

다중 volume 자료 처리 기술을 기반으로 한 3차원 유사 일기도 검색 시스템

정창성*, 이성환, 김지윤, 황치정
*충남대학교 컴퓨터학과
e-mail : csjung@ipl.cnu.ac.kr

3-D similar weather retrieval system based on multiplex volume data processing technology

Chang-Sung Jung*, Sung-Hwan Lee, Ji-Yoon Kim, Chi-Jung Hwang
*Dept. of Computer Science, Chungnam National University

요 약

본 연구의 목적은 동아시아 지역의 NCEP 데이터를 대상으로 1000hPa, 850hPa, 700hPa, 500hPa에서의 기온, 고도와 300hPa에서의 풍향, 풍속과 지상의 기온, 기압 데이터를 이용하여 유사 일기도 검색 시 동일한 파라미터를 사용하여 유사성 정도를 검색하는 알고리즘을 개발하는데 있다. 또한 각 파라미터에 해당하는 격자 데이터로부터 특징값들을 추출하고 이를 이용하여 유사도 측정을 수행하여 유사일기도를 검색하는 알고리즘을 개발하였다.

1. 서론

최근 과학 기술과 컴퓨터 통신망의 발달로 인해 넓은 지역의 지상과 상층에서 관측된 기상자료를 분석 처리하여 기상 변화를 예측, 정확한 날씨를 판단함과 동시에 집중 호우, 태풍 경보 등 자연 재해에 대비할 수 있는 유용한 정보를 제공하는 기술이 활발히 연구 활용되고 있다.

일반적으로 기상 변화에 영향을 주는 고기압, 저기압, 기단, 전선 등은 각각 특이한 일기 현상을 동반하며 규칙적인 변화와 이동을 하므로, 고/저기압 위치 및 수, 등압선/등고선의 간격 및 기압값에 의해 일기 유형을 분류할 수 있으며 고저기압의 특징적 위치에 의해서 유사패턴을 분류할 수 있고 이를 주관적 예보 기술로 활용하여 온 것은 잘 알려진 사실이다.

상층 일기도 유형은 지상 일기 유형에 비해 단순하여 쉽게 유형을 분류할 수 있으며 현재 강수량 및 하층 대기의 기온 변화 예상에 유용한 정보를 제공하는 1000hPa, 850hPa, 700hPa, 대규모적인 기류 분석에 필요한 500hPa, 제트 기류 분석에 필요한 300hPa 등 특정 부위에 대해 유형 분류기법이 사용 되고 있다.

지상 일기 유형 분류의 경우, 지형이나 밤낮의 변화에 따라 시간별로 복잡하게 변화하며, 작은

규모의 다양한 일기 현상이 발생/소멸하는 등 변화가 심해 상층 일기도 패턴에 비해 활용을 못하고 있는 실정이다. 그러나 날씨의 분포를 파악, 향후 변화를 예측하여 정확한 일기 예보를 하기 위해서는 상층 대기 상태 파악과 함께 지상 부근 대기 상태를 파악하는 것이 필수적이다. 그러므로 본 연구의 목적인 현재의 지상 상황 일기도와 패턴이 유사한 과거의 일기도를 검색 추출하는 알고리즘의 개발은 정확한 일기 예보를 하는데 있어 매우 중요하다.

2. 격자점 데이터를 이용한 특징 추출

본 연구에서 유사일기도 검색을 위하여 동아시아 지역에 해당하는 격자점에서의 데이터를 입력으로 사용하였다. 전체 데이터는 격자 45 x 25의 수치 데이터로 매 6시간 간격으로 측정되어져 있다. 이들 격자점의 데이터는 굴곡이 심하지 않도록 부드럽고, 중간에 단절되지 않은 연속성이 보장되며, 같은 값의 선이 마주보며 평행하지 않으며, 서로 교차하거나 한 등압선이 둘로 갈라지지 않는 성격을 가지고 있다. 또한 지역 전체를 볼 때 격자점의 데이터는 연속적인 성질이 있다. 각각의 데이터간의 유사도를 측정하기 위해서는 각 데이터의 특징을 추출하고 이를 수치화 하여 그 값을 비교하여야 한다. 이를 위해서 각

데이터의 특징이 되는 Peak/Valley 지점을 검출하고 그 지점과 한반도와와의 거리 및 주변의 등고선을 계산하여 데이터간의 유사도를 측정하는데 사용하였다.

2.1 Peak/Valley 지점 검출

기압데이터가 아닌 고도나 기온데이터 등은 지리적 위치에 영향을 받으므로 Peak/Valley 지점을 검출하기 위해서는 단순히 인근 주변데이터와 비교하는 것은 무의미하다. 그러므로 같은 위도상의 데이터들은 각 값에 동일 위도상의 값들의 평균값(Zonal mean)을 뺀 값을 이용하였다.

