Image processing for smart browsing of ocean colour data products and subsequent incorporation into a multi-modal sensing framework

E. O'Connor[1], J. Hayes[2], C. O'Conaire[1], A.F. Smeaton[1], N.E. O'Connor,[1]D. Diamond[1]

[1]CLARITY: Centre for Sensor Web Technologies, Dublin City University, Dublin 9
[2]IBM Innovative Environmental Solutions, Mulhuddart, Ireland

Abstract

Ocean colour is defined as the water hue due to the presence of tiny plants containing the pigment chlorophyll, sediments and coloured dissolved organic material and so water colour can provide valuable information on coastal ecosystems. The 'Ocean Colour project' collects data from various satellites (e.g. MERIS, MODIS) and makes this data available online. One method of searching the Ocean Colour project data is to visually browse level 1 and level 2 data. Users can search via location (regions), time and data type. They are presented with images which cover chlorophyll, quasi-true colour and sea surface temperature (11 μ) and links to the source data. However it is often preferable for users to search such a complex and large dataset by event and analyse the distribution of colour in an image before examination of the source data. This will allow users to browse and search ocean colour data more efficiently and to include this information more seamlessly into a framework that incorporates sensor information from a variety of modalities. This paper presents a system for more efficient management and analysis of ocean colour data and suggests how this information can be incorporated into a multi-modal sensing framework for a smarter, more adaptive environmental sensor network.

1 Overview

Remote sensing from satellite and airborne sensors has proved to be a tremendous tool for studying our environment at large temporal and spatial scales. It offers unique large scale synoptic data to capture the range and variability of many complex processes. The use of satellite ocean colour sensing has a range of potential applications. One of its primary uses is monitoring the spatial extent and duration of harmful algal blooms (HABs) and the environmental conditions surrounding them. Some phytoplankton blooms can be toxic which may have many negative impacts including fish, bird, and marine mammal deaths. It has also been known to irritate human eyes and respiratory systems. Research has been carried out into satellite image analysis techniques to distinguish between toxic species and other species (e.g. Amin et al. 2009; Astoreca et al, 2009; Kahru and Mitchell, 1998; Millie at al., 1997; Staehr and Cullen, 2003). Algal blooms are a natural phenomenon; however the frequency, duration and distribution appear to have increased in recent years (Mueller-Karger et al, 2007). Many studies are investigating the use of satellite imagery and in-situ field measurements to predict the occurrence of algal blooms in order to mitigate their effects in a timely fashion (e.g. Heisler et al, 2008; Pitcher et al., 2010; Pitcher and Weeks, 2007; Zingone and Enevoldsen, 2000).

Despite the huge benefits that satellite remote sensing has brought to marine monitoring applications, there are also many challenges with using this type of data. Generally ocean colour satellite sensors only operate in the solar reflective spectral range; hence they only gather useful data on cloud-free days during periods of daylight (i.e. when illumination conditions are suitable). In regions where cloud cover is a predominant weather condition such as in coastal regions around Ireland and the UK, this can prove hugely problematic for obtaining satellite ocean colour information at high temporal scales. There are also issues in the coastal zone for the reliable estimation of parameters especially that of chlorophyll concentration. Thus a marine monitoring/event detection network that incorporates a number of sensing modalities is desirable. Work carried out in (O'Connor et al, 2009) demonstrates the necessity for both satellite and in-situ information sources in an environmental sensor network monitoring SST related events.

The Ocean Colour project (Ocean Colour Web, 2010) collects data from various satellites (e.g. from MERIS aboard ESA's Envisat platform and from MODIS aboard NASA's Terra and Aqua satellites etc.) and makes this data available online. Users can visually browse data through searching via location (regions), time and data type. Users are presented with JPEG images which cover quasi-true colour, chlorophyll, and sea surface temperature (11 μ), and links to the source data (See Fig. 1). These images provide a summary of the source data. However often it would be preferable for a user to browse by event (e.g. find data where there are events in a certain region), and to analyse JPEG images and trends over a particular time period before deciding to carry out an in depth analysis of the raw data sources or to decide which data sources need to be examined. It is also often preferable for the user to be able to browse by data availability especially in regions where cloud cover is highly predominant. Also due to the limitations of satellite information such as limited temporal resolution and data reliability issues, a framework which incorporates satellite information alongside in-situ sensors and context information can provide greatly improved event detection and analysis. Both information sources are necessary for reliable event detection and sufficient monitoring of a coastal marine environment. The more efficient management of ocean colour data allows the implementation of algorithms which consult data from other modalities for similar time periods to provide context and validation of events and a smarter, adaptive sensor network.The following sections outline the development of such a system for Irish coastal waters. These techniques could easily be adapted for developing a similar system for other coastal regions.

