

気象画像を用いた時空間変動における相関性マイニング

Mining of the spatio-temporal correlation from the weather imagery

坂口 祥太^{*1}

本田 理恵^{*2}

Shota Sakaguchi

Rie Honda

^{*1} 高知大学総合人間自然科学研究科 ^{*2} 高知大学理学部

^{*1}Graduate school of integrated Arts and Sciences, Kochi University, ^{*2}Faculty of Science, Kochi University

Two approaches for mining of the spatio-temporal correlation from the weather satellite imagery are examined: the presentation of the imagery as an animation with their blocks colored corresponding to the cluster labels based on the clustering of the texture information, and three-dimensional visualization of correlation coefficients between the time series of the average intensity of the standard point and the reference point in consideration of time lag. Both approaches are examined for MTSAT-1R IR1 images in September 2007, and the effectiveness of both approaches was confirmed.

1. はじめに

地球観測衛星によって取得されるデータは、観測機器や地上処理システムの性能向上により、多様化し、かつ膨大になっている。こうしたデータを効率的に要約して相関性の高い時空間変動パターンを抽出することができれば、地球環境変動や気象現象の理解・予測に活用することができる。しかし、処理体制の整備に比べてデータ量の増加が急激であるため研究者の処理が追いつかず、一部のデータが解析されないまま保管されてしまうといった問題も指摘されてきている。

この問題に対し、生駒ら[生駒 04]は NASA(米国航空宇宙局)や JAXA(宇宙航空研究開発機構)などの協力により得られた質、量共に充実したデータから地球環境に関する異種統合データベースを構成し、さらにこの集約データからグラフィカルユーザインターフェースを通してインタラクティブに時空間の相関関係を表示・分析できるシステムを構築した。ただしこのシステムは数値データを対象としたものであり、可視・近赤外画像の雲分布など、空間パターンを持ちながら時間的に変動する特徴については考慮されていなかった。

一方、時系列気象画像の画像特徴から知識発見をしようとする試みには、台風に着目した研究[北本 02]や、クラスタリングによって記号化された時系列からの共起ルールの抽出[Honda 02]などがあるが、本格的な時空間の相関性を画像特徴だけから求めることを目的とした研究は行われていない。

本研究では、画像集合からの時空間変動の相関性発見を支援するシステムの構築を時系列気象画像をテストベッドとして検討する。また、時空間の相関性を評価するには、基準点、参照点、時間ずれを、どのように統一的に扱うかという点が問題となる。よって、可視化によるヴィジュアルデータマイニングを考慮に入れて、抽出された相関係数などの情報を動画や、空間に時間を加えた 3 次元空間で等値面として可視化することにより、インタラクティブな発見支援を支援することを考慮する。

2. 時空間データにおける相関性

本研究では時空間での相関性を、生駒ら[生駒 04]と同様に基準点の時間変動に対する参照点の時間変動の相関性としてとらえる。生駒らの研究では、ユーザが Web インターフェース上

で基準点を指定することによって参照点との相関係数をカラーマップとして可視化し、ユーザはさらにその中から興味深い参照点を選択して、基準点と選択した参照点の時系列データを可視化して詳細に比較することができる。この手法はインドモンスーンのデータに適用されて成果をあげている。ただし、雲のように空間的パターン自体も意味を持つデータについては、生駒らのよう点对応の時系列データの比較だけで時空間の相関性の比較を行うことは難しい。

よって本研究では図 1 のように画像をブロック化し、ブロックの特徴を用いて相関性又は類似性を分析していくこととする。なお、ブロックの特徴としては、輝度の平均値とテクスチャに相当するFFTパワースペクトラムを採用するものとする。具体的な相関性の分析手法として、本研究では 2 種類のアプローチを考える。

1. ブロックごとのテクスチャのクラスタリングとクラスタIDに応じた着色を用いた動画表示
2. 時間ウィンドウ内の時系列データの相関係数とその 3 次元表示

なお、前者については著者らにより一部が発表されている[坂口 09]。以下、それぞれの手法について述べていく。

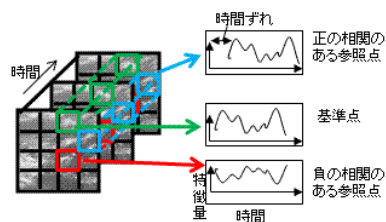


図 1. 時系列気象画像における時空間相関関係の概念図

3. テクスチャのクラスタ ID に応じて着色した動画による表示

まずブロックごとの雲の濃淡等の形状や分布パターンを表す情報としてテクスチャに着目する。図 2 に考慮した手法の概要を示す。ここでは、ブロックのテクスチャ情報として FFT パワースペクトルを用いる。テクスチャ情報をクラスタリングによっていくつかのグループに振り分け、時系列画像のブロックにクラスタIDのラベル付けを行いこのラベル情報の時空間相関関係を見ていくことによって、雲パターンの時空間の相関性を調べることができる。なお、ここでは時空間の相関性の発見については相関係数のような評価量を用いず視覚的発見に頼ることとする。具体的

