

相似海水鹽度資料查詢系統

A Query System for Similar Sea Salinity Data

謝尚琳

大同大學資訊工程研究所
slhsieh@ttu.edu.tw

邱峰村

大同大學資訊工程研究所
morris0001@gmail.com

蘇明雄

大同大學資訊工程研究所
milo1003@gmail.com

Abstract

The sea salinity is closely related to the sea temperature, which consequently causes the variations of the weather. This study presents a query system specifically designed for searching similar sea salinity data. The presented query system has been developed to provide information for experts to study the changes in the sea salinity so they can further explore the weather anomalies. The query system, when given the current sea salinity data, is able to discover the most related sea salinity data in the past years and then display the results on the screen. The user enters the query data through the user interface provided by the system, which next divides the query data into several blocks and then extracts the features from the blocks. After that, the system calculates the differences between the extracted features and the recorded features in the feature database, and then sorts the feature records accordingly to find the most similar results. The experimental results showed the system not only can obtain high search effectiveness, but also can save plenty of search time when compared to some other method.

Keywords: sea salinity, similar data search, Argo

摘要

因為海水鹽度與海水溫度息息相關，而海水溫度又會造成氣候的變化，本研究提出了一套為了找尋相似海水鹽度資料而特別設計的查詢系統，此系統係被開發來提供資訊讓專業人員研究海水鹽度的變化，以便進一步探討全球氣候的異常現象。本文提出的查詢系統，可根據目前的海水鹽度的資料去找出最相關的歷年海水鹽度資料，並於畫面上顯示其結果。當使用者透過本系統所提供的使用者介面輸入欲查詢的資料後，本系統會將查詢資料進行區塊切割並從中擷取出特徵值，再計算與特徵資料庫內的特徵記錄之間的變化值，並據此將特徵記錄排序，以找尋最相似的結果。從實驗的結果可以看出，本系統不僅可以獲得相當高的有效搜尋率，同時相較於他法亦可節省相當多的時間。

關鍵字：海水鹽度、相似資料查詢、Argo

1. 前言

由於人類過度開發，破壞了大自然的環境，造成全球暖化、北極的冰層逐漸融化及海平面漸漸升高。根據聯合國 2007 年公佈的評估報告顯示，年平均溫度正逐年上升，造成全球氣候的異常變化。

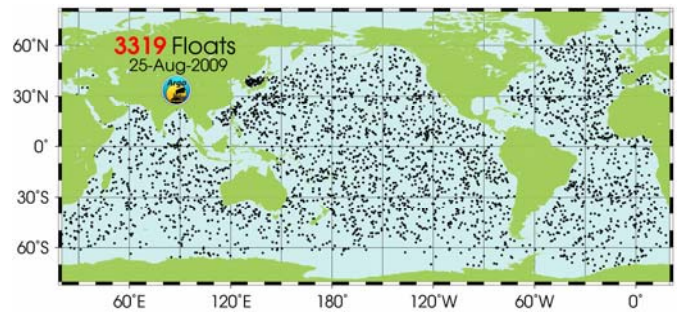
在 1998 年美國、法國等國的氣象、海洋學者，推出一個 Array for Real-time Geostrophic Oceanography(簡稱 Argo)計畫，其目的在快速準確的搜集海洋表面和剖面的鹽度與溫度資料，希望利用這些資料，有效防止全球氣候性災害的發生及提高氣象預測的準確度。因為海水溫度與鹽度的變化息息相關[1]，而海水溫度又會造成氣候的變化，因此本研究特別發展一套相似海水鹽度資料的查詢系統，供專業人員研究海水鹽度的變化，進而探索目前全球暖化的程度，並預防異常氣候帶來的重大災害。

