社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

台風時系列画像のマルチプルアラインメントに基づくデータマイニング

北本　朝展†

† 国立情報学研究所　〒 101–8430 東京都千代田区一ッ橋 2–1–2
E-mail: †kitamoto@nii.ac.jp

あらまし　台風は複雑な時空間パターンを示す自然現象であるが，複数の台風時系列画像が似たような時間発展を示すこともあることから、そのような事例を何らかの方法でまとめることにより、台風の時空間パターンをより明確に理解することができるようになると期待できる。このような複数の時系列信号を比較しまとめるための手法として、一般的に動的計画法やクラスタリングなどの方法が知られているが、本論文では特に複数の時系列を並べて比較するというマルチプルアラインメントの問題に、動的計画法やクラスタリングを発展させた方法を提案し、それを実際の台風時系列画像のアラインメントに適用する。その結果、4 万件以上の大規模台風画像データベースの中から、10 系列以上が似たような変化をしている部分系列を発見し、それが台風の典型的な発達過程に対応していることを示した。
キーワード　台風画像コレクション、時系列画像、マルチプルアラインメント、動的計画法、階層的クラスタリング

Data Mining from the Multiple Alignment of Typhoon Image Sequences

Asanobu KITAMOTO†

† National Institute of Informatics,　Hitotsubashi 2–1–2, Chiyoda-ku, Tokyo, 101–8430 Japan
E-mail: †kitamoto@nii.ac.jp

Abstract　Typhoon is one of the most severe atmosphereic phenomenon with highly complex spatio-temporal patterns — some of the typhoon sequences, however, shows relatively similar course of evolution. This suggests that the grouping of such similar cases may lead to clearer understanding of the spatio-temporal pattern of the typhoon by comparing multiple sequences. Generally known methods of this kind includes dynamic programming and clustering techniques, and in this paper, we develop those dynamic programming and clustering techniques for the problem of multiple alignment, the comparison of multipe time series through the alignment of sequences, and applied this proposal to the alighment of typhoon image sequences. As a result, we discovered a multiple alignment of more than 10 sequences out of the typhoon image collection of 40,000+ images. Moreover, the result of multiple alignment illustrates a typical process of typhoon evolution over multiple typhoon sequences.
Key words　Typhoon Image Collection, Image Sequence, Multiple Alignment, Dynamic Programmming, Hierarchical Clustering

1. 序　　論

複数の時系列画像が部分的に似たような時間変化を示すとき、これら部分系列を類似例としてまとめ、典型的な時間変化を見出したい、というニーズが生まれる。本論文はこのような問題に対して、時系列画像のマルチプルアラインメント法を提案する。また本論文では、台風時系列画像から典型的な発達過程を発見するという問題にこの方法を適用することで、台風のライフサイクルをモデル化するという問題にも取り組む。

台風のライフサイクルは次章で述べるように、通常はいくつかの段階にわけて説明されることが多い。ゆえに、ライフサイクルをいくつかの段階に分割することは自然な発想である。しかしこの問題の難しさは、時系列画像中に分割の基準となる明確なサインが存在しないところにある。例えば、似たような問題に映像信号の分割という問題があるが、その場合は「シーンチェンジ」などの明確なサインが存在することもあり、映像信号の分解結果はだいたい直観的である。それに対して台風時系列の場合は、そもそもが連続的な時系列信号であるため、ある場所で分割することに明確な根拠を見出すことは難しい。

そこで本論文では、まず時系列画像のアラインメント、すなわち時間方向に伸縮可能な対応付けを用いて、複数の時系列画像の類似部分系列を取り出すことを試みた。次にその対応付けが本当に信頼できるものかを確かめるために、階層的クラスタリング、他系列を用いたアラインメントの検証、コンセンサス

フラグメントの計算など、複合的な処理を組合せたマルチプルアラインメント法を試した。その結果、台風の典型的なライフサイクルを示す部分時系列画像を、大量の台風画像データベースから発掘することができた。このような手法を用いて規則的な時間発展と不規則的な時間発展とを区別することによって、台風の解析や予測に有用な知見を見出すことが研究の最終的な目標である。

なお本論文の手法は、本質的には台風時系列画像のみを対象とする手法ではなく、データ間の距離尺度を定義すれば任意のデータにも適用可能なものである。しかし本論文の関心が台風時系列画像という具体的な対象にあることから、以下では台風時系列画像に適用した場合のみに話題を絞って議論を展開する。

