

オーロラの出現・形状の予測に向けた 全天オーロラ画像の自動分類への試み

田中 孝宗^{*1}, 田中 良昌^{*2}, 佐藤 由佳^{*2}, 池田 大輔^{*1}

Toward Automatic Classification of Auroral All-Sky Images to Forecast Auroral Behaviors

Takanori TANAKA^{*1}, Yoshimasa TANAKA^{*2}, Yuka SATO^{*2}, Daisuke IKEDA^{*1}

Abstract

Aurora, which is attractive for many people, is an astronomical phenomenon related to many fields, such as interplanetary space, magnetosphere and ionosphere, and thus it is difficult to predict behaviors or shapes of aurora. The big goal of this research is to forecast them using observed data in different formats from different fields. To do that, we need training data which shows when and what types of aurora have appeared. In this paper, we evaluate three popular methods of automatic classification of images to classify auroral all-sky images. We found that two methods based on the local feature and the color histogram, both of which are expected to classify auroral all-sky images in detail, fail to capture characteristics of aurora in the preliminary experiment, while we obtained 92.3% of the classification accuracy based on auroral area, which can only classify whether an image contains aurora or not.

Keywords :Aurora, e-Science, Image classification, オーロラ, データ科学, 画像分類

概 要

オーロラは多くの人を魅了する自然現象だが、惑星間空間、磁気圏、電離圏などの多くの領域にまたがる現象であり、その物理モデルは完全には構築されていない。そこで、我々は観測分野を横断し、関連するデータを組み合わせながらデータ指向型科学の手法を用いて、オーロラの出現や形状等の予測を行うための研究を進めている。このような予測を実現するためには、いつ、どのようなオーロラが発生したのかという正解データを準備し、これを訓練データとして用いる必要がある。そこで本論文では、将来の機械学習によるオーロラ画像自動判定において良質な訓練データを得る準備として、国立極地研究所が公開している全天オーロラ画像に対して、オーロラの有無や、規模、雲の有無によって自動的に分類する画像処理の手法を適用し評価を行う。形状特徴による分類が可能になる局所特徴量を用いた手法と色のヒストグラムを用いた手法は、予備実験の段階で必要な精度がでないことが分かった。一方、HSV カラーモデルの閾値を満たす画素数でオーロラの有無を分類した場合、正答率が92.3%であり良好な結果を得た。

1 はじめに

オーロラは、その美しさにより多くの人々を魅了すると同時に、現代の我々の生活に関わりの強い天体現象である。例えば、大規模なオーロラが発生した場合には、人工衛星に障害が発生するなど、我々の生活にも

^{*1} 九州大学システム情報科学府 / 研究院 (Graduate School and Faculty of Information Science and Electrical Engineering, Kyushu University)

^{*2} 国立極地研究所宙空圏グループ (Space and Upper Atmospheric Science Group, National Institute of Polar Research)

影響がある [8]. しかし, 自然現象は一般に非定常な振る舞いを見せるため, その物理モデル化は難しく, 正確な予測は極めて難しいとされており [7], オーロラの出現予測に向けた様々な研究 [3] が行われてきたが, 完全に解明されていない事柄も多く, 詳細な出現地域の予測や, オーロラサブストームの開始時刻の予測には至っていない.

従来, 科学を支えてきたのは理論 (モデル), 実験, シミュレーションであったが, 近年のセンシング技術の発達やネットワークの充実等に支えられ, データ指向型科学 (Data-intensive Science) が科学を支える第 4 の柱として注目を集めている [1, 4, 9]. データ指向型科学の大きな特徴の一つとして, 精緻なモデルを構築せずに, 既存のデータを大量に用いて未知のデータを説明しようとする点にある. 例えば, コンピュータ将棋の分野では, 保木邦仁氏の作成した Bonanza[16] 以降, 大量の棋譜データをもとに評価関数のパラメータを自動的に学習することで, トッププロにも迫る強さを獲得している. 従来のコンピュータ将棋アルゴリズムが, プログラマが構築する将棋モデルの精緻さに依存していたことと対照的に, 保木氏は駒の動きを知っている程度であった. 他にも, 画像認識や音声認識等のパターン認識, 画像検索等のマルチメディアを対象とした情報検索, 化合物の活性予測分野にも適用され, 従来の精度を大きく改善している [17, 18].