にはクラスタIDに対応してブロックを色づけし、その動画を利用者が参照しながら発見することを支援するものとする。

得られたクラスタIDを着色による可視化に利用する際には、特徴の類似したものには似た色を与えることが必要である。よってクラスタIDの値は入力特徴の類似性を反映して連続的に変化するものでなければならない。このような性質をみたすものとして Kohonen の自己組織化マップ (SOM) がある。この手法では、入力データに基づく学習過程で、入力データに最も類似した競合層のユニットを発火させるとともに、そのユニット(勝者ユニット)と近傍のユニットが入力データに近づくように弱い学習をさせる。これにより競合層上での距離に近いユニットほど特徴が類似することになる。よって、競合層上のユニットの座標に対応させて連続的に色を割り振ることにより、特徴が似たブロックを類似した色で示すことが可能になる。

なお、この際に求めるクラスタの個数が未知であることが問題となる。SOM は本来はクラスタリングの手法ではなく、結果が 2 次元的な競合層に投影されることから、その配置が実行ごとに変わり結果の解釈が難しく、この手法だけから最適なクラスタの個数を判別することは困難である。そこで、前処理として、まず複数のクラスタ数を用いて K-means 法によるクラスタリングを反復試行し結果を評価することによって最適クラスタ数を求めることにする。この際、各クラスタ数で反復試行を行い、誤差関数で評価した最適解を採用することで初期値の影響を緩和する。また、さらに、クラスタ数を増大させた時の解の変化や、空のクラスタの出現などから最適なクラスタ数を判別する。こうして得られた最適クラスタ数に近いサイズの競合層を設定して SOM による最終的なクラスタリングを実行するものとする。

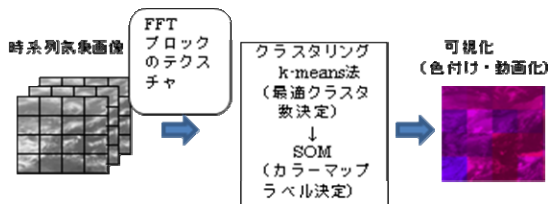


図 2. テクスチャによるカラーラベル付き動画化の概要

4. 輝度平均値の時空間相関係数の 3 次元表示

前節ではユーザによる視覚的な発見に強く依存するアプローチをとったが、ここでは直接的に相関係数を用い、正・負の相関のある地点を時間ずれも考慮に入れながら抽出する手法を検討する。

図 3 にこの手法の概要を示す。特徴量については、簡単のため、ブロックの平均輝度値を採用、すなわちブロック内の雲の総量の評価するものとして、前節で述べたような周期的パターンなどの空間分布は扱わないこととする。

時刻 t を視点とする基準点 (i, j) と、時間差 Δt の参照点 (i', j') を考え、それぞれの時系列データを

$$X = \{P(i, j, t) \mid k = 1, 2, \dots, n\},$$

$$Y = \{P(i', j', t + \Delta t + k) \mid k = 1, 2, \dots, n\},$$

で表すものとする。ここで $P(i, j, t)$ は座標 (i, j) のブロックの時刻 t の平均輝度値、 n は時間ウィンドウ幅とする。このとき、基準点と参照点の間の相関係数は下記のとおりとなる。

$$r = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum (x_i - \bar{X})^2} \sqrt{\sum (y_i - \bar{Y})^2}}$$

このようにして得られた相関係数を、2 次元の空間に時間差を加えた 3 次元空間の中に等値面表示することにより、時空間の相関性の高い箇所とその連続性を可視化することにする。