2. 台風の一生とライフステージ

台風は中心付近の風速が大きい熱帯低気圧であり、日本付近に接近・上陸するたびに、大きな被害（と恩恵）をもたらす気象現象である。ゆえに社会的な関心・重要度も非常に高いことから、気象現象としては特別の監視体制が敷かれており、また過去の記録についても網羅的に整備が進んでいる。特に気象衛星の出現以降、台風の一生を宇宙から克明に追跡することが可能となったため、台風が発生から消滅までの間にどのような時間発展を示すか、という問題について、少なくとも現象論的には大きな蓄積がある。そのような豊富な経験から実用レベルの手法として確立してきたのが、ドボラック (Dvorak) 法である [1]。

ドボラック法は、台風の見かけの雲パターンから台風の実際の強度を推定するパターン認識手法であり、特に専門家が目視でパターン認識するための手引として使われるものである(注1)この方法では、台風の雲パターンをいくつかの典型的なパターンに分類し、それらが台風の発達過程にしたがってどのように形態変化するかなど、長年の経験に基づく典型的時間発展パターンを取り上げている。

では典型的な時間発展パターンとは何だろうか。一般的にはこのパターンを以下の 4 段階にわけて理解することが多い。

（1） 発生期 (birth)
（2） 発達期 (growth)
（3） 成熟期 (maturity)
（4） 衰弱期 (decay)

実際に台風の雲パターンをこれらの段階にわけることができれば、例えば発達の直前にあるような台風の雲パターンを識別し、あらかじめ警戒を強めておくことなどが可能になるため、このような分類に関する実用的な価値は大きい。しかしそれぞれの期間に特有のパターンなどはあまりよくわかっておらず、しかも上記の期間が全体の何割を占めるかは台風によって異なるため、台風の一生を機械的に分割することはできない、という点には注意すべきである。より確かな方法は、各台風の個性をも考えながら、何らかの統計的基準を用いて分割する方法であると考える。このような統計的基準の信頼性を高めるためには、まず大量のデータが必要である。

(注1)：なぜ台風の勢力を雲パターンから解析する必要があるのかというと、台風が大洋上にある場合、中心気圧等を実測できないためである。

表 1　ベストトラックおよび台風画像コレクションの概要。2002 年 11 月現在で約 42,200 件の画像を収集している。

	北半球	南半球
ベストトラック		
観測機関	気象庁 (JMA)	オーストラリア気象局 (BOM)
緯度範囲	赤道より北	赤道より南
経度範囲	$100°E \sim 180°E$	$90°E \sim 170°E$
台風画像コレクション		
台風シーズン	8 (1995–2002)	6 (1995–2000)
台風系列数 / 総画像数	182 / 31,400	70 / 10,800
台風系列当たり画像数	53 ～ 433	25 ～ 480

3. 台風画像コレクション

本論文ではそのような大量のデータとして、表 1 に示す台風画像コレクションを用いる [2]。これは、気象衛星ひまわり 5 号の画像から台風中心を決定し、ランベルト天頂等積図法を用いて投影するという方法で収集した台風画像のコレクションであり、これまでに収集した台風画像は 42,200 件を越えている。また現在では、半自動的にではあるが、台風発生と同時に自動的にデータベースを更新する機能も備えている(注2)。

図 3. はこの台風画像コレクションの多様性を示すために、全画像にクラスタリング手法を適用してみた例である。いわゆる台風のような典型的な渦巻き状のパターンだけではなく、斜めに伸びたパターンや、渦巻きが完全に崩れたパターンなど、かなり多様な雲パターンが実際には存在することが明瞭である。

ここでは、あくまで個々の画像をばらばらのものとしてクラスタリングを適用したが、実際にはこれらは時間的に連続する画像信号であり、しかも発生初期に頻出するパターンや、消滅期に頻出するパターンなど、ライフサイクルに依存するパターンが多い。よって、パターンの時系列情報から、そのようないわば「状態遷移」を検出し、その情報を台風解析や台風予測に活用できないか、と考えている。これが本論文の目的である。