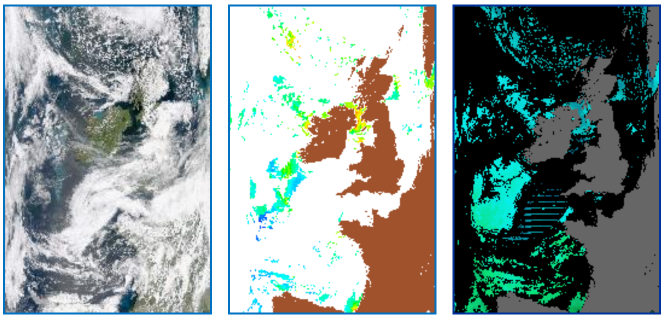

Figure 1: Summary images of raw data representing quasi-true colour, chlorophyll, and Sea Surface Temperature (11 μ)

2 Detection and Classification of Ireland in the Imagery

2.1 Data

For the initial development of the system 2251 images were downloaded from the Ocean Colour Web (Ocean Colour Web, 2010) . The 'English Channel' with swaths containing at least 75% was selected as the region of interest. Only MODIS images were considered for this investigation however the incorporation of data from other satellite sensors such as MERIS could also be seamlessly included if required. The downloaded images included quasi-true colour, chlorophyll and SST images. There was a fourth SST related image that was sometimes available to the website scraping application we used however it was not used in the context of this work. The chlorophyll images were used for initial region modelling. Out of the 2251 images that were downloaded, 711 were chlorophyll images ranging from May 31 2008 to October 28 2009. The focus of this system is satellite image browsing/analysis for Irish coastal waters. However the images for the English Channel include images which do not contain Ireland at all, images where only parts of Ireland can be seen and images where a full view of Ireland is contained in the image, however the satellite has viewed it from different positions and angles, therefore it changes shape, size and location in the image (See Fig. 2). Therefore the first step in developing this system is to develop a method to detect and classify Ireland in the incoming satellite image downloaded by the scraper application.

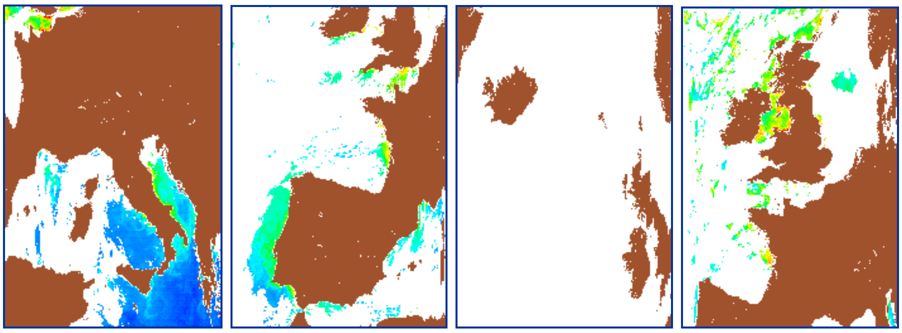

Figure 2: Example of the variety of images downloaded from the region denoted 'English Channel'. In some images Ireland does not appear at all, in other images it partially appears and in other images it appears with varying location, shape and size.

2.2 Training and Testing

Chlorophyll images were divided into a training set and a test set. The training set consisted of 500 images and the test set consisted of the remaining 211 images. All image analysis was carried out using the Matlab Image Processing Toolbox (V2009a). A variety of standard features (area, major axis length, minor axis length, eccentricity, convex area, filled area, Euler number, solidity, extent and perimeter) were extracted from the connected component representing Ireland in relevant images and stored. These stored features provide a suitable model to represent Ireland. In testing, each of the connected components in the image are tested against the model in order to detect if they represent Ireland. If the area of the component is less than ten pixels (this was the general area of various small islands scattered in the region) or if more than five pixels of the component is located on a border (these thresholds were chosen from analysing a variety of images and only images where all regions of Ireland are in the image are required), then the component is disregarded, otherwise it is tested against the model of Ireland developed from the training images. The likeliest component from the image is selected as the component that has features within the range of the max and min of the most features of the Ireland model. If this likeliest component is within the range of at least 8 of the features then it is classified as Ireland. The 211 test images were manually annotated as ‘contains Ireland’ or ‘no Ireland’. When verified against the ground-truth data, the algorithm produced an excellent classification rate of 99.52%. When examined, the single false positive image was actually due to a difficulty associated with the annotation whereby part of Ireland was contained on the border but it was less than five pixels and the annotator had deemed it to be greater than five pixels. Each component classified as Ireland was also saved as an image so it could be verified that the classified component was actually Ireland and not another component in the image.