在已經發表的影像檢索系統文章[2-4][9-10]中，大部份都是針對一般的影像檢索，尚未看到針對海水鹽度特性而開發的資料查詢系統。這些檢索系統雖然可以套用在海水鹽度影像的檢索上，但由於未針對海水鹽度而設計，其檢索出來的資料未必準確。因此本文提出一套專門根據海水鹽度資料而設計的查詢系統，可利用海水鹽度的關係去找出最相關的歷年海水鹽度資料。

2. 相關背景知識

2.1 Argo 計畫

Argo 是英文 Array for Real-time Geostrophic Oceanography 的縮寫，它是在 1998 年由美國邀集日本、法國、德國等主要國家的氣象與海洋學者所發展出一個全球海洋氣象觀測計畫，藉由這個計畫建立一套即時的海洋資訊監測系統，並規劃在全球放置三千多個監測浮標點(如圖一)，用來收集全球海洋鹽度與溫度表面和剖面資料。這些資料可以有效地幫助即時了解海洋變化並提高氣象預報準確度，更進一步有效地防止嚴重的氣候異常災害。目前許多國際組織的觀測研究計畫與發表的學術論文[1][5-8][11-12]都使用 Argo 提供的資料。



圖一. Argo 在全球海洋放置的浮標

資料來源: Argo - part of the integrated global observation strategy[13]

2.2 鹽度刻度範圍

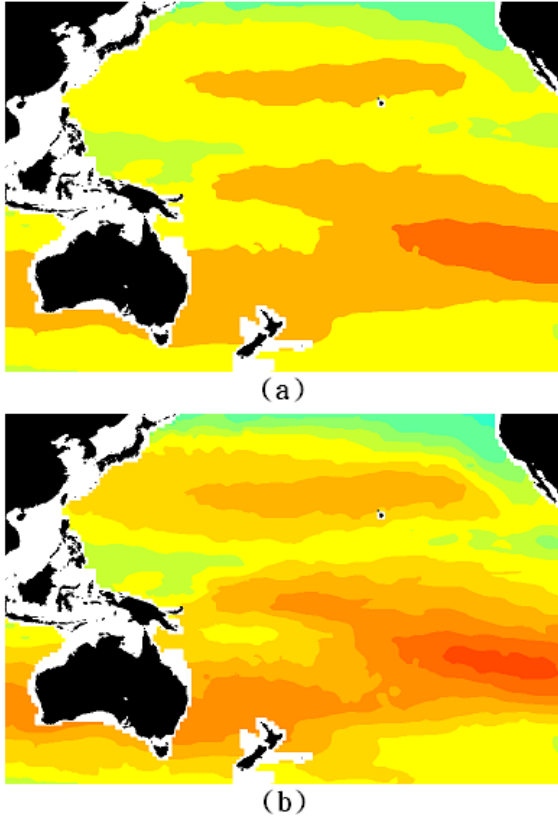
鹽在海水中的含量稱之為鹽度，目前最常用的測量方法是導電法，導電法係由海水的導電度所推算出來的鹽度數值，以 psu (practical salinity unit) 為單位。

圖二(a)(b)為相同年月但是不同 psu 的海水鹽度資料，由圖中可以看出以 1psu 為單位的(a)影像品質明顯低於以 0.5psu 為單位的(b)影像品質。因此如果將鹽度的刻度定義的太大，會因為量化太多而影響資料查詢出來的品質及準確性，因此本研究將鹽度刻度範圍定義在 28psu 到 37.5psu 間，以 0.5psu 為單位共 20 階。另外陸地與未測量鹽度為 0psu，所以整張的海水鹽度資料共有 21 階。

