick of channels. The insert shows the same data on a log-log scale.

We can examine the frequency population of this 3D space more carefully. Figure 6 shows the ranked occupation values. It shows that there is a considerable (logarithmic) variation between the most common combinations and the least likely to occur. The limiting behaviour of the histogram is a straight line when plotted on a log-log scale (shown in the insert.) This suggests that at least in the high frequency limit, there is a power-law scaling relationship[18] (a least-squares fitted exponent of approximately 0.7 ± 0.1 is shown as the red straight line.) This is not indicative of scale-free behaviour over the whole range, but does indicate a surprisingly strong variation between scales over part of the histogram range.

The departure from the straight line or power-law region suggests that there might be different clusters present in the data that could be revealed by thresholding the data shown as a super-clump in Figure 7. We therefore apply a population threshold to the voxel values and identify and count the component clusters that arise when we only include voxels with the channel combination population of values above the threshold.

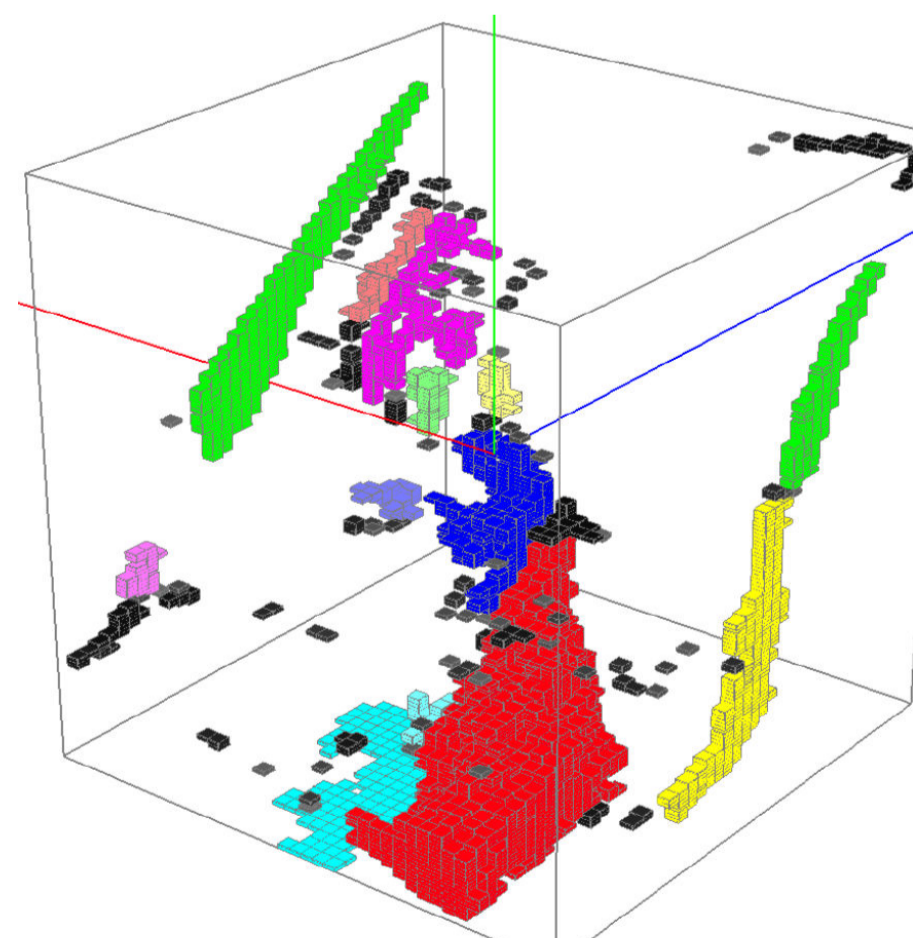


Figure 7: Frequency Threshold-clustered voxel hyper-brick showing separable clusters in channel space. Colours are assigned to distinguish separate components. The Red-Green-Blue axes shown correspond to increasing directions for Vis, IR1, IR3.

Figure 7 shows how the super-clump separates into component clusters when we apply a threshold of value 150. This value is obviously entirely dependent on the GMS5 characteristics and is obtained empirically. We can further analyse this idea by examining the number of clusters and the size of the biggest cluster that occurs when we vary the threshold.