3 Regionalisation of Ireland

In order to detect events in specific regions around the coast of Ireland, six main regions were delineated – North West, West, South West, North East, East, South East. This is quite a difficult task due to the fact that the shape, size and orientation of Ireland changes from image to image. An algorithm which uses the centroid, bounding box and orientation of Ireland in the image under consideration is used in order to regionalise Ireland. This algorithm works well in most cases with a possible improvement desirable in cases where Ireland is located in the far left of the image. In this algorithm, the orientation of the connected component representing Ireland is used in order to orient the image whereby Ireland is in an upright position. The centroid and bounding box are then used to calculate each region as a percentage increase from the borders of the bounding box with the centroid delineating a central point. Sample images that have been regionalised can be seen in Fig. 3.

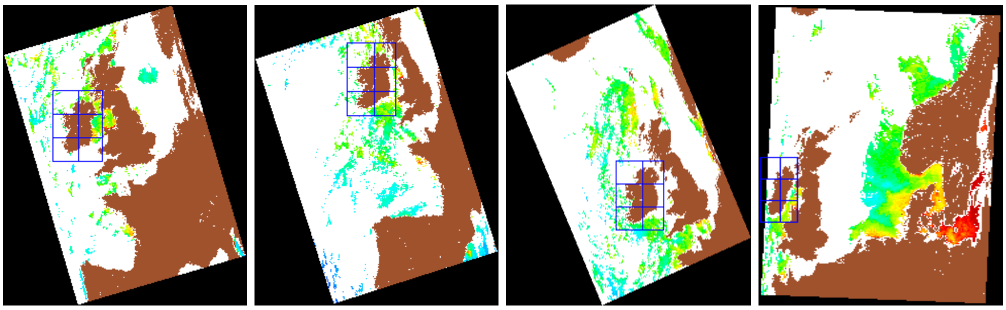

Figure 3: Sample images which have been regionalised. Overall the algorithm is very effective. However the last image is an example image where the algorithm is least effective and could possibly be improved.

4 Satellite Image Analysis and Browsing System

Having developed a system for detection, classification and regionalisation of Ireland, the next step involves providing a system which allows the quick and easy analysis of this data in order to determine the raw data sources that are desirable to explore further, and the incorporation of satellite information into a framework which includes information from multiple modalities.

4.1 *Regional Histograms*

An analysis of the RGB values used in the chlorophyll images was carried out on the 500 training images. In total 63 RGB values were used to represent the various conditions in the underlying raw data sources. Brown represents land, white represents no data available, and other colours then range from purple to red, with purple representing the lowest amount of chlorophyll and red representing the highest.

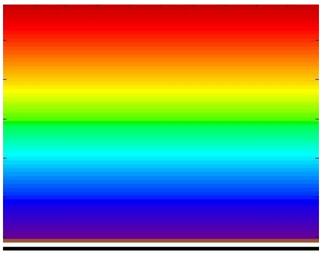

Figure 4: The range of colours used in the Chlorophyll images. Brown represents land, white represents no data available, and the colours then range from purple to red, with purple representing the lowest amount of chlorophyll and red representing the highest.

Histograms are developed for each region with 63 bands representing each of the colours of the chlorophyll images. Histograms with a lower resolution composed simply of the five main colours representing chlorophyll concentration (i.e. blue, green, yellow, orange, red) are also developed in order to provide the user with a high level overview of events in the region. An analysis of colour over time was also carried out through extracting high level colour information for each region in each image analysed and storing it. This allows the quick analysis of colour over time, in order to identify event 'hotspots', and subsequently when the raw data values need to be monitored very closely in the future. Potential trends and patterns may also be identified, and further sensing modalities may be consulted for event validation or to investigate the correlation between modalities or other environmental variables.

