

衛星画像データ解析システムにおける 地理情報の解析と視覚化

岩崎 信也^{†1} 森 駿^{†2} 遠藤 省吾^{†1}
山口 崇志^{†1} 花田 真樹^{†1} 朴 鍾杰^{†1} 布広 永示^{†1}

衛星画像解析システムでは数値情報解析による研究成果を衛星画像として Web を介し配信している。現在、地理情報の収集を逐次的に行い可視化・分析する Web GIS 機能を開発している。本稿では Web GIS でのリクエストごとに地理情報の視覚化及びデータ提供を行うオンデマンド衛星画像データ機能における大規模データ処理について議論する。

Analysis and Visualization of Geographic Information in Satellite Image Data Analysis System

SHINYA IWASAKI^{†1} HAYAO MORI^{†2}
SHOGO ENDO^{†1} TAKASHI YAMAGUCHI^{†1} MASAKI HANADA^{†1}
JONG GEOL PARK^{†1} EIJI NUNOHIRO^{†1}

Satellite Image Data Analyze System broadcast the satellite image of visualized geographical information and sensor data via web site. In this research, we develop the Web GIS functions consist of visualization and analysis for the acquisitions geographical information. In this paper, a distributed computation system was investigated in order to handle big data of on-demand satellite images for each user request.

1. はじめに

東京情報大学では文科省の戦略的研究基盤形成支援採択事業として、NASA の地球観測衛星 Terra および Aqua に搭載された中分解能撮像分光放射計 (MODIS: Moderate Resolution Imaging Spectroradiometer) センサで受信されるデータ (衛星画像データ) を利用した「環境変動」の解析に関する研究を進めている [1][2][3][4][5][6][7]。その一環として現在受信した衛星画像データと解析結果の Web 配信を行う衛星画像データ解析システム (SIDAS: Satellite Image Data Analysis System) を開発している。本報告では Web GIS の開発を主軸とし、大規模な衛星データの地理情報の視覚化及びデータ提供を行う際のデータストレージ・大規模データ処理について衛星データの特徴を考慮しながら検討を行った。

2. SIDAS (衛星画像解析システム)

2.1 システム概要

SIDAS は、戦略的研究基盤形成に関わる研究の一環として開発された衛星画像データの解析処理を行う衛星画像データ解析システムである。現在公開している Ver1.2 は東日

本大震災をきっかけに日本の東北地方を範囲とした衛星データを逐次解析し、研究に利用しやすいフォーマットにデータ変換すると共に RGB 画像に加え、植生指数、地表面温度、海表面温度の可視化画像を Web 上で毎日配信している。現在研究開発中である Ver2.0 以降では、解析範囲を東アジア一帯に拡大すると共に Web 上に構築する地理情報システム (Web GIS) の導入、ユーザのリクエストごとの地理情報の視覚化・データ提供を行うオンデマンド型衛星画像提供の 2 点を中心とした拡張を行なっている。その際、衛星データが大規模になることから衛星データの特徴を踏まえながらデータ形式の検討及びストレージシステム・データ処理システムの設計を行った。

2.2 衛星画像データ

SIDAS では MODIS センサで受信したデータ及び加工されたデータを全て衛星画像データとして取り扱う。

MODIS センサデータは NASA の地球観測衛星 Terra および Aqua に搭載された MODIS センサで受信され 1 日 2 回地球を周回する。MODIS センサの観測幅は 2330km で 36 バンド (可視域から熱赤外域まで: 0.405~14.385 μ m) と 3 つの空間分解能 (250m, 500m, 1000m) を持っている [8]。

東京情報大学では沖縄、北海道、千葉の三箇所に受信局を設置し、東経 80 度から 180 度まで、北緯 0 度から 80 度の範囲を受信しており SIDAS では MOD02 の一部、MOD11、MOD13、MOD14、MOD28 の 5 種類のデータを解析する。