本研究でも, データ指向型科学の立場から, 物理モデルを極力仮定せずに, 観測データから未知のデータの分類を目指す. しかし, オーロラは, 発生機構が惑星間空間, 磁気圏, 電離圏などの多くの領域にまたがる天体現象であるため, 様々な関連するデータが存在する (図 1 参照). 例えば, 太陽面画像等の画像データや, 太陽風パラメータや地磁気等の時系列データなどが存在する. さらに, 例えば画像データなどのデータ量は膨大で, 全天カメラの画像データの場合, そのサイズは 10 年間で 1.3 テラバイトに達する. 先述の通り, デー

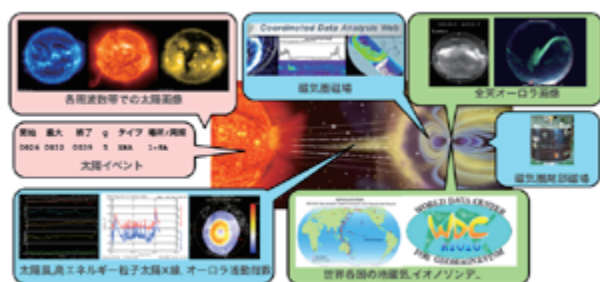


図 1: オーロラに関連するデータの例

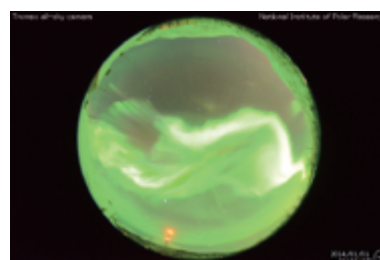


図 2: 全天オーロラ画像

タ指向型科学的手法において最も重視すべき点は, 大量のデータを用意する事である. 特に, いつ, どのような規模のオーロラが発生したかという情報を他の観測データと紐づけて, 訓練データとして用意する必要がある. しかし, 現在そのようなデータは存在していないため, オーロラの全天観測画像をオーロラの有無, 規模, 雲の有無によって分類する必要がある.

そのために, 本稿では一般的な画像処理技術をオーロラ画像の分類に適用し, その精度を評価する. 評価のために, ノルウェー・トロムソにおいて国立極地研究所によって撮影されたオーロラの画像 (図 2 参照) を用いる. これらの画像は国立極地研究所のホームページ [13] 上で公開されている. この画像は, トロムソの観測所内の地上に設置したカメラに魚眼レンズを装着する事によって, 一枚の画像で全を撮影しており, 今回の目的に好適である. 本研究では, この画像データに対して, 局所特徴量を用いたベクトルの類似度による分類と色のヒストグラムを用いた手法, 及び, HSV カラーモデルの閾値を満たす画素数によるオーロラの面積抽出の手法を適用し, それぞれの結果を評価する. これらの手法は, 順番に, 形状や色などの画像特徴をより細く分類できる手法で, 局所特徴量を用いた手法は画像検索に用いられる. ヒストグラムは色の分布に, オーロラの面積抽出はオーロラに対応する色のピクセル数にのみ着目している.

本稿では, 全天オーロラ画像をオーロラの有無・規模・雲の有無によって分類する手法の結果と考察について第 2 節で述べ, 今後の展望について第 3 節で述べる.