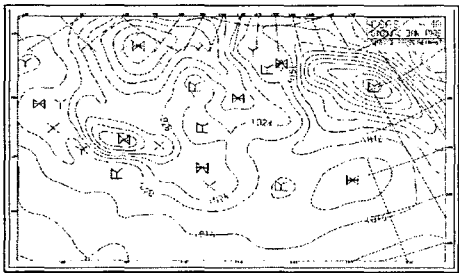
Zonal mean 값에 의해 다시 계산된 격자 데이터를 2 차원 배열로 간주하고 이에 대해 3X3 크기의 마스크 윈도우를 적용하여 Peak/Valley 의 위치를 찾는 방법을 구현하였다.

3x3 윈도우 중심의 데이터가 주변 8 개의 데이터에 비해 가장 작으면 Valley 로 가장 크면 Peak 로 판단하는 방법이다. 단, 주변의 여러 데이터들이 중심의 데이터와 같은 경우 이들이 일정 개수 이상이면 이를 Peak/Valley 로 판단하지 않는다.

X1	X2	X3
X4	X5	X6
X7	X8	X9

if(X5 == min(X1,...,X9)) then X5 is 'V'(Valley)
if(X5 == max(X1,...,X9)) then X5 is 'P'(Peak)

이 알고리즘의 정확도를 확인하기 위해 일기도 생성시 사용되는 알고리즘에 이 알고리즘을 추가하여 기존 방법으로부터 검출되는 Peak/Valley 는 'H'와 'L'로 표시하고 제안 알고리즘으로 검출된 Peak/Valley 는 'X'와 'Y'로 표시하여 일기도를 생성한 결과가 [그림 1]에 나타나 있다. 기존 알고리즘에서 추출했던 Peak/Valley 는 물론 추출하지 못한 추가적인 Peak/Valley 정보를 찾아낸 것을 확인할 수 있다.



[그림 1] Peak/Valley 의 위치 검출(3x3 마스크)

2.2 Peak/Valley 지점과 한반도 사이의 거리 측정

일기도 전체에 분포하는 Peak/Valley 지점이 서로 동일한 값의 중요도를 지니게 하는 것은 유사 일기도를 검출하는데 있어 불합리하다. 이는 우리나라에 전혀 영향을 미치지 못하는 Peak/Valley 지점들의 정보로 인해 유사도에 영향을 미쳐 원치 않는 검색 결과를 초래할 수 있기 때문이다. 그렇기 때문에 유사 일기도 검색을 위해서는 검출된 Peak/Valley 지점에 대해서 한반도를 중심으로 한 거리를 측정하여 가중치를 주는

방법이 필요하다. 한반도로부터 거리가 먼 곳에서 매칭이 일어난 Peak/Valley 지점의 정보는 한반도 인근지역에서 매칭이 일어난 것보다 가중치를 적게 주는 방법으로 같은 개수의 Peak/Valley 지점의 매칭이 일어났다 하더라도 한반도 주변에서 매칭이 일어난 일기도가 보다 높은 유사도를 갖게 하였다. 한반도에 영향을 주는 것은 한반도 주변의 Peak/Valley 지점들이기 때문에 이러한 한반도와의 거리에 대한 가중치를 부여하는 방법은 적절한 방법이라 할 수 있다.

각 Peak/Valley 의 한반도 중심으로부터의 거리에 대한 가중치는 [식 1]에 의해서 구해진다.

$$d = \sqrt{|C_x - x|^2 + |C_y - y|^2} \quad [식 1]$$

$$W = w_c * \frac{C}{d}$$

d: 거리, C_x, C_y: 한반도 중심의 좌표
x, y: 해당 Peak/Valley 의 좌표, w_c: 가중치 상수
C: 거리 상수, W: 거리에 대한 가중치

[식 1]에 의하면 거리가 멀어질수록 가중치가 작아져서 전체 유사도 측정에 미치는 영향이 작아지게 된다. 가중치/거리 상수를 조절해서 그 영향을 조절할 수 있다.

2.3 Peak/Valley 주변의 등고선 측정

유사 정도를 비교할 두 일기도의 Peak/Valley 의 분포의 유사 정도와 함께 Peak/Valley 중심점 주위의 등고선의 모양(간격)이 유사하면 더욱 높은 유사도를 갖는다. 이를 위해 검출된 Peak/Valley 에 대해 주위의 등고선 정보를 측정하는 알고리즘을 개발하였다.