^{†1} 東京情報大学総合情報学部
Tokyo University of Information Sciences

^{†2} 東京情報大学大学院総合情報学研究所
Graduate School of Tokyo University of Information Sciences

2.2.1 衛星画像データの種類(MODIS データ)

SIDAS では、5 種類のデータを解析している。各データの詳細を表 1 に示す。

表 1 衛星画像データ詳細

データ名称	概要	バンド数	空間解像度
MOD02	反射率データ	7	500m
MOD11	地表面温度	1	1km
MOD13	植生指数	2	500m
MOD14	異常温度	1	1km
MOD28	海面温度	1	1km

バンドはセンサの反射帯域ごとに存在している。MOD02 には 36 バンドのデータがあるが RGB 画像の視覚化に必要な 3 バンドを含む利用度の高い 7 バンドのみを解析している。

2.2.2 衛星画像データのフォーマット

SIDAS ではデータフォーマットとして HDF-EOS (Hierarchical Data Format-Earth Observing System) データ、RAW データ、PNG データの三種類を利用している。下記に 3 種類のそれぞれの特徴を記述する。

1) HDF-EOS データ

HDF-EOS は受信した直後のデータフォーマットであり、複数のデータを階層的に格納できるため衛星データに利用しやすい形式である。特徴として位置・時間・撮影データが階層に分かれて保存されており 1 データ毎に位置と時間が取得できる反面、データ量が非常に多くなる。

2) RAW データ

SIDAS では研究における利便性やデータ量の観点からバンド毎に RAW データに変換して保存・提供を行なっている。RAW データは一般的な画像フォーマットに近いデータ形式であるがデータ部しか存在せず、データ部内は撮影データが等間隔・直列に並んでいる。位置情報や時間等の付加情報は別ファイル(ヘッダファイル)に保存しておく必要があるが、データ量は HDF-EOS に比べ少ない。

3) PNG データ

SIDAS では Web 上で衛星画像を公開する際、PNG データとして視覚化を行なっている。また Ver2.0 以降では公開する範囲が東アジア全域と多くなりデータ量が巨大になることからピース化と呼ばれる画像を解像度で細分化する操作を行い衛星画像を保存している。

2.2.3 SIDAS が提供する衛星画像データ

SIDAS が提供する衛星画像データについては、以下のとおりとなる。

・データの種類

データの種類については MOD02, MOD11, MOD13, MOD14, MOD28 の 5 種類であるが内部のバンドを分解し 11 種類のデータとして扱う。各データの認識にはデータ ID を使う。データ ID はそれぞれ MOD02 がバンド順に 001~007, MOD11 が 101, MOD13 が 102, 103, MOD14 が 104, MOD28 が 105 となる。

・データサイズ

データサイズに関しては、東経 80 度から 180 度まで、北緯 0 度から 80 度の衛星画像データとなるが、経度緯度 1 度辺り 100km の等間隔投影影であり、解像度は 500m で統一する。また 1 解像度あたり 4Byte のデータを持ちフォーマットは RAW データとなる。そのためストレージに保存される衛星画像データは 1 ファイルあたり 640MByte となる。

衛星画像データは 1 日に 2 機の衛星がそれぞれ昼・夜の 2 種類とするため 1 データ ID ごとに 1 日 4 つの衛星画像データファイルが保存される。

2.2.4 衛星画像データの受信

東京情報大学では沖縄、北海道、千葉の 3 箇所に受信所を設け 1 日 2 回周回する Terra 及び Aqua の衛星画像データを受信しているが、その際の特徴として下記の 2 点が上げられる。

- 1) Terra 及び Aqua は 14 日周期で同じ軌道を迎えるため日にち毎に撮影範囲が異なる。
- 2) 待機等の受信状態により受信が複数回にわたり、衛星 1 周回のデータファイルが複数になる場合がある。