2 全天オーロラ画像の画像分類

本節では、ヒストグラム類似度と局所特徴量を用いて全天オーロラ画像を、巨大なオーロラが出現している画像、薄いオーロラが出現している画像、雲がオーロラを遮っている画像、悪天候でオーロラが確認できない画像/オーロラが確認できない画像の4つのカテゴリに分類することで、訓練データの生成に利用できない画像データを取り除き、オーロラが巨大になった時間のデータを収集することを目標とする。また、オーロラの面積を計算することによって全天オーロラ画像を「オーロラ有」と「オーロラ無」の二種類に分類する手法と専門家による評価結果について述べる。

2.1 ヒストグラムを利用したカテゴリ分類

撮影された全天オーロラ画像には、オーロラがはっきりと写っている画像や、薄いオーロラが出現している画像、オーロラが写っていない画像、悪天候によりオーロラが確認できない画像などが混在している。図3は、オーロラがはっきりと写っている画像と薄いオーロラが出現している画像の例である。そして、図4の右のグラフは、画素の持つ青、緑、赤の輝度値(0-255)を横軸に、出現数を縦軸とした輝度値ヒストグラムであり、異なっている両画像間のヒストグラムに差が生じている事がわかる。本節では輝度値ヒストグラムを利用して、オーロラの規模、雲の有無によるオーロラ有画像のカテゴリ分類を行う。以後、全天オーロラ画像をカテゴリへ分類するための指標となる画像を分類基準画像、分類する対象となる画像を未分類画像と呼ぶ。本手法ではオーロラの規模、雲の有無による4つの分類基準画像を各6例ずつ用意し、未分類画像を4

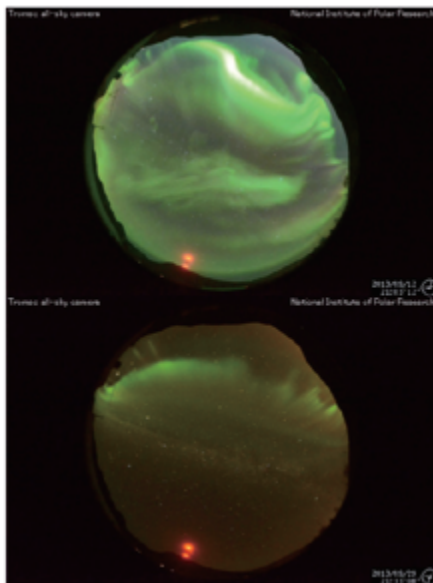


図 3: オーロラの規模の異なり

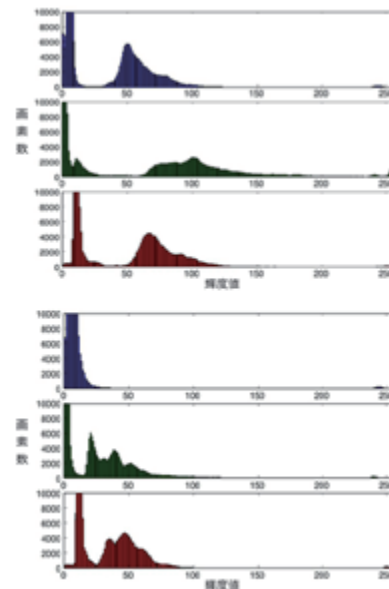


図 4: 出現規模によるヒストグラムの異なり

つのカテゴリ(図5)へ分類する。図5は、分類する4つのカテゴリ、[第1行]: 巨大なオーロラが出現している画像、[第2行]: 薄いオーロラが出現している画像、[第3行]: 雲がオーロラを遮っている画像、[第4行]: オーロラが確認できない画像の分類基準画像となっている。未分類画像のヒストグラムを算出し、分類基準画像のヒストグラムとの類似度を0-1範囲の数値で算出する。その後、最も類似度の高い分類基準画像が所属するカテゴリへ未分類画像を分類するプログラムを作成した。類似度の算出にはOpenCVに実装されて