なお、以下の実験ではすべて北半球台風画像コレクションを用いるため、系列数は $N = 182$ となる。

4. 台風時系列画像のアラインメント

4.1 動的計画法

まずは台風時系列画像のアラインメント（整列）について検討する。ここで、台風雲パターンは、発生期から衰弱期まで変化が「逆戻り」することはない(注3)と仮定し、形態が時間方向に単調に変化すると仮定する。このような仮定のもとでは、2 つの時系列のアラインメント問題は、動的計画法とよばれる効率的な計算法を用いて解決できることが知られている。

(注2)：http://www.digital-typhoon.org/

(注3)：例えば、熱帯地方に存在する間には、いったん衰弱したあと再び発達期に入ることはあるが、温帯低気圧への変化期から発生期の形態に戻ることはまずない。

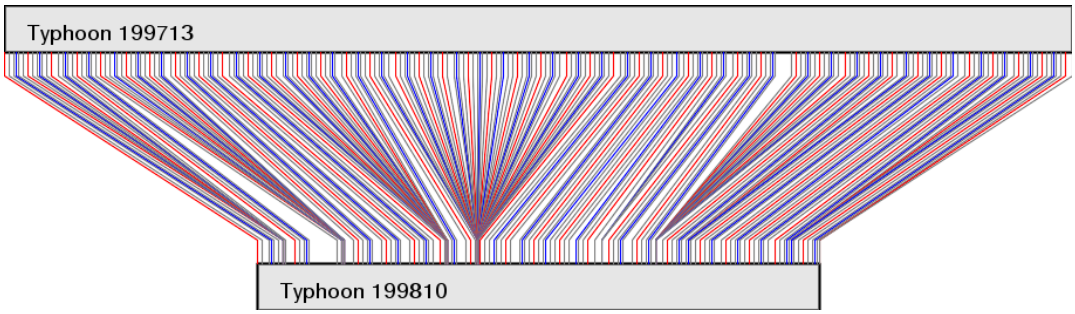


図 2　動的計画法による台風 199713 号と台風 199810 号とのアラインメント。

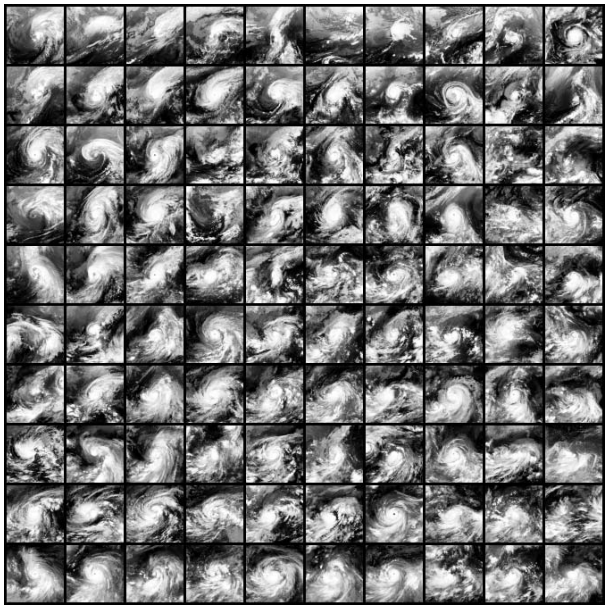


図 1　K-平均法による台風画像のクラスタリング。クラスタリング後には、Sammon のマップ法を用いて 2 次元平面上に配置している。

まず系列 S_a と系列 S_b とをグローバルに比較する問題を考える。ここで各系列の要素は必ず相手方のどれかの要素とマッチングしなければならない、という制約をつける。両者の長さが異なる場合には、複数の要素とマッチング M する要素を出現させることになる。このときこのアラインメントの最適対応付けとその時のマッチングスコアを計算したい。

まず系列 S_a、系列 S_b とをアラインメントするとし、それぞれの長さを m、n とする。すると、部分系列 $S_a[1..i]$ と $S_b[1..j]$ のアラインメントスコア $D(i,j)$ は、以下の再帰式を解けば得られる。

$$D(i,j) = p(i,j) + \min \begin{cases} D(i-1,j) \\ D(i-1,j-1) \\ D(i,j-1) \end{cases} \qquad (1)$$

ここで $p(i,j)$ は要素 $S_a(i)$ と要素 $S_b(j)$ とをマッチングさせるためのコストである（付録 1 参照）。この再帰式を解くことで、系列間のグローバルなアラインメントが可能となる。