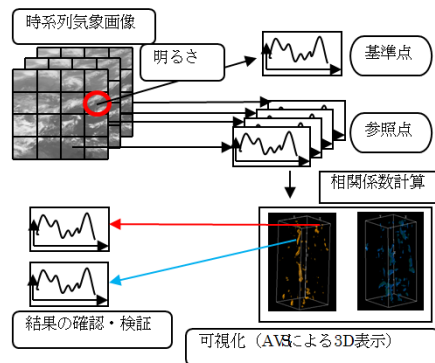


図 3. 時空間相関の可視化の概要

5. 実験

5.1. 実験データ

実験には気象衛星ひまわり 6 号 (MTSAT-1R) により得られた東経 70 度-160 度、北緯 70 度から南緯 20 度 (GAME 領域) の画像を利用した [菊地 09]。なお、雲の動きをとらえるために特に $10.5 \mu\text{m} - 11.5 \mu\text{m}$ の赤外チャネル IR1 を選択した。また、台風が頻繁に発生する時期として、2007 年 9 月の 1 か月分の画像を選択した。サンプリング間隔は 2 時間、画像サイズは 1800×1800 pixel である。また、処理にあたっては画像サイズをあらかじめ 640×640 pixel にまで縮小した。

5.2. テクスチャの可視化実験

テクスチャの可視化実験に関しては、ブロックサイズを 64×64 pixel (1 枚当たり 10×10 ブロック) とした。K-means 法による最適クラスタ数の検討では、クラスタ数を 10-70 の範囲で変動させてそれぞれ 100 回の反復で誤差関数を用いて最良の解を比較した。この結果、クラスタ数 40 以上で空のクラスタ個数が急激に増加したため、最適クラスタ数としては 40 を選択し、SOM のマップサイズ(ユニット数)は最適クラスタ数に近い 6×6 を選択した。なお、SOM による学習の際にも、初期値の影響を削減するため、異なる初期値での複数回の試行を実施した。実験より得られた結果を分類すると 8 パターンの解が出現したが、その中から各ユニットに割りつけられたブロック画像の性質、個数、距離の分布の評価から最適な結果を決定した。

図 4 の左図に得られた最適解を示す。この図は SOM の出力層の各ユニット(クラスタ重心に対応)に各代表ブロック画像を貼り付けたものであり、特徴マップと呼ぶものとする。ここでは左下から右上に向かうほど雲量が増加する一方、左上から右下の頂点に近づくほど塊状のはっきりとした雲が現れることがわかる。よって、雲量やその分布パターンの特徴を反映したクラスタリングが得られていることがわかる。さらに、図 4 左図の特徴マップの各クラスタに図 4 右図のような、右上に近づくほど赤色に左下青色になるように作成したカラーマップを対応させて、着色した後、時系列画像を動画化した。

図 5 に色づけされた動画のスナップショットと、動画の観測から相関性のある可能性がある点として注目された 2 点のクラスタ ID の時間変化をグラフで示す。この 2 点のグラフを比較すると、時間ラグを有して正の相関のある区間が 1-2 か所存在することがわかる。この結果、ブロックの特徴ごとの色分けによる画像

特徴の可視化に概ね成功し、動画からの視覚的な相関性のある点の発見にも効果的であることを確認することができた。

しかし、得られた動画を参照するにあたって利用者は大量のデータを空間的に一気に観測すると同時に、その時間変化を記憶して比較しなければならず、大きな負担がかかる。よってこのような可視化手法は次節で述べるような手法によって相関性のある個所の絞り込みを行った後に、追検証手段として利用することが適当であると考えられる。

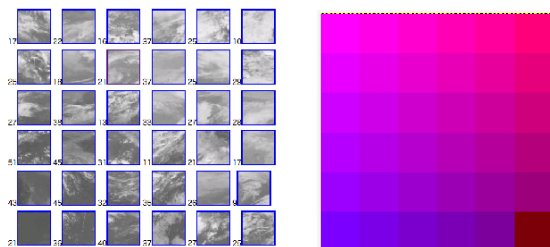


図 4. 各クラスターの代表ブロック画像 (SOM 学習結果) とそのカラーマップ

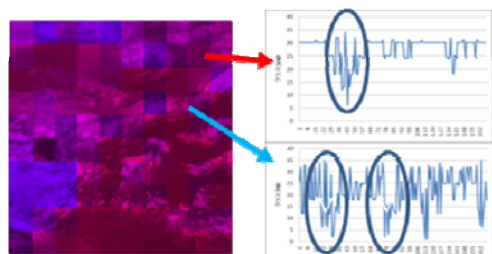


図 5. 色つき動画と 2 点のクラスター ID の時間変化

5.3 時空間相関の可視化実験

時空間の相関性の可視化実験では、特徴量として輝度平均値のみを使用した。そのため、ブロックサイズはテクスチャの実験よりも縮小して 16×16 pixel とした。また、時間ずれは 4 時間間隔、ウィンドウの長さは 15 日間とし、任意の基準点に対して、すべての参照点 (ブロック) と時間ずれに対して総当りで相関係数を求め、その値が正負の閾値を超えたものを等値面表示した。なお、基準点としては、台風発生地域に近く、大陸の影響を受けにくい太平洋上の点を設定した。