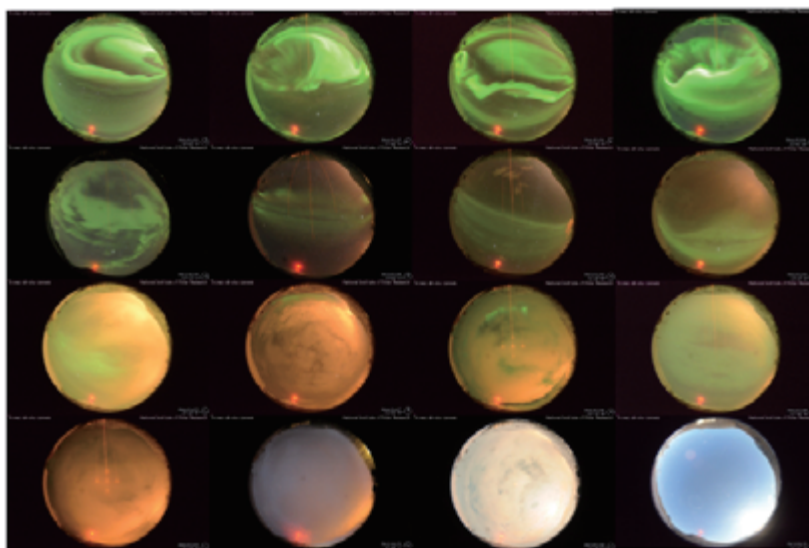


図 5: 分類する 4 つのカテゴリとその基準画像 :

[第 1 行]: 巨大なオーロラが出現している画像, [第 2 行]: 薄いオーロラが出現している画像,
[第 3 行]: 雲がオーロラを遮っている画像, [第 4 行]: オーロラが確認できない画像

いるヒストグラム相関を利用して、ランダムに選択した画像 392 枚の分類を行った。図 5 の分類基準画像に類似していると判断された、上位 5 枚の画像を図 6 に示す。

分類結果を目視で確認したところ、どのカテゴリにも雲がオーロラを遮っている画像が上位に位置する結果となった。これは、雲の持つ輝度値が高いため、画像全体の輝度値が高くなり、オーロラの出現している画像のヒストグラムに類似するためであると考えられる。分類基準画像を増加させた場合でも同様の結果が得られたため、本研究で実施したヒストグラム類似度による分類では、オーロラ全天観測画像の分類は困難であることがわかった。

2.2 局所特徴量を用いた画像分類

より詳細な画像分類を実現する為に、局所特徴量を利用した画像分類を行う。この手法は現在一般画像認識分野で広く用いられており、近年はこの手法による高精度な画像検索のシステムも登場している。本節ではオーロラの出現している画像、オーロラの出現していない画像のそれぞれから特徴点を抽出し、抽出された特徴点を利用した画像分類を目標とする。

代表的な特徴点抽出手法として、SIFT(Scale Invariant Feature Transform)[11] と SURF(Speeded Up Robust Features)[10] がある。SIFT は SURF と比較して特徴点の抽出時の計算量は増加するが、画像の明暗の変化、微小な形状変化に対して頑健な手法であるため、SIFT を採用した。全天オーロラ画像から抽出した特徴点の位置を図 7 に丸で示す。抽出の結果を見ると、画像上部のオーロラの周囲は特徴点として抽出されているが、画面中央の薄いオーロラについては特徴点として抽出されず、電灯などの人工物や山が特徴点として抽出される結果となった。オーロラは緩やかな濃度勾配を持つため、特徴のある点ではないと判定され、局所特徴量を用いた画像分類に利用できないことがわかった。そのため、今後は緩やかな濃度勾配をもつ対象についても特徴点を抽出できるような手法を検討する必要がある。

2.3 オーロラの面積を利用した画像分類

本節では、画像に写るオーロラの面積を算出することで全天オーロラ画像中のオーロラの有無を判定する(図 8 参照)。本手法では、目視でオーロラが確認できる画像を「オーロラ有り画像」、目視ではオーロラが確