例えば台風 199713 号と台風 199810 号との対応について計算した結果が図 2 である。両方の時系列は左から右に進行しており、両台風の寿命が異なるためにその全体的な長さは異なっているが、両系列の要素どうしの距離の総和が最小になるような対応付けを、動的計画法により計算することができる。

図 3　ピボット系列とのアラインメント。

5.　ピボット系列とのアラインメント

上記の方法を用いると、任意の 2 系列間のアラインメントが可能となったため、次にある系列（本論文ではピボット系列と呼ぶ）を固定したときに、他の系列がどのようにアラインメントできるのかを調べる。すなわち、全体で N 系列があるとすれば、あるピボット系列と残りの $N-1$ 系列とのアラインメントを実行することになる。これは動的計画法を $N-1$ 回適用すれば得られるものである。

さてここで、グローバルなアラインメントだけではなく、部分系列に対するアラインメントも行いたい。なぜなら、全体としては異なる変化をしていても、部分ごとには似た時間発展を示す場合も多いためである。このような部分系列のことを、本論文ではフラグメント（断片）とよぶ。そのような断片ごとのアラインメントの方法はいくつか考えられるが、本論文の目的は、同じような時間変化を同じような時間スケールで示しているフラグメントを抽出することにある。そのような観点から図 2 に着目すると、アラインメント結果には、両系列でほぼ平

行にマッチングしている部分と、1 点に集中してマッチングしている部分が存在していることがわかる。この平行にマッチングしている部分は、同じような時間スケールで同じような時間変化を示しているフラグメントの候補になると考えられる。

そこで、ほぼ平行(注4)にマッチングしている部分をフラグメントとして抽出し、これを以後の処理に用いる方法を提案する。こうして抽出したフラグメントを、ピボット系列を基準にしたアラインメントとして描いたのが図 3 である。ここで上下の順序は、系列間の平均距離、すなわち系列間のマッチングコストをマッチング要素数で割った値、の昇順に並べ替えており、平均して類似した系列が上部に現れている。

6. 階層的クラスタリング

こうして得られたフラグメントについて、次に似たフラグメントどうしをグルーピングすることを考える。つまり、系列間のグローバルな類似度ではなく、むしろフラグメントどうしの類似度を考え、似たような時間変化のフラグメントをグルーピングすることで、台風のライフサイクルに対応する類似部分系列を発見することを目論む。そのような方法は一般にクラスタリングとよばれ、例えば図 3. で用いた K-平均法もその一つの代表的な方法であるが、ここでは階層的クラスタリング手法を適用する。

この階層的クラスタリングとは、距離の近いものから同じクラスタであると判定し融合していく時に、K-平均法とは異なり、距離が最も小さい 2 つのデータを融合してボトムアップにクラスタを形成していく点が異なる。系列数を N とすると、まずそのすべてを対にして $\frac{1}{2}N(N-1)$ 組の距離を計算しておき、後で新しく生成されたクラスタと個体間の距離は漸化式で更新する。その方法としてよく知られているのが、(1) 最短距離法、(2) 最長距離法、(3) メディアン法、(4) 重心法、(5) 群平均法、(6) Ward 法である [3]。

本論文では、フラグメント間の距離尺度としては付録 2 に示すものを用い、特に最短距離法および Ward 法を適用した場合を比較した結果が図 4 である。よく知られているように、最短距離法は連続したクラスタを生み出しやすく、あまり自然なクラスタリング結果が得られない。その他の距離尺度に関してはまだ定量的な比較はできていないが、以下では Ward 法を用いてクラスタリングをおこなう。なお、クラスタリング数は 10 と設定している。、

7. マルチプルアラインメント

こうして、ピボット系列を固定した場合のアラインメントについては、

（1） フラグメントの切り出し

（2） フラグメントの距離の計算

（3） フラグメントのクラスタリング

という手順によって計算することができた。次にピボット系列を系列全体に動かした時のフラグメントのアラインメント、すなわちマルチプルアラインメントについて考える。

その基本は「多数決による決定」である。つまり多くのフラグメントが同じような位置でアラインメントしていれば、それを意味のあるアラインメントであると仮定する方法である。具体的には以下のようなアルゴリズムを提案する。

a） 不要なアラインメントの除去

まず平均距離が大きいアラインメントは、たまたまマッチングしただけのものであると考え、これらを除去する。具体的には、個々のピボット系列に対してそれぞれ、平均距離が小さい順に上位 50 件(注5) のアラインメントを取り出す。

b） 他系列を用いたアラインメントの検証

2 系列間のアラインメントが「たまたま」でないことを示すために、ここでは他系列の助けを借りることにする。具体的には以下のアルゴリズムを提案する。

（1） ピボット系列として S_p を選び、S_p に対して適用したクラスタリング結果から、同じクラスタに属する 2 つのフラグメント F_q と F_r を選び出す。