Peak/Valley 의 등고선 정보는 격자 5x5 크기의 중심으로부터 8방향에 대해 거리가 1인 지점과 2인 지점의 중심 값과의 차이를 상수 c로 나눈 값을 이용한다. 중심 값과 차이가 많이 날수록 더 많은 수의 등고선이 존재하여 조밀한 분포를 나타내고, 차이가 적으면 적은 수의 등고선이 존재하므로 등고선 간격이 넓은 분포를 나타낸다.

X1	X2	X3
X11	X22	X33
X4	X0	X55
X66	X77	X88
X6	X7	X8

X0: Peak/Valley 로 검출된 위치
X1~X8: 8 방향의 거리가 2인 지점의 데이터
X11~X88: 8 방향의 거리가 1인 지점의 데이터

[그림 2] 등고선 정보 추출을 위한 5x5 마스크윈도우

3. 유사도 측정 알고리즘

일기도의 유사도 측정은 Peak/Valley 위치 및 한반도 중심으로부터의 거리 그리고 등고선 정보를 사용해서 이루어지고 있다. 유사도 측정을 통해 검색하고자 하는 원본 데이터와 나머지 데이터들과의 특징 값들을 비교하여 그 값이 가장 높은 순으로 정렬하여 원본과 가장 유사한 데이터를 찾아낸다.

3.1 Peak/Valley의 위치 정보를 이용한 유사도 측정

Peak/Valley의 위치정보를 이용한 가장 기초적인 유사도 측정 방법은 원본 데이터에서 추출된 Peak/Valley의 위치 정보와 나머지 데이터들의 위치정보를 서로 비교하여 같은 위치에서 같은 Peak/Valley가 존재하면 그 수를 누적하여 이 수가 가장 높은 순으로 정렬하는 것이다.

하지만 이는 정확한 위치에 Peak/Valley가 있어야만 매치가 일어나므로, 인접위치에 있는 Peak/Valley와는 매치되지 않는다. 이를 해결하고자 5x5 크기의 주변을 검사하여 정확한 위치에 있는 경우와 인접 주변에 있는 경우도 검출할 수 있도록 하였다. 가중치를 두어 정확한 위치에서의 매치의 경우 1을, 주변 3이면 w1, 주변 5이면 w2의 가중치를 부여하였다. w1, w2의 값은 1보다 작은 임의의 값으로 부여하였다.

W2	W2	W2	W2	W2
W2	W1	W1	W1	W2
W2	W1	1.0	W1	W2
W2	W1	W1	W1	W2
W2	W2	W2	W2	W2

[그림 3] 영역에 따른 가중치

이러한 가중치를 이용해서 검출된 Peak/Valley 위치 매칭값을 전체 기압 개수로 나누어 전체 Peak/Valley 중에 어느 정도 매칭되었는지를 실수값으로 표현하였다.

유사도 측정시 원본과 대상의 상호 유사도 값을 평균하여 이 두 데이터 간의 유사도 값으로 사용하였다.

$$M_{ab} = \{(\sum MatchValue_a[i]/N_a) + (\sum MatchValue_b[i]/N_b)\} / 2$$

[식 2]

M_{ab}: 데이터 A와 B의 상호 매칭 유사도

MatchValue_a[i]: 데이터 A의 Peak/Valley 매칭시 가중치값

MatchValue_b[i]: 데이터 B의 Peak/Valley 매칭시 가중치 값

N_a: 데이터 A의 Peak/Valley 개수

N_b: 데이터 B의 Peak/Valley 개수

3.2 한반도와외의 거리 정보를 이용한 유사도 측정

한반도에 영향을 미치지 못하는 곳에서의 고/저기압이 매칭이 일어난 경우와 한반도 주변에서 매칭이 일어난 경우, 한반도에서 가까운 곳에서의 매칭에 높은 가중치를 주어 검색 시 효과적인 검색 결과가 나올 수 있도록 하기 위한 방법이다. 즉, 거리가 먼 곳에서 매치가 되면 가중치가 줄고 가까운 곳에서 매치가 되면 그만큼 가중치가 증가된다. 각 고/저기압의 거리 정보와 그에 따른 가중치는 [식 1]을 통해 구해진다. 그러므로 유사도는 [식 3]와 같이 계산된다.

$$M.D_{ab} = \{(\sum MatchValue_a[i]/N_a) * W_1 + (\sum MatchValue_b[i]/N_b) * W_2\} / 2$$

[식 3]