図 5 に基準点 (40,20,0: 東経 160 度、北緯 25 度) に対して求められた相関係数を AVS によって 3 次元空間内に等値面表示した図を示す。図 5 左は相関係数 +0.7、図 5 右は相関係数 -0.6 の等値面を表示している。これにより時空間に正負の相関の高い箇所とその分布を容易に確認できる。特に、正の相関では強い相関を表す箇所がチューブ上に複数現れていることが分かる。一方、負の相関については相関係数 -0.7 を下回るような箇所が

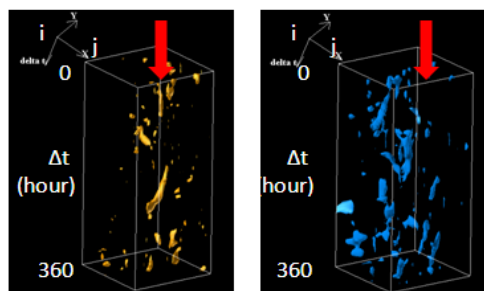


図 5: AVS による相関係数の等値面の 3 次元表示。x 軸 y 軸は空間軸、z 軸は時間軸で長さは 30 日。左は相関係数 0.7、右は -0.6 の等値面を表す。赤い矢印は基準点 (40,20,0)。

観測されることは稀であった。

図 6 は最も相関係数の大きい参照点 ((28,24,5step=20 時間)、相関係数 0.75) について、基準点とその点の時系列における輝度平均値の変動をグラフにしたものである。この 2 つのグラフからも 2 点の雲の総量の時間変化は 20 時間の時間差で正の相関をもって変動していることが確認できる。これは基準点に初め存在していた台風が時間とともに偏西風の影響を受けながら北上しその過程が対象点に影響を与えていることが原因と思われるが、より大量のデータを用いた再現性の評価と、物理的な意味についての専門家による評価が必要である。

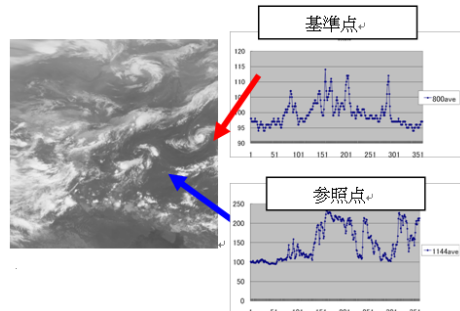


図 6. 基準点と高い相関を示した参照点での明るさの時間変化。縦軸は輝度平均値、横軸は時間で 30 日。

6. まとめ

本研究では時系列画像から時空間における相関関係を発見することを目的として、画像の特徴量から視覚的、解析的に相関性の高い箇所の発見を支援する手法を検討した。テクスチャのクラスタリング結果を用いて色付けされた動画は、ユーザによる空間パターンの変動の似通った部分の検出に効果があることがわかった。一方、輝度の平均値を用いた相関解析においては空間に時間ずれの次元を加えた 3 次元空間で等値面表示することによって、時空間における相関性を一度に確認できることがわかった。この結果、興味深いポイントをさらに時系列データとしてグラフ化することにより、詳細な調査を行うことができることがわかった。

今後はテクスチャ情報と平均輝度の両方の相関性あるいは類似度を 3 次元化するとともに、ユーザがインタラクティブに操作して評価を行うことをより積極的に支援する GUI を備えたシステムを作成し、その有効性を定量的に評価していく必要がある。

謝辞

本研究は科学研究費補助金 (18500114) の援助を受けて実施された。ここに謹んで感謝する。

参考文献

[生駒 04] 生駒栄治、谷口健司、小池俊雄、喜連川優: 大規模地球環境データアーカイブシステムにおけるデータマイニングツールの構築, DEWS2004 6-A-05, DEWS, 2004.
 [北本 02] 北本朝展: 台風コレクションからの予兆発見, 人工知能学会研究会資料, Vol. SIG-FAI- A, 2002
 [Honda 01] R. Honda and O. Konishi: Temporal rule discovery for time-series satellite images and integration with RDB, PKDD, 204-215, 2001
 [菊地 09] 菊地時夫: 気象情報頁, <http://weaheer.is.kochi-u.ac.jp>
 [坂口 08] 坂口祥太、本田理恵: 気象画像からの時空間変動マイニングの試み, 電気通信関係学会四国支部連合大会予稿, 徳島大学, 2008