2.3 衛星画像データ利用法

SIDAS の目的の一つとして、衛星画像データを利用した環境の変化や地表面の変化などに関する研究を支援することがある。Ver1.2 ではメール等によって連絡を受けた場合のみ衛星データの提供を行なっていたが Ver2.0 以降では Web 上でリクエストに応じてデータ提供を行う。これまでに SIDAS にて提供された衛星画像データの研究の一部を下記に記す

- ① 東アジアの土地被覆変化抽出の傾向分析[3]
2001 年から現在まで東アジアの土地被覆変化を抽出し、その傾向を分析する。
- ② 東アジアの災害図(火災、洪水)の作成[6]
洪水や火災など自然災害の自動抽出を行い東アジアの災害図を作成する。
- ③ 自己回帰分析を用いた NDVI の時系列予測[5]
衛星画像データを元に自己回帰分析を用い定期化植生指数の変化を予測する。
- ④ 知能情報処理を用いた土地被覆領域の境界抽出[2]
ニューラルネットワーク、自己組織化マップなどの知能情報処理手法を用いた土地被覆分類を行うと共に精

度の向上を目指す。

2.4 システム構成

SIDAS の構成を図 1 に示す。解析システムは Web サーバ、データベースサーバ、アプリケーションサーバ群(AP サーバ群)及び画像ストレージとデータストレージから構成される。データ受信サーバは SIDAS 内部に含まず独立している。

AP サーバ群については、単体処理用の AP サーバと並列分散用の PC クラスタから構成される。事前処理といった定期処理は AP サーバ単体で実行されるが、オンデマンド型衛星画像データ配信などの処理は PC クラスタで高速実行が行われる。現状 AP として提供している処理はモザイク処理・コンポジット処理・ピース化処理であるが、林野火災探索・類似画像探索・土地被覆変化解析や気象変動解析などの衛星画像解析アプリケーションも研究開発中である。

新規衛星画像データの受信後、事前処理を行いデータストレージに格納すると共に DB サーバ内にヘッダ情報を格納、視覚化画像を画像ストレージに格納している。また、ユーザは Web 上で視覚化された地図情報を Web GIS によって閲覧できると共に、リクエストを出すことで自分の意図したデータをデータストレージから抜き出し計算ノードによって AP の実行、結果を Web 上に表示できる。

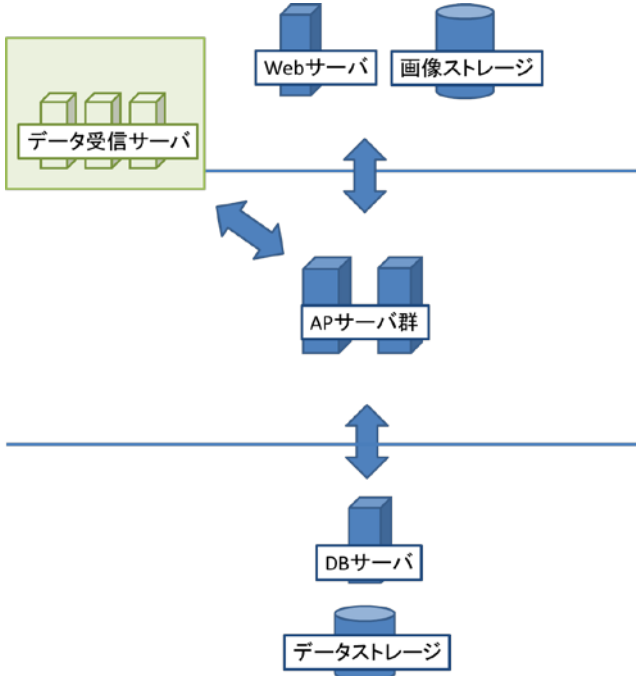


図 1 SIDAS システム構成図

2.4.1 事前処理

SIDAS では AP サーバにおいて受信した衛星画像データを保存するために事前処理を行なっている。受信した衛星画像データは HDF-EOS フォーマットであることから RAW