M.D_{ab}: 거리정보 가중치를 이용한 상호매칭 유사도

MatchValue_a[i]: 데이터 A의 Peak/Valley 매칭시 가중치 값

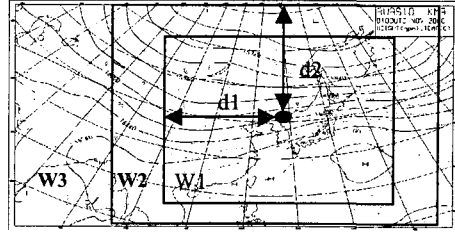
MatchValue_b[i]: 데이터 B의 Peak/Valley 매칭시 가중치 값

N_a: 데이터 A의 Peak/Valley 개수

N_b: 데이터 B의 Peak/Valley 개수

W₁: 매칭 Peak/Valley에 대한 거리 정보를 이용한 가중치 값

거리 가중치는 [그림 4]와 같이 적용된다.



[그림 4] 거리에 따른 가중치

3.3 등고선 정보를 이용한 유사도 측정

매칭된 Peak/Valley에 대해 각 Peak/Valley 간의 8 방향에 대한 등고선 정보를 서로 비교하여 그 차이를 누적하고 이 값이 가장 작은 것이 가장 큰 유사도를 갖도록 하였다.

For I=0 to I<count

For J=0 to J<8

sum = sum + |Matched_a[I].C[J] - Matched_b[I].C[J]|

Matched_a[I].C[J]: 데이터 A의 매칭이 일어난 I번째의 Peak/Valley의 J번째 등고선 정보

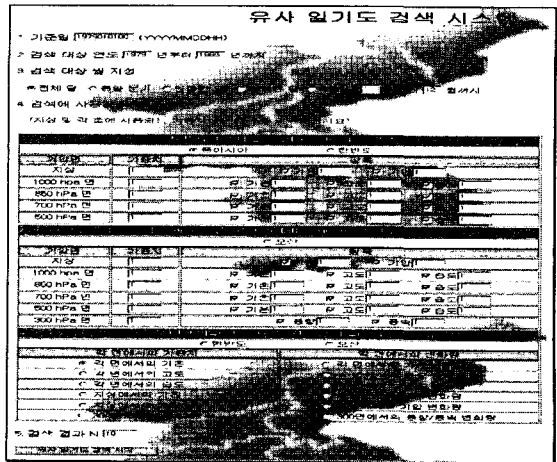
sum: 등고선 정보의 차이값의 합

$$M.D_{ab} = M.D_{ab} - sum * C$$

C: 가중치 상수

등압선 정보가 유사도에 미치는 영향을 반영하는데 있어 너무 큰 영향을 주지 않게 하기 위해 가중치 상수 C를 두어 적절히 조절하였다. 최종적으로 유사도 측정은 Peak/Valley의 위치 정보를 이용하여 인접 위치에서의 매칭 정보 및 원본과 대상의 유사도의 평균값과 한반도에서의 거리 정보를 반영하고 각 Peak/Valley의 등고선 정보를 반영한 것을 유사도로 사용하였다.

4. 구현 결과



[그림 5] 유사 일기도 검색 프로그램 실행 화면

유사 일기도를 검색하기 위해 Explorer 나 Navigator 등의 웹 브라우저를 사용해서 유사 일기도 검색 웹 사이트에 접속하면 [그림 5]의 화면이 나타난다.

검색하고자 하는 target 이 되는 날짜와 검색 대상 연도와 월 및 검색에 사용될 특징과 특징에 대한 가중치 등을 입력한 후 검색을 시작한다. 시스템에서는 target 에 대해서 전처리를 하고 격자데이터 DB로부터 위치정보와 한반도 중심으로부터의 거리정보, 등고선 정보를 이용하여 매칭을 하고 유사 일기도 검색 결과를 보여준다.

4.1 검색 예

1979년에서 1995년까지의 일기도 중에서 1979년 1월 1일 0시의 일기도와 유사한 일기도를 검색하고자 하였다. 검색 대상 월은 입력된 달 즉, 1월의 일기도만을 대상으로 검색을 하도록 하였다.

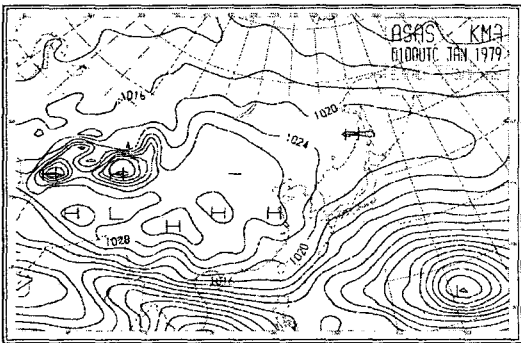
동아시아 지역의 각 면(hpa)에서의 Peak/Valley 및 가중치에 의한 검색 방법으로 5개의 검색 결과를 얻도록 선택사항을 설정하도록 하였다. 지상의 가중치는 2이고 나머지 면은 1의 가중치를 주었으며 850 hpa 면에서의 습도와 700 hpa 면에서의 고도/습도, 500 hpa 면에서의 고도/습도는 검색 대상에서 제외시켰다.