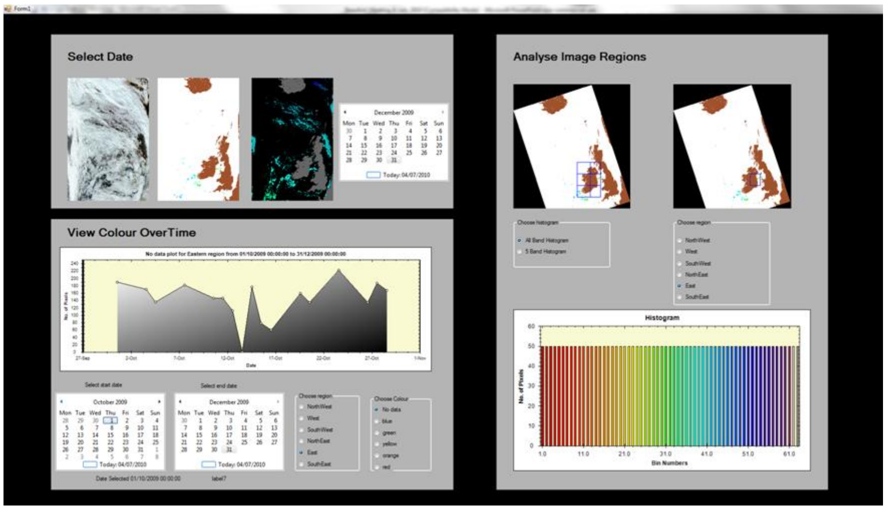

Figure 5: Image Analysis System for examining individual images and viewing colour trends over time

4.1 Image Analysis System

Figure 5 above shows the image analysis system for analysing the summary images for the chlorophyll raw pixel data. If for the date selected, a satellite image of Ireland is available, the quasi-true colour image, the chlorophyll image and the sea surface temperature image are displayed in the top left panel. The top right panel allows the user to select the region they want to analyse and the histogram they want to view for this image (i.e. All Band, 5-Band). The bars of the histogram are coloured according to the colour value that particular bin represents. Finally the bottom left-hand pane allows the user to select a start date, an end date, a region and a colour and to examine its distribution over time. This can be extended to allow the distribution of data from other modalities to be examined, and subsequently to examine peaks in data and the corresponding relationships between modalities. It also allows for the extension into an event detection framework where signals from various modalities are compared, source data is incorporated where relevant and events are disputed or validated. "Event hotspots" can be identified and the conditions surrounding them which can help predict these events in the future and mitigate the effects in a timely fashion.

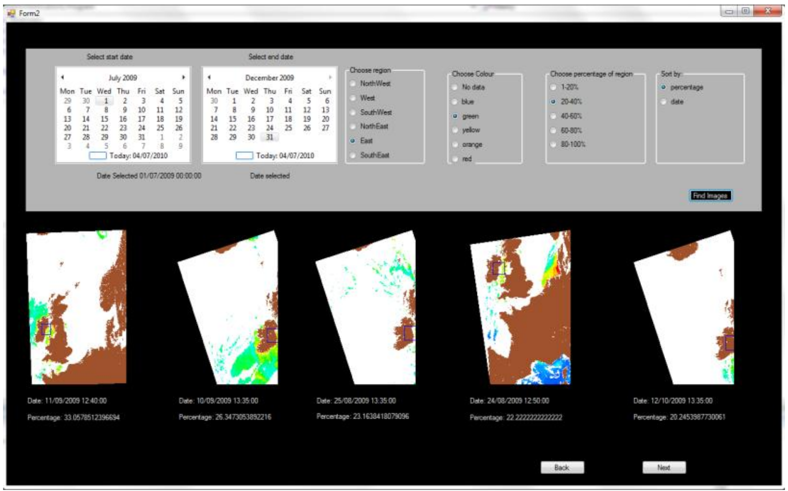

Figure 6: Search System I- View images for a selected time period that match certain criteria.

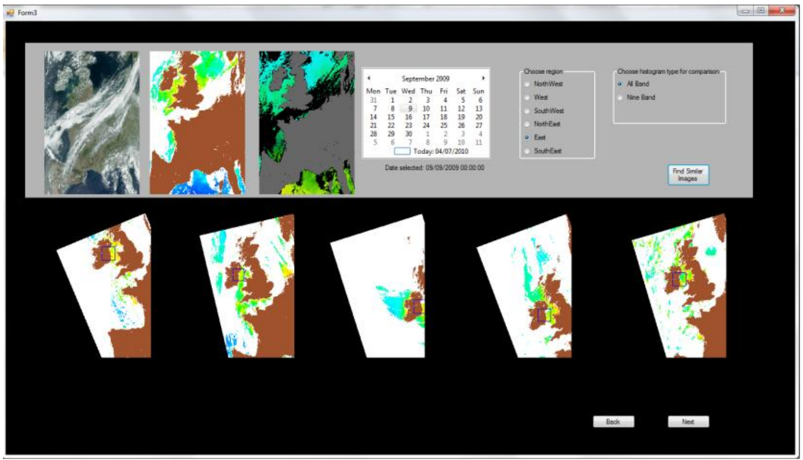

Figure 7: Search System II- View images with similar events to the selected image region.