（2） F_q と F_r は S_p を基準としてたまたまアラインメントしているが、それが意味のあるものであるために、以下の条件を満たすものとする。

（a） S_q をピボット系列としたときに、F_r と S_p に属するフラグメント F_p がアラインメントしている。

（b） 同様に S_r をピボット系列としたときに、F_q と S_p に属するフラグメント F_p がアラインメントしている。

、また、それぞれの場合のフラグメントが実際に十分に重なり合っているかも同時に確認する。

（3） 以上の条件が満されるとき、F_p、F_q、F_r は意味のあるアラインメントを構成しているとみなす。

つまり常に 3 系列を参照することで、より確かなアラインメントを得ようとするのが、このアルゴリズムの狙いである。

c） 連結成分の探索

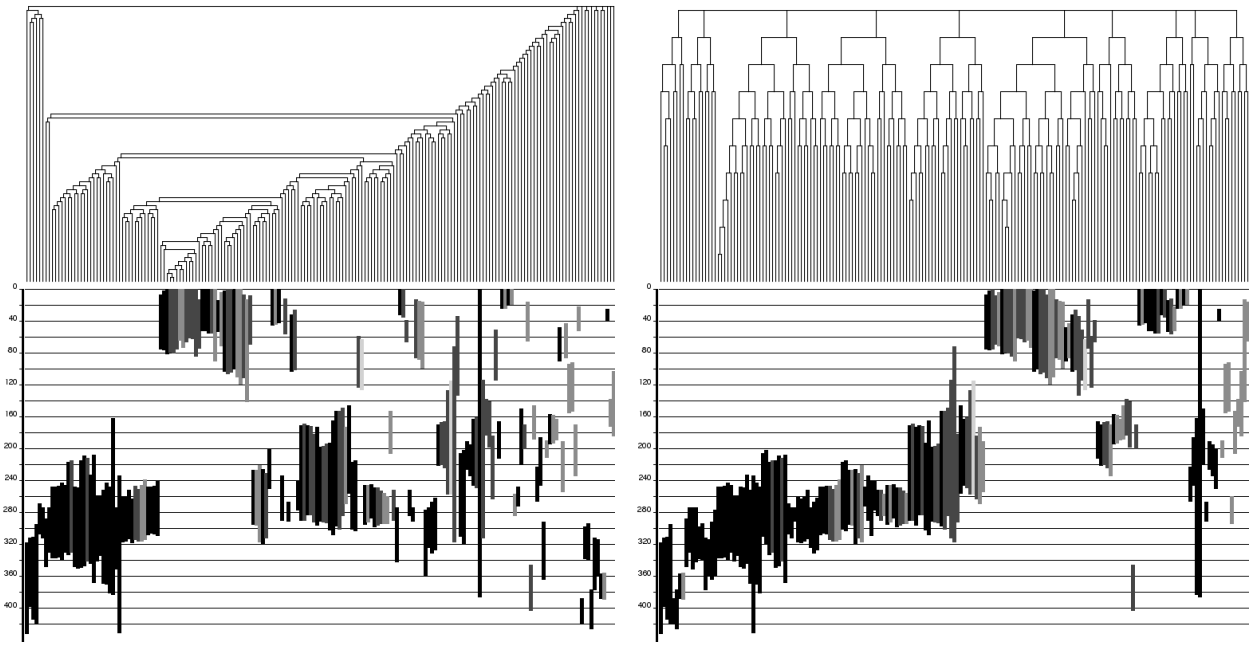
以上の操作を、ピボット系列を全系列にわたって動かすことで、3 フラグメントを単位とするアラインメント・トリオが多数出現する。これはフラグメントをノードとする 3 頂点のクリークを生成することに対応すると考えてもよい。ゆえに、ノードを共有するクリークを連結成分とみなすことによって、形成された多数のアラインメントの中にいくつの連結成分があるかを計算し、一つの連結成分に含まれるフラグメントを相互にアラインメントしているものとみなす手法が考えられる。このような連結成分の探索は、深さ優先探索により簡単に実現できる。