그 결과 다음과 같은 유사일기도 검색 결과를 얻을 수 있었다.

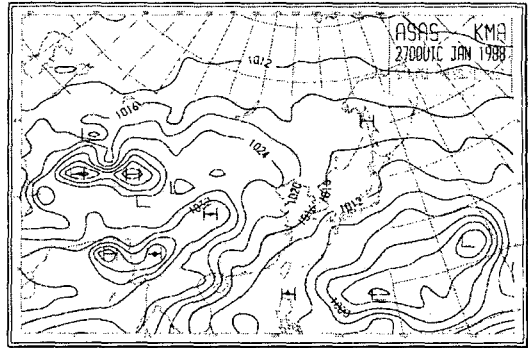
순위	매치된 날짜(YYYYMMDDHH)
1	1979010118
2	1979010200
3	1979012400
4	1988012700
5	1979010400

[표 1] 유사 일기도 검색 결과

[표 1]의 검색 결과를 보면 4번째 우선순위로 나온 일기도를 제외하고는 입력된 날짜 주변 시간대의 일기도임을 확인할 수 있다. 비슷한 시간대의 일기도가 서로 유사할 가능성이 높다는 것은 일반적이다. 약 9년 정도의 차이가 있는, 입력된 날짜와 4번째 우선순위로 검색된 날짜의 일기도를 다음에서 보이고 있다.



[그림 6] 기준 일기도(1979년 1월 1일 0시)



[그림 7] 검색된 일기도(1988년 1월 27일 0시)

5. 결론

본 연구는 동아시아 지역에 대한 NCEP 데이터의 1000hPa, 850hPa, 700hPa, 500hPa 에서의 기온과 고도, 300hPa 에서의 풍향과 풍속, 지상의 기온, 기압을 대상으로 지상 일기 유형을 Peak/Valley 의 수 및 위치, 주변 등고선의 유사성을 분류하고 유사 일기도 검색 시 특징 값의 비교를 통해서 유사성 정도를 검색하는 알고리즘을 개발하였다. 본 연구의 결과물은 각 격자점에서 획득한 수치 데이터를 이용하여 패턴을 분류함과 동시에 유사 패턴을 추출하기에 지상, 상층 구별 없이 사용 가능하며, 기압, 온도, 습도 등 다른 형태의 일기도에 대해 적용 가능한 범용성을 갖는 장점이 있다.

보다 정확한 검색을 위해서는 본 연구에서 사용한 알고리즘 외에도 다양한 특징들을 추출하는 알고리즘이 요구된다. 또한, 검색 성능 향상을 위해 클러스터링 등의 기법이 필요할 것으로 생각된다.

참고문헌

- [1] 최성식, "동아시아 일기 유형에 따른 우리나라 강수 분포 특성", 석사 학위 논문, 부산대학교, 1998
- [2] 서은진, "동아시아 일기 유형과 관련된 한반도의 산성 강우 현상 특성 연구", 석사 학위 논문, 부산대학교, 1999
- [3] Oliver Monga et al., "Using Partial Derivatives of 3D Images to Extract Typical Surface Features", CVGIP, vol. 61, no., 2, pp. 171-189, March, 1995
- [4] A Guezic and N. Ayache, "Smoothing and Matching of 3D space curves", Int. J. Computer. Vision 12, pp. 79-104, Feb. 1994
- [5] P. Sander and S. W. Zucker, "Inferring surface trace and differential structure from 3D images". IEEE Trans., PAMI, 12(9), 1990
- [6] M. Do Carmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, Englewood Cliffs, NJ, 1976
- [7] K. Normizu and S. Kobayashi, Foundations of Differential Geometry, Wiley, NY, 1969
- [8] R. Gonzalez, and R. Wood, Digital Image Processing, Addison-Wesley Pub. Co., 1992
- [9] A. Jain, Algorithms for Clustering Data, Prentice-Hall, Englewood Cliffs, NJ, 1988
- [10] B. Jahne, Digital Image Processing, Springer, 1997
- [11] G. Ritter, Computer Vision Algorithms in Image Algebra, CRC Press., 1996
- [12] J. Foley and A. Dam, Fundamentals of Interactive Computer Graphics, Addison-Wesley, 1984