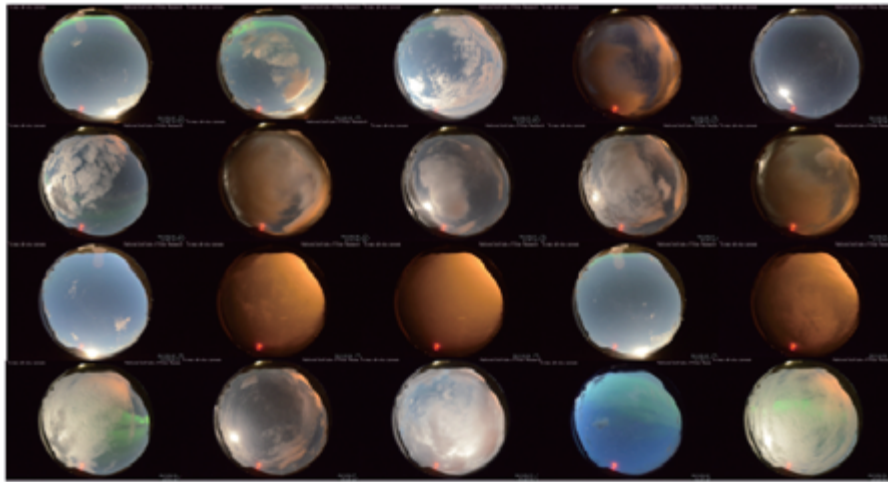


図 6: ヒストグラム類似度による分類結果 :

[第 1 行]:巨大なオーロラが出現している画像, [第 2 行]:薄いオーロラが出現している画像,
[第 3 行]:雲がオーロラを遮っている画像, [第 4 行]:オーロラが確認できない画像

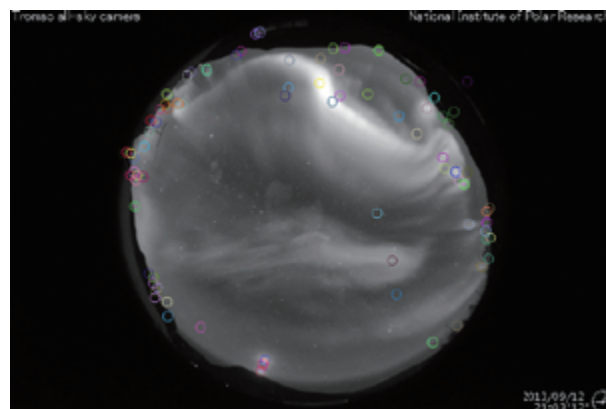


図 7: SIFT による全天オーロラ画像からの特徴点抽出結果

認できない画像を「オーロラ無画像」として、算出された面積をもとに全天オーロラ画像を二種類に分類する事を目標とする。目視で確認できるオーロラは、鮮やかな緑色を伴って出現するため、画像中のオーロラ