Discussion

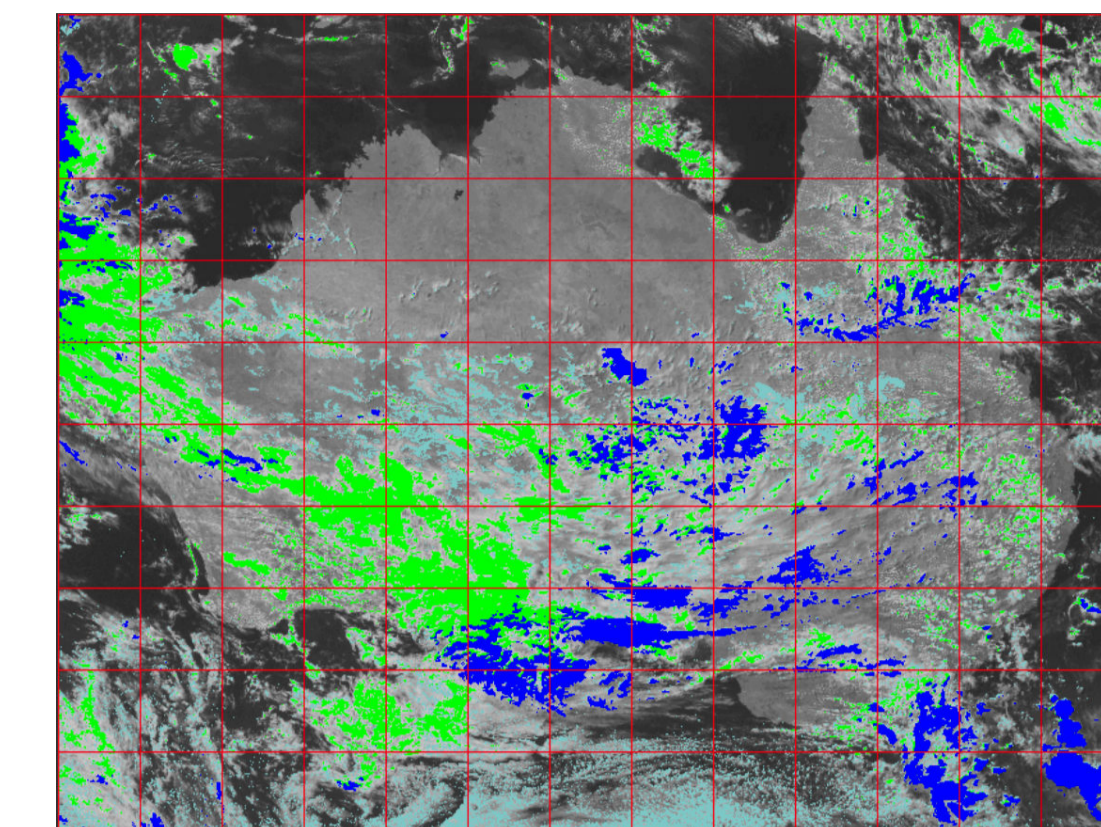


Figure 8: Artificially-coloured region around Australia. Colours correspond to different cloud types, identified by component clustering in VIS-IR1-IR3 voxel-space

This analysis technique - which can be done on a modern workstation in near interactive time, allows us to separate out the component clusters of multi-channel-space voxels from the super-clump. It allows us to identify features semi automatically and use this information to artificially colour the pixels in the original imagery according to their feature type. Figure 8 shows the visible, water-vapour and an artificially coloured image of the same spatial region over Australia. We have used the component identification technique to identify three different sorts of cloud cover, that would otherwise be hard to separate out quantitatively in the original raw imagery. Our technique generates actual voxel clusters of arbitrary complexity and shape and specified solely by the histogram thresholding parameter.

A custom computer program in C++ using OpenGL[19] was used to generate the 3D visualisation of the 3D voxel space. A fast cluster component labelling and identification program was also developed[20]. A separate image visualisation program was used to apply the artificial colouring to features in the original imagery. It would be possible to combine these two techniques into a single integrated program and also to extract statistics on the area coverage for particular features.

In summary, we have used a fast component cluster identification algorithm to identify separate clusters in the 3D channel voxel space. The component labelling technique can be used on the voxel data so-generated but it reveals a connected super-clump of voxels that is not otherwise easy to separate visually. The thresholding appears to give good visual cues to a human operator trying to identify clusters in the data. This technique works particularly well if done semi-interactively. The operator tries a range of threshold values - chosen near the knee of the cluster count curve and can see where the dominant clusters appear and persist. This information, combined with a suitable knowledge of meteorological parameters of different cloud types - or other specialist knowledge of other feature parameters such as vegetative land cover properties[21], would assist a data analyst set up an automatic feature detection filter.

We believe this technique has some promise in identifying features in present generation multi-spectral satellite data but - with judicious choice of which dominant three channels to focus upon - also for new generation hyper-spectral data sources. An area of possible further work would make use of multiple temporal images and super resolution techniques combined with multi-spectral analysis[22, 23]. Finally, we have only used unregistered pixel imagery. Greater value can be extracted from the time-sequence of images if they can be registered to coincide spatially.