4.2 Search Systems

As previously outlined often it is preferable for a user to browse by event (e.g. find data where there are events in a certain region), and to search for images which meet varying requirements. For example the user may want to search for images where there has been a high chlorophyll event (e.g. red pixels) or where at least 75% of the region is not covered by cloud. Figure 6 shows a search system which was developed which allows the user to browse data by choosing a time period, a region, a colour, and a percentage of the region. This allows the user to search for images which are not contaminated by cloud cover and which contain a particular colour representing a certain percentage of the water pixels in that region. The user can choose to sort the results by percentage or by date. In the second version of the search system (Fig. 7), the user browses the image data and if they see an event of interest in a particular region, the system will retrieve other images with similar events in this region. Similar regions are retrieved through calculating the Euclidean distance between the histograms (the user can choose this comparison to be a low-level or a high-level comparison through selecting All-Band or 5-Band). This is a standard technique in the image processing literature, however the system will be extended to include other histogram comparison techniques and investigate those which are more suitable.

These search systems allows quick and easy access to data which meets certain criteria for further analysis. It can also easily be extended into a framework whereby images which meet certain criteria can be compared against other modalities. If it seems there is the possibility of the onset of an event, source data can subsequently be requested and a more in depth study can be carried out, resulting in early warning of marine events. It also allows for the quick analysis of trends in a region over time. Other modalities can be investigated for dates or periods where specific types of events have been detected in a certain region and an in depth analysis can be carried out.

5 Conclusion

This paper has demonstrated a satellite image analysis and search system for quick and easy browsing of satellite ocean colour imagery. This allows the more effective browsing and analysis of ocean colour data and the seamless inclusion of satellite imagery into a multi-modal event detection framework. This system can easily be extended to incorporate alternative sensing modalities which are necessary for reliable and effective monitoring of coastal locations. This system has been based around Irish coastal waters. However similar techniques can easily be applied to adapt this system for use in other regions.

Acknowledgements

Based on research funded by the Dept. of Communications, Marine & Natural Resources under the Strategy for Science, Technology and Innovation (2006-2013) and by Science Foundation Ireland under grant 07/CE/I1147.

References

AMIN, R., ZHOU, J., GILERSON, A., GROSS, B., MOSHARY, F., and AHMED, S., 2009, Novel optical techniques for detecting and classifying toxic dinoflagellate Karenia brevis blooms using satellite imagery, *Opt. Express*, **17**, 9126-9144.

HEISLER, J., GILBERT, P.M., and BURKHOLDER, J.M., et al., 2008. Eutrophication and harmful algal blooms: A scientific consensus. *Harmful Algae*, **8**, 1, pp. 3-13.

KAHRU, M. AND MITCHELL, B. G. (1998), Spectral Reflectance and absorption of a massive red tide off Southern California. *Journal of Geophysical Research*, 103, 601-621.

MILLIE, D. F., SCHOFIELD, O. M., KIRKPATRICK, G. J. ET AL., 1997. Detection of harmful algal blooms using photopigments and absorption signatures: a case study of the Florida red tide dinoflagellate Gymnodinium breve. *Limnology and Oceanography*, **42**, 1240-1251.

MULLER-KARGER, F.E., ANDRÉFOUËT, S., VARELA, R., and THUNELL, R., 2007. The colour of the coastal ocean and applications in the solution of research and management problems, *Remote Sensing of Coastal Environments*, chapter 5, pp. 101-127, Springer, 2007.

OCEAN COLOUR WEB, http://oceancolor.gsfc.nasa.gov/ [Last accessed July 7 2010]

O'CONNOR, E., HAYES, J., SMEATON, A.F., O'CONNOR, N.E., and DIAMOND, D., 2009. Environmental monitoring of Galway Bay: Fusing data from remote and in-situ sources. In *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology IX*, SPIE Europe's International Symposium on Remote Sensing, Berlin, Germany, 31 August - 3 September 2009.

PITCHER, G.C., FIGUEIRAS, F.G., HICKEY, B.M. and MOITA, M.T., 2010. The physical oceanography of upwelling systems and the development of harmful algal blooms, *Progress in Oceanography*, **65**, 1:2, pp. 5-32.

PITCHER, G.C. and WEEKS, S.J., 2007**.** The variability and potential for prediction of harmful algal blooms in the southern Benguela ecosystem, *Large Marine Ecosystems*, **14**, pp. 125-146

STAEHR, P.A. and Cullen, J.J. (2003). Detection of Karenia mikimotoi by spectral absorption signatures, *Journal of Plankton Research*, **25**, pp. 1237-1249.

ZINGONE, A., and ENEVOLDSEN, H.,O., 2000. The diversity of harmful algal blooms: a challenge for science and management. *Ocean and Coastal Management*, **43,** 8:9, pp. 725-748