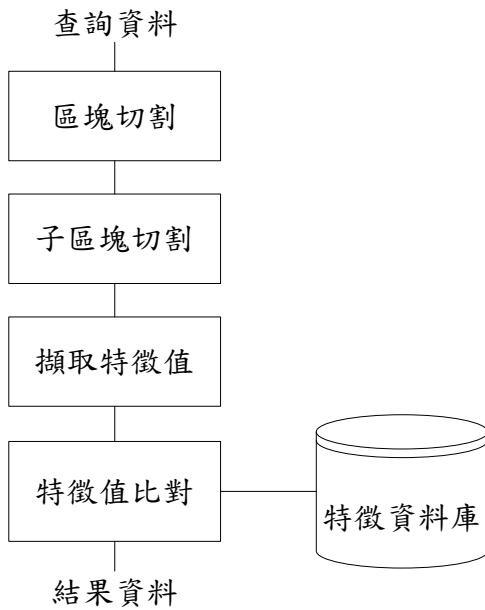
3. 相似海水鹽度資料查詢系統

本節將介紹相似海水鹽度資料查詢系統，以及所使用的資料特徵值擷取方法，並且提出一個特徵值的變化計算方式來達到查詢效果。

圖三是整個相似海水鹽度資料查詢系統的架構圖，當使用者選擇查詢資料後，由查詢資料擷取出特徵值，再與特徵資料庫內的特徵記錄計算變化值，得到最相似的結果。



圖二. 不同量化階的海水鹽度資料(a)單位為 1psu (b)單位為 0.5psu



圖三. 查詢系統架構圖

3.1 特徵值擷取方法

首先將資料如圖四切割成 $M \times N$ 個不重複的區塊 $B_l(l=1, \dots, M \times N)$ ，每個區塊大小為 $A \times A$ 。接著如圖五(a)和(b)將每個區塊分別垂直切割成 A/K 個 $A \times K$ 的子區塊 $X_i(i=1, \dots, A/K)$ 以及水平切割成 A/K 個 $K \times A$ 的子區塊 $Y_i(i=1, \dots, A/K)$ 。最後使用公式(1)與(2)分別計算鹽度特徵值 FX_i 與 $FY_i(i=1, \dots, A/K)$ 。

$$FX_i = \sqrt{\sum_{a=1}^A \sum_{k=1}^K S_{x,a,k}^i} \quad (1)$$

$$FY_i = \sqrt{\sum_{a=1}^A \sum_{k=1}^K S_{y,a,k}^i} \quad (2)$$

其中， $S_{x,a,k}^i$ 與 $S_{y,a,k}^i$ 分別為 X_i 子區塊與 Y_i 子區塊中位於座標 (a, k) 的像素所代表的鹽度值。

3.2 特徵值比對方法

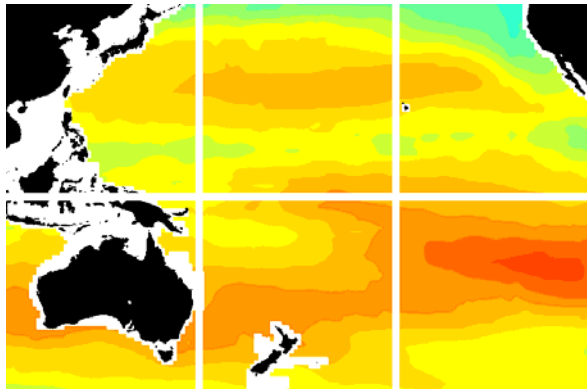
查詢的資料依前述特徵值擷取方法取得鹽度特徵值後，接著按照區塊的順序與區塊內 FX_i 與 FY_i 的順位，將這些特徵值與特徵資料庫內每筆被查詢的資料特徵值依公式(3)計算其變化值。