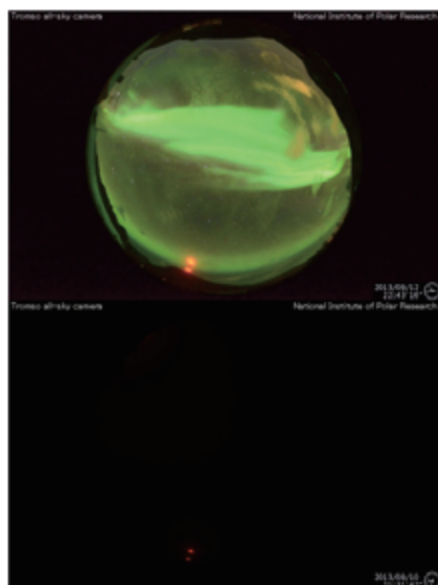


図 8: オーロラ有画像 (上) とオーロラ無画像 (下) の例

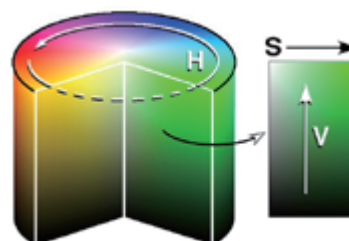


図 9: HSV カラーモデルの視覚的イメージ

の面積を算出する前にオーロラの持つ緑色の範囲を定義した。加法混色である RGB カラーモデルでは緑色の定義が困難であるため、HSV カラーモデル [19](図 9) を利用した。HSV カラーモデルは、色相を表す H 、彩度を表す S 、輝度を表す V の値を定める事で色を定義し、RGB カラーモデルよりも直感的に色を定義する事が出来る。オーロラの面積を算出するために、オーロラの H , S , V の値の範囲を画像処理ソフトウェアによって調査した。全天オーロラ画像について、 H , S , V の全ての値が定めた範囲内である画素をカウントし、カウントされた画素の数をオーロラの面積とする。面積 [pixel] が閾値を超えた場合にオーロラが出現しているものと判定する。 $H = 60 \sim 180$, $S = 25 \sim 97$, $V = 20 \sim 90$ として 2.1 節で利用した画像 392 枚について、オーロラの出現判別を行った。面積が 100[pixel] を超えた場合、オーロラが出現しているものとし、専門家による目視にて評価を行った。出現判別の結果を表 1 に示す。判別結果を見ると、オーロラ的面積による画像分類は単純なアルゴリズムながら、良い結果が得られた。しかし、空が明るく薄いオーロラが出ている全天オーロラ画像や、やや赤みがかったオーロラはオーロラが出現していないと判断されてしまった。

表 1: 識別結果

	正答率	誤識別率
オーロラ有	94.3%	4.7%
オーロラ無	91%	9%

本節では、全天オーロラ画像に写るオーロラの面積を算出した。このオーロラ的面積の変化をグラフ化したものを図 10 に示す。図 10 は、2014/01/01 00:00-23:59 に撮影された、全天オーロラ画像に写るオーロラ的面積を時系列グラフにしたものであり、縦軸がオーロラ的面積 [pixel]、横軸は 00:00 からの経過時間 (秒) を表している。このグラフから 21:30 に急激にオーロラ的面積が変化している事がわかった。この時間の前後の全天オーロラ画像を確認したところ、巨大なオーロラが短時間で出現していた事から、オーロラ的面積の時系列データからオーロラサブストームの開始時刻を発見できる可能性を発見した。今後はこのデータを

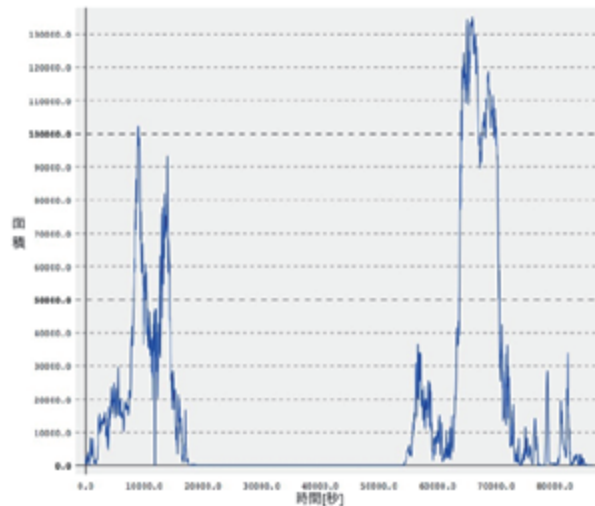


図 10: オーロラ面積の時系列グラフ：100px を超えた場合にオーロラが出現していると判定した。

利用してオーロラサブストームの開始時刻を収集し、オーロラサブストーム開始前の地磁気、太陽風等の時系列データなどを入力、オーロラの面積を出力として学習させる事で、オーロラサブストームの発生時刻の予測に向けた研究を進める。

3 まとめと今後の課題

形状特徴による分類が可能になる局所特徴量を用いた手法と色のヒストグラムを用いた手法は、予備実験の段階で必要な精度がでないことが分かった。オーロラの有無に関しては、実装したプログラムによって全天オーロラ画像から判断することが可能となり、平均正答率が 92.3% という良好な結果が得られた。ただし、薄いオーロラや雲による影響までは分類することができなかった。今後は撮影時の太陽の高さや雲量を考慮し、暗い緑や、明るい緑に分けて H , S , V のパラメータを変化させ、面積の計算を行うことでより細かい画像分類を実現し、10 年間の全天オーロラ画像の分類を行うことが課題である。