d） コンセンサスフラグメントの計算

こうして探索した連結成分には、同じピボット系列とアラインメントしているフラグメントが複数含まれている。そこで最後に、これらのフラグメントを用いて投票をおこない、ピボット系列上のフラグメントを確定する。例えばピボット系列 S_p とアラインメントしているフラグメントを $F_m(m=1,\ldots,M)$ とし、フラグメント F_i とアラインメントしているピボット系

(注4)：具体的には最大許容できる時間ずれをしきい値として与える。

(注5)：この 50 という数字には本質的な意味はないが、この程度の値より大きくすると、無意味なアラインメントが増加してくるというのが観察結果である。



(a) 最短距離法を用いた場合　　(b) Ward 法を用いた場合

図 4　階層的クラスタリングの結果の比較。クラスタ数は 10 であり、フラグメントは図 3 と同じである。

列の座標を $[x_m^s, x_m^e]$ とすると、ピボット系列上のコンセンサスフラグメントは $[\frac{1}{M}\sum_{m=1}^{M} x_m^s, \frac{1}{M}\sum_{m=1}^{M} x_m^e]$ と定義する。この計算を、連結成分に含まれるすべてのピボット系列に対しておこなうことで、フラグメントのマルチプルアラインメントが得られる。

8. 実験および考察

以上の方法を北半球台風画像コレクションに適用し、台風時系列画像のマルチプルアラインメントを探索した。その結果として得られた 10 系列のマルチプルアラインメント結果を図 5 に示す。これらはいずれも約 80 要素長のフラグメントであることから、実時間でおよそ 3 日半の間、類似した時間発展を示した台風系列群であることを示している。この他にも 24 系列のマルチプルアラインメントなど、合計で 90 個程度のマルチプルアラインメントを発見することができた。

図 5 をもう少し詳しく分析すると、この結果は典型的な台風のダイナミクスを表現しているように見える。すなわち、一連の時系列は、台風のライフサイクルの典型的な推移、すなわち発生期、発達期、成熟期、衰弱期を示していると考えることができる。左端の画像は発達期に入りかけの段階であるが、中間あたりで成熟期に入り、右端の段階では衰弱期（温帯低気圧化）に入っている。ゆえに、このように整列することによって、個々の画像の意味づけや時空間的なパターンの変化をより明確に把握することができる。また、このように人間が見るだけではなく、その特徴付けをさらにコンピュータにまかせるのも、これからの課題である。

ただし、これらのアラインメント結果は、さまざまなパラメータの選択や距離尺度の選択などに依存していることは注意しておく必要がある。パラメータの選択でアラインメント結果が変わる可能性は高く、特に階層的クラスタリングの部分はノイズに弱い部分として注意を払うべきである。また根本的な問題としては、本論文で用いている台風の固有画像表現が、限定的な有効性しか持たない点であろう。例えば台風の眼の有無といった情報は全く考慮されていないため、このような情報を考慮できるような画像表現モデルを用いることによって、より気象学的にも意味のあるアラインメントが可能となると期待できる。