$$Diff_i = \sum_{i=1}^{A/K} (|FX_i - FX'_i| + |FY_i - FY'_i|) \quad (3)$$

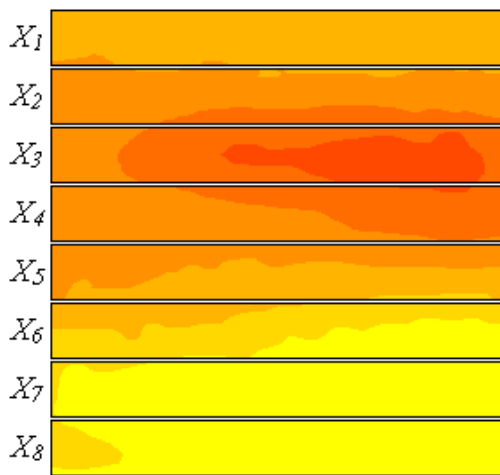
其中， $Diff_i$ 是區塊 $B_l(l=1, \dots, M \times N)$ 的變化值， FX_i 與 FY_i 是查詢資料的子區塊特徵值， FX'_i 與 FY'_i 是被查詢資料的子區塊特徵值。依公式(4)將 $Diff_i$ 加總後除以 $M \times N$ ，即可得到總平均變化值 T 。

$$T = \frac{1}{M \times N} \sum_{l=1}^{M \times N} Diff_l \quad (4)$$

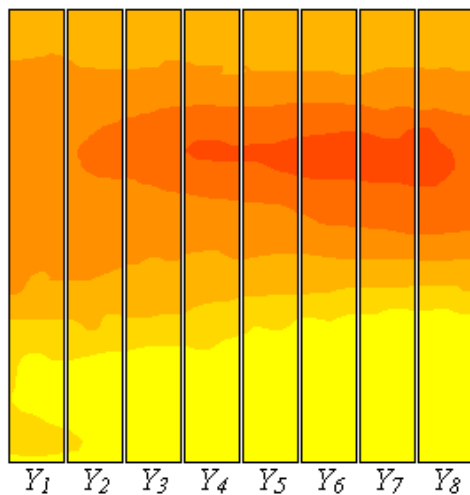
全部特徵資料庫內的紀錄都與查詢資料計算後，所得到 T 值最低的即為與查詢資料最相似的被查詢資料。



圖四. 資料切割區塊切割
M=2, N=3



(a)



(b)

圖五. 區塊個別切割成子區塊(A/K=8)(a)垂直切割子區塊(b)水平切割子區塊

4. 實驗結果

本實驗使用 2004 年 1 月至 2008 年 9 月的海水鹽度資料進行實驗，其資料係使用 Pacific Marine Atlas[14]網站提供的工具所產生。該網站的資料係以影像表示，每一種像素顏色分別對應一種海水鹽度值。實驗資料的範圍涵蓋經度 110°E 至 110°W 及緯度 45°N 至 50°S，海水鹽度值則介於 28psu 到 37.5psu 間，資料影像共 57 張，大小皆為 768 × 512 像素。

本實驗將每張影像分別當成查詢的輸入資料，並與其他 56 張不同年份的資料進行比對，找出最相似的資料影像。

4.1 實驗結果的評估方式

為評估本論文提出方法之有效性，本實驗另外設計一比對方法(稱為 MSE 法)，並將結果與其他比較。MSE 法計算每張影像的像素鹽度值與其他比對影像的像素鹽度值兩者間之 MSE(mean square error)值，MSE 值最小者即為最相似的資料。MSE 的計算公式如下：

$$MSE = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H (X_{i,j} - X'_{i,j})^2 \quad (4)$$

其中， W 和 H 分別為資料寬和高， $X_{i,j}$ 和 $X'_{i,j}$ 分別為查詢資料與結果資料的鹽度值。

MSE 法最大的缺點是計算量龐大，因此需耗費許多時間才能得到計算結果，所以實際上並不適合當成查詢的比對方式。

本實驗假設 MSE 法的排名結果是正確的，實驗結果的評估係將本文方法的查詢結果與 MSE 法得到的排名加以比較，若查詢出來的該筆資料位於 MSE 法排名前三名中，即視為有效的查詢。將有效的查詢筆數除以全部影像的張數(57 張)即為本實驗的有效率，如公式(5)。