References

- [1] K. Hawick, "Clusters in hyper-cubic multi-channel satellite imagery," Tech. Rep. CSTN-071, Massey University, September 2010.
- [2] T. Lillesand and R. Kiefer, *Remote sensing and image interpretation*. John Wiley and Sons, Inc., 3rd ed., 1994. ISBN 0-471-30757-6.
- [3] K. A. Hawick, R. S. Bell, A. Dickinson, P. D. Surry, and B. J. N. Wylie, "Parallelisation of the unified model data assimilation scheme," in *Proc. Fifth ECMWF Workshop on Use of Parallel Processors in Meteorology*, (Reading), European Centre for Medium Range Weather Forecasting (ECMWF), November 1992.
- [4] Q. Xiao, X. Zou, M. Pineda, M. Shapiro, and C. Velden, "Impact of gms-5 and goes-9 satellite-derived winds on the prediction of a norpex extratropical cyclone," *Monthly Weather Review*, vol. 130, pp. 507-526, March 2002.
- [5] M. Nordén, D. Doelling, M. Khayyer, A. Rapp, P. Minnis, and L. Nguyen, "Gms-5 satellite-derived cloud properties over the tropical western pacific," in *Proc. Eleventh ARM Science Team Meeting*, (Atlanta, Georgia), pp. 1-8, March 2001.
- [6] Japanese Meteorological Satellite Center, "The GMS user's guide," 3-25 Nakajiyoto, Kiyose, Tokyo 204, Japan, 1989.
- [7] K. Hawick, H. James, K. J. Maciunas, F. Vaughan, A. Wendelborn, M. Buchhorn, M. Rezny, S. Taylor, and M. Wilson, "Geostationary-satellite imagery applications on distributed, high-performance computing," in *Proc. of High Performance Computing (HPC) Asia '97*, pp. 50-55, IEEE, 1997.
- [8] M. A. Lombardi and D. W. Hanson, "The goes time code service, 1974/2004: A retrospective," *J. Res. Nat. Inst. Standards and Tech.*, vol. 110, pp. 79-96, March-April 2005.
- [9] H. James and K. Hawick, "Eric: A user and applications interface to a distributed satellite data repository," Tech. Rep. DHPC-008, Computer Science, The University of Adelaide, South Australia, April 1997.
- [10] K. Hawick and H. James, "Distributed high-performance computation for remote sensing," in *Proc. Supercomputing '97*, no. ISBN 0-89791-905-0, (San Jose, California, USA), pp. 1-13, ACM/IEEE, November 1997.
- [11] National Center for Supercomputing Applications, "Getting started with HDF - user manual," University of Illinois at Urbana-Champaign, May 1993.
- [12] K. Hawick and P. Coddington, "Interfacing to distributed active data archives," *Journal of Future Generation Computer Systems*, vol. 16, pp. 73-89, 1999.
- [13] K. A. Hawick, P. D. Coddington, and H. A. James, "Distributed frameworks and parallel algorithms for processing large-scale geographic data," *Parallel Computing*, vol. 10, p. 1297, 2003.
- [14] K. Kerry and K. Hawick, "Kriging interpolation on high-performance computers," in *Proc. High-Performance Computing and Networking (HPC/N'98)*, vol. 1401/1998 of LNCS, (Amsterdam), pp. 429-436, Springer, 1998.
- [15] O. Bosman, P. Fletcher, and K. Tsui, "K-tiling: A structure to support regular ordering and mapping of image data," in *APRS Workshop on Two and Three Dimensional Spatial Data: Representation and Standards*, (Perth, Western Australia), 7-8 December 1992.
- [16] P. Flanders and S. Reddaway, "Parallel Data Transforms," DAP Series, Active Memory Technology, 1988.
- [17] K. A. Hawick and D. P. Playne, "Hypercube Storage Layout and Transforms in Arbitrary Dimensions using GPUs and CUDA," Tech. Rep. CSTN-096, Computer Science, Massey University, 2010. Accepted for and to appear in *Concurrency and Computation: Practice and Experience*.
- [18] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509-512, October 1999.
- [19] K. Hawick, "3-d projection and simulation model visualisation," Tech. Rep. CSTN-082, Computer Science, Massey University, April 2009.
- [20] K. A. Hawick, A. Leist, and D. P. Playne, "Parallel Graph Component Labelling with GPUs and CUDA," *Parallel Computing*, vol. 36, pp. 655-678, 2010. CSTN-089.
- [21] R. Chinchuluan, W. S. Lee, J. Bhanania, and P. M. Pardalos, *Advances in Modeling Agriculture Systems*, ch. Clustering and Classification Algorithms in Food and Agricultural Applications: A Survey, pp. 1-22. Springer, 2009.
- [22] A. Gilman, D. G. Bailey, and S. R. Marsland, "Interpolation models for image super-resolution," in *4th IEEE Int. Symp. on Electronic Design, Test and Applications*, no. 0-7695-3110-5/08, (Hong Kong), pp. 55-60, IEEE Computer Society, January 2008.
- [23] F. Li, X. Jia, and D. Fraser, "Superresolution reconstruction of multispectral data for improved image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, pp. 689-693, 2009.