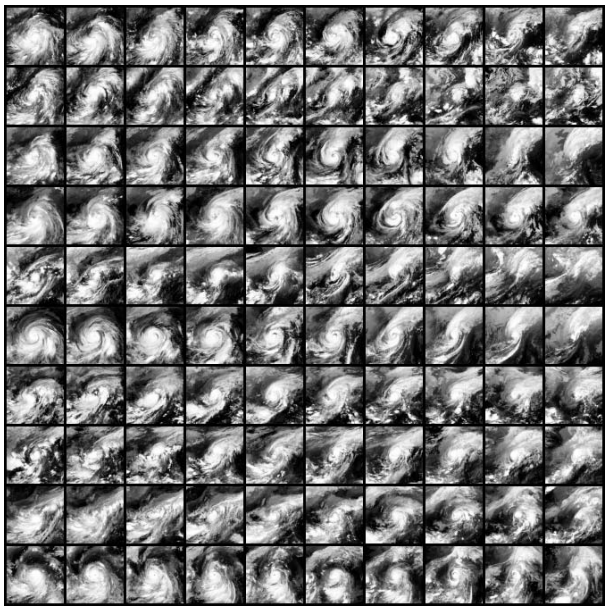


図 5　台風時系列画像のマルチプルアラインメント。各行が台風系列に対応しており、時間は左から右に流れている。

9. 結　　論

本論文では台風時系列画像を対象とした、時系列信号のマルチプルアラインメントのための方法を提案し、この方法によって得られた台風時系列フラグメントのマルチプルアラインメント結果を示した。そしてそれが台風の典型的なライフサイクルを示しており、複数の台風時系列画像の比較がより明確になることも示した。

本論文の実験はまだ初期段階にあり、まだ詰めなければならない点がたくさんある。例えばパラメータや距離尺度の選択などもそれらの課題であるが、もっと興味深い方向性としては、これらのアラインメントに特徴ベクトルの距離尺度をのまま使うのではなく、特徴ベクトルから計算される（隠れ）状態の間のアラインメントとして定式化する、という方向性がある。このように特徴ベクトル系列から隠れ状態を計算する方法については、隠れマルコフモデルを用いた議論をすでに始めているが、これらのモデルとの統合によって、異なる距離尺度を用いたマルチプルアラインメントを比較することも今後の課題である。

文　　献

[1] V.F. Dvorak. Tropical cyclone intensity analysis using satellite data. *NOAA Technical Report NESDIS*, Vol. 11, pp. 1–47, 1984.

[2] A. Kitamoto. Spatio-temporal data mining for typhoon image collection. *Journal of Intelligent Information Systems*, Vol. 19, No. 1, 2002. 25–41.

[3] 高木幹雄, 下田陽久（編）. 画像処理ハンドブック. 東京大学出版会, 1991.

[4] A. Kitamoto. Interpretation of typhoon cloud patterns by holistic analysis. *Technical Report of the Institute of Electronics, Information and Communication Engineers*, Vol. PRMU2000-240, pp. 129–136, 2001. (Japanese).

付　　録

1. 要素のマッチングコスト

本論文では台風画像時系列として、具体的には台風雲パターンの特徴ベクトルの時系列を用いた。この特徴ベクトルは、台風画像コレクションの（地域別）全画像から、台風の空間的パターンの固有ベクトル（固有台風画像）を求め、そこから特徴ベクトルを計算するという方法に基づいている [4]。ゆえに各要素は特徴ベクトルとなっており、要素同士のマッチングコストは特徴ベクトルのユークリッド距離としている。なお北半球画像の場合、特徴ベクトルの次元数は 66 である。

2. フラグメント間の距離尺度

基準系列を i とし、その基準系列とアラインメントされた二つの系列を j および k とする。系列 j が基準系列とフラグメント $[x_s^j, x_e^j]$ でアラインメント（距離 d_{ij}）し、一方系列 k は基準系列とフラグメント $[x_s^k, x_e^k]$ でアラインメント（距離 d_{ik}）しているとする。このとき、系列 j のフラグメントと系列 k のフラグメントとの距離を以下のように定める。

$$O = \frac{\max\left(\min(x_e^j, x_e^k) - \max(x_s^j, x_s^k), 0.0\right)}{\max(x_e^j, x_e^k) - \min(x_s^j, x_s^k)} \qquad \text{(A·1)}$$

$$R = \frac{\min\left(d_{ij}, d_{ik}\right)}{\max\left(d_{ij}, d_{ik}\right)} \qquad \text{(A·2)}$$

$$S = (1.0 - O)^{y_o} (1.0 - R)^{y_r} \qquad \text{(A·3)}$$

ここで O はフラグメント同士の重なり具合を計算する式であり、完全に一致する場合には 1、重なりがない場合には 0 になる。一方で R は距離の比を計算するものであり、同程度の距離の場合には 1 となる。ゆえに、最後の S では、完全に一致し、しかも距離が同じ場合には 0、最も異なる場合には距離 1 となる。最後に y_o と y_r は両項の影響を制御するパラメータである。本論文では $y_o = 1.0$、$y_r = 0.3$ というパラメータを用いた。これは距離の類似性をやや過小評価するような類似度になっている。