$$\text{有效率} = \frac{\text{有效筆數}}{\text{全部影像張數}} \quad (5)$$

4.2 評估結果

本實驗依圖五的方式進行切割比對，表一為其實驗結果以及相對於 MSE 法排名之比率，從表中可以看出其有效率高達 98.2%。

表二為 2005 年各月份的資料查詢結果，由表中可以看出本方法的查詢結果一半以上與 MSE 法的第一名相同，其中唯一一筆查詢結果落在三名外的資料為 2005 年 7 月。圖六為選擇查詢資料為 2004 年 2 月，使用本查詢系統得到的查詢畫面，得到最相似資料為 2004 年 3 月；圖七為圖六中各結果之放大圖，其中(a)為查詢資料，(b)-(d)為 MSE 第 1-3 名，其年月份分別為 2004 年 3 月、2005 年 2 月與 2005 年 3 月。

4.3 查詢時間比較

前面提到 MSE 法需耗費許多時間才能得到計算結果，本節特別將本方法所需的時間與 MSE 法比較，以展現本方法之可用性。

如表三所示，將查詢影像與其他 56 張被查詢影像做 MSE 計算，共需花費 112 秒(每張需花費 2 秒)。將查詢影像使用本文提出的方法擷取特徵值後再與特徵資料庫比對所需時間為大約為 6.5 秒時間，其中特徵值擷取時間需要花費約 6.3 秒，而與特徵資料庫比對時間部分約花費 0.14 秒(每筆需花費約 0.0025 秒)。

由表三可以看出，使用本查詢系統，不僅有相當高的查詢有效率，且查詢的時間也比使用 MSE 方法快很多，且隨著資料庫資料筆數的增加，兩者的時間差距也會逐漸增大，茲分析如下：

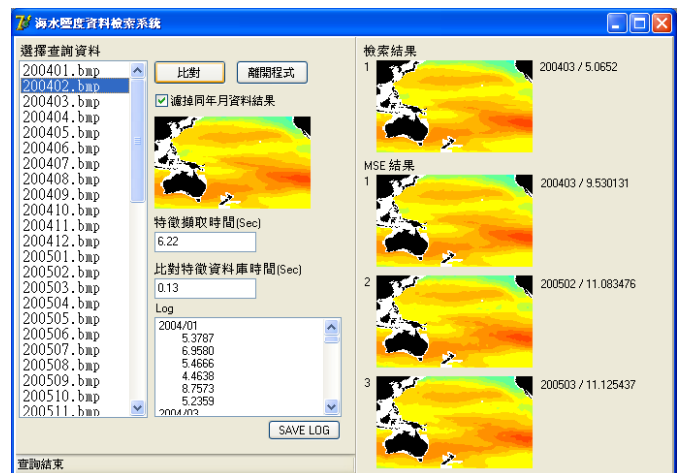
假設資料筆數是 n ，MSE 法花費的時間為 t_1 ，其每筆比對時間為 c_1 ，而本文方法花費的時間為 t_2 ，且特徵值擷取的時間為 c_2 ，每筆比對的時間為 c_3 ，則 t_1 與 t_2 可分別表示為

表一. 實驗結果之相對比例

MSE 法名次	1	2	3	> 3
相對於 MSE 法排名之筆數	30	17	9	1
所佔比例	52.6%	29.8%	15.8%	1.8%

表二. 2005 年資料查詢結果

查詢月份	檢索結果	MSE 分數最低三名			在 MSE 排名
		1	2	3	
2005.01	2005.02	2005.02	2004.12	2004.01	1
2005.02	2005.01	2005.03	2005.01	2004.02	2
2005.03	2004.02	2005.02	2005.04	2004.02	3
2005.04	2005.03	2005.05	2005.03	2004.04	2
2005.05	2005.04	2005.04	2004.05	2005.06	1
2005.06	2004.06	2005.05	2005.07	2004.06	3
2005.07	2006.07	2005.06	2005.08	2004.07	4
2005.08	2005.09	2005.09	2005.07	2004.08	1
2005.09	2005.08	2005.01	2005.08	2005.11	2
2005.01	2005.11	2005.11	2005.09	2004.01	1
2005.11	2005.01	2005.01	2005.12	2005.09	1
2005.12	2005.11	2005.11	2006.01	2005.01	1