この手法を利用することによって、専門家が研究に利用する画像データ収集の際の負担を軽減し、物理学におけるモデル構築等の支援も可能になると考えられる。今後は、オーロラの有無の精度向上に加えて、オーロラの分類、オーロラの出現予測に向けたアプローチを行う予定である。

謝 辞

本研究に際して、画像データを提供して頂いた国立極地研究所の宮岡宏准教授、共同研究者の小川泰信准教授にこの場を借りて深謝いたします。

本研究は JSPS 科研費 25540039 の助成を受けたものです。

参考文献

- [1] 村田健史, “科学研究用クラウドシステム (NICT サイエンスクラウド) の提案”, 宇宙航空研究開発機構研究開発報告書 宇宙科学情報解析論文誌 第 3 号, pp. 39 – 56, (2014).
- [2] 小川泰信, 門倉昭, 元場哲郎, 田中良昌, 細川敬祐, “トロムソ/ロングイアビンにおけるオーロラ観測用並列イメージャーの大量データ処理と可視化”, 宇宙航空研究開発機構研究開発報告書 宇宙科学情報解析論文誌 第 2 号, pp. 51 – 61, (2013).

- [3] Sakaguchi, K., Miyoshi, Y., Saito, S., Nagatsuma, T., Seki, K., Murata, K. and T. “Relativistic electron flux forecast at geostationary orbit using Kalman filter based on multivariate autoregressive model”, Space Weather Volume 11, Issue 2, pp 79 – 89, February, (2013).
- [4] ビクター・マイヤー＝ショーンベルガー, ケネス・クキエ, “ビッグデータの正体”, 講談社, (2013).
- [5] F. Sigernes, M. Dyrland, P. Brekke, “Two methods to forecast auroral displays”, Space Weather Space Clim vol.1, (2011).
- [6] 岩崎雅二郎, “近似 k 最近傍グラフによる距離空間の近傍検索”, 情報処理論文学会誌 Vol. 3, pp. 91 - 110, (2010).
- [7] Tokunaga, T., Ikeda, D., Nakamura, K., Higuchi, T., Yoshikawa, A., Uozumi, T., Fujimoto, A., Morioka, A. and Yumoto, K. “Detecting Precursory Events in Time Series Data by Extension of Singular Spectrum Transformation”, Proceedings of the 10th WSEAS International Conference on Applied Computer Science, pp.366 – 374, (2010).
- [8] 篠原学, “宇宙天気”, 誠文堂新光社, (2009).
- [9] Tony Hey, Stewart Tansley, and Kristin Tolle, “The Fourth Paradigm: Data-Intensive Scientific Discovery”, Microsoft Research, (2009).
- [10] Herbert Bay, Tinne Tuytelaars, Luc Van Gool, “SURF: Speeded Up Robust Features”, Lecture Notes in Computer Science Volume 3951, pp.404–417 (2006).
- [11] David G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, International journal of computer vision ,60 (2), pp. 91– 110 (2004).
- [12] H. Gleisner, “Response of the auroral electrojets to the solar wind modeled with newral networks”, Journal of Geophysical Research: Space Physics . Vol.102(A7), pp. 14269 – 14278, (1997).
- [13] 国立極地研究所 “<http://polaris.nipr.ac.jp/acaaurora/aurora/Tromso/>”
- [14] NASA SWPC ACE “http://www.swpc.noaa.gov/ace/ace_rtsw_data.html”
- [15] Alaska University Aurora Forecast “<http://www.gi.alaska.edu/AuroraForecast>”
- [16] Bonanza-The Computer Shogi Program “http://www.geocities.jp/bonanza_shogi/”
- [17] Merck Molecular Activity Challenge “<https://www.kaggle.com/c/MerckActivity>”
- [18] Large Scale Visual Recognition Challenge 2012
“<http://www.image-net.org/challenges/LSVRC/2012/>”
- [19] 画像出典: “<http://ja.wikipedia.org/wiki/HSV>”