圖六. 查詢與 2004 年 2 月相關資料之結果畫面

$$t_1 = c_1 n$$

$$t_2 = c_2 + c_3 n$$

故兩者之時間比為

$$\frac{t_1}{t_2} = \frac{c_1 n}{c_2 + c_3 n}$$

當資料筆數 n 大到一定程度後，特徵值擷取時間 c_2 相較於比對時間 c_3 ，將顯得微不足道而可忽略，所以可得到兩者之時間比為

$$\frac{t_1}{t_2} \cong \frac{c_1 n}{c_3 n} = \frac{c_1}{c_3}$$

根據表三之時間比較表， $c_1 = 112/56 = 2$ ， $c_3 = 0.14/56 \approx 0.0025$ ，故可得到

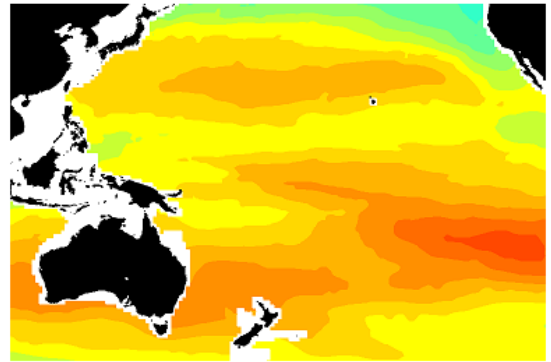
$$\frac{t_1}{t_2} \approx 800$$

由上述分析結果可得知，當資料筆數 n 大到一定程度後，本文方法較 MSE 法約可節省 800 倍的時間。

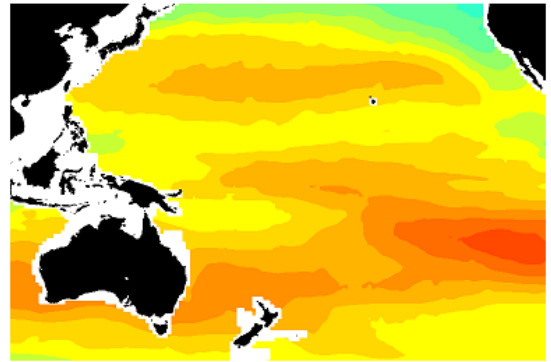
5. 結論與未來方向

本研究針對 Argo 的海水鹽度資料設計了一個相似海水鹽度資料查詢系統，讓相關研究人員得以找出最相關的歷年的 Argo 海水鹽度資料。從實驗的結果可以看出，本文方法不僅可以獲得相當高的有效查詢率，同時相較於 MSE 法亦可節省相當多的時間。

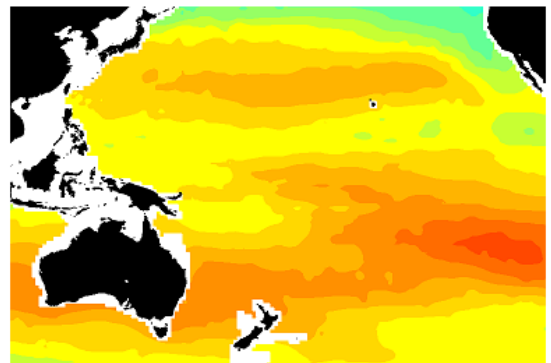
不過，本文的方法其花費時間仍會隨特徵資料庫筆數的增加而線性成長，雖然本文方法的時間常數(c_3)很小，成長的速度相對緩慢，但當特徵資料庫筆數眾多時，仍無法避免地會花費較多的時間。因此如何以更快速地得到查詢結果，而仍能維持原來的有效率，尚有很大的探討空間。



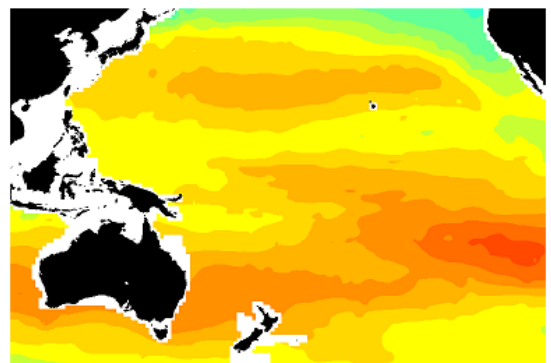
(a)



(b)



(c)



(d)

圖七. 查詢資料與 MSE 結果前三名 (a)查詢資料 (b)MSE 第 1 名 (c)MSE 第 2 名 (d)MSE 第 3 名

表三.資料筆數為 56 筆時的查詢時間比較表

單位:秒

查詢年月	MSE	相似海水鹽度資料查詢系統	
		特徵值擷取	資料比對
2004.01	112	6.28	0.13
2004.07	112	6.36	0.13
2005.12	112	6.39	0.14
2006.05	112	6.37	0.14
2007.02	112	6.36	0.13
2008.05	112	6.33	0.12
2008.07	112	6.31	0.14

參考文獻

- [1] 高立仁,「利用資料探勘及模糊推論技術預測海水溫度與鹽度變化之研究」,資訊工程研究所博士論文,大同大學,2008年。
- [2] 林立謙,「利用影像分類法之快速影像檢索系統」,電機工程學系碩士班碩士論文,逢甲大學,2002年。
- [3] 董志偉,「利用階層式比對之影像內容檢索」,資訊經營研究所碩士論文,大同大學,2005年。
- [4] 黃文彬,「以顏色樣式與空間樣式為主的影像搜尋系統」,資訊管理系碩士論文,朝陽科技大學,2006年。
- [5] 許建平、劉增宏、孫朝輝、朱伯康,「全球Argo即時海洋觀測網全面建成」,海洋技術,2008年3月,第27卷,第1期。
- [6] 寧鵬飛、孫朝輝、劉增宏、滕駿華、許建平,「Argo網路資料庫視覺化平臺技術及其應用」,海洋技術,2007年12月,第26卷,第4期。
- [7] 朱伯康、許建平,「國際Argo計畫執行現狀剖析」,海洋技術,2008年12月,第27卷,第4期。

[8] 許雯婷,「使用視覺化策略以呈現海水溫度與鹽度變化資訊之研究」,資訊工程研究所碩士論文,大同大學,2008年。

[9] E. Petrakis, C. Faloutsos, and K.I. Lin, "ImageMap: An image indexing method based on spatial similarity," IEEE Transactions on Knowledge and Data Engineering, Vol. 14, No. 5, pp. 979-987, Sep.-Oct. 2002.

[10] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: image segmentation using expectation-maximization and its application to image querying," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 8, pp. 1026-1038, Aug. 2002.

[11] J.J. Park, K.A. Park, K. Kim, and Y.Y. Youn, "Statistical Analysis of Upper Ocean Temperature Response to Typhoons from ARGO floats and Satellite data," IEEE Transactions on Geoscience and Remote Sensing Symposium, Vol. 4, pp. 2564-2567, Jul. 2005.

[12] J. Boutin, and N. Martin, "ARGO upper salinity measurements: perspectives for L-band radiometers calibration and retrieved sea surface salinity validation," IEEE Transactions on Geoscience and Remote Sensing Letters, Vol. 3, pp. 202-206, Apr. 2006.

[13] Argo - part of the integrated global observation strategy. <http://www.argo.ucsd.edu/> (2009/08/26)

[14] Pacific Marine Atlas. http://sio-argo.ucsd.edu/Marine_Atlas.html (2009/05/15)