フォーマットに変換すると共にモザイク処理・コンポジット処理・視覚化処理を図 2 の手順で行なっている。

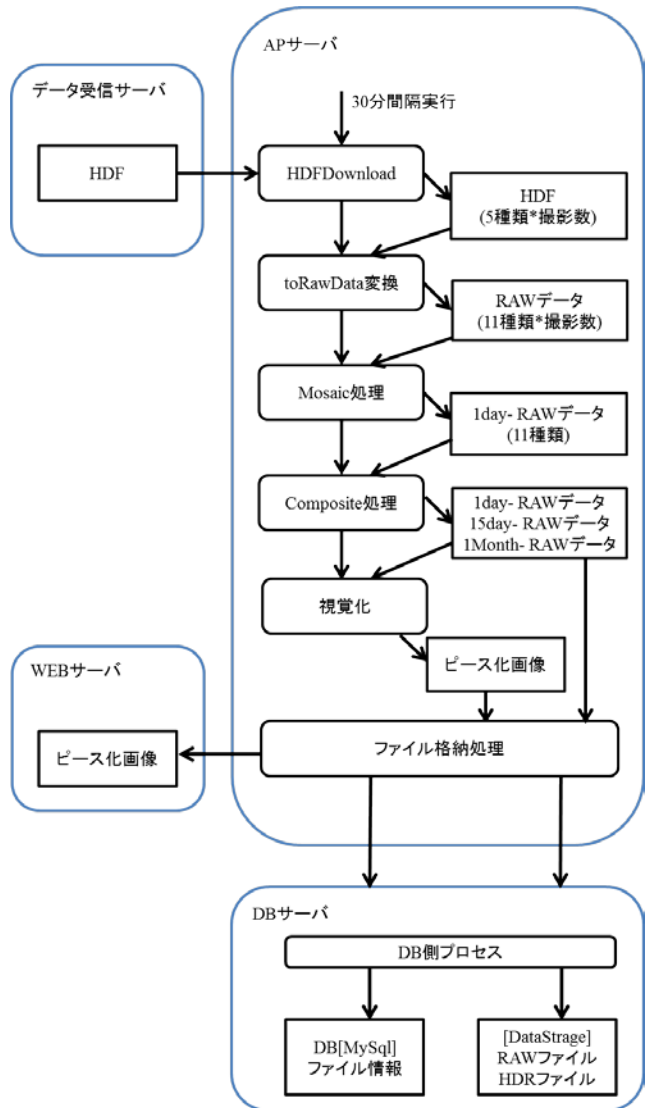


図 2 事前処理手順,データフロー

HDFDownload 処理では 30 分毎にデータ受信サーバから新規衛星画像データのダウンロードを行う。

toRawData 処理では受信した新規衛星画像データの RAW フォーマットへの変換を行うと共にバンド毎にデータを分解する。

モザイク処理では変換した RAW データを衛星周回単位で 1 つのデータに図 3 のように纏める。この際、3 つの受信所のデータを纏めることになるが、同一地点は同一データであるためデータの欠損は起こらない。

コンポジット処理では衛星画像データの視認性向上及び雲・帯域の削減のため 15 日及び 1 ヶ月をコンポジットした 15day-RAW データ及び 1Month-RAW データを作成する。

視覚化処理ではモザイク及びコンポジット処理で作成した RAW データの PNG 画像への変換を行うと共に、Web GIS で利用するために解像度の縮小及びピース化を行う。

ピース化した画像に関しては表2のようになる。

表2 ピース化解像度詳細

解像度 LV	画像サイズ[px]		解像度	ファイル数
	横	縦		
1	640	512	16000	4
2	1280	1024	8000	16
3	2560	2048	4000	64
4	5120	4096	2000	256
5	10240	8192	1000	1024
6	20480	16384	500	4096
			計	5460

LV6 が解像度 500m の RAW データをそのまま画像化したものとなり、Web GIS で利用するため解像度を縮小した LV1~5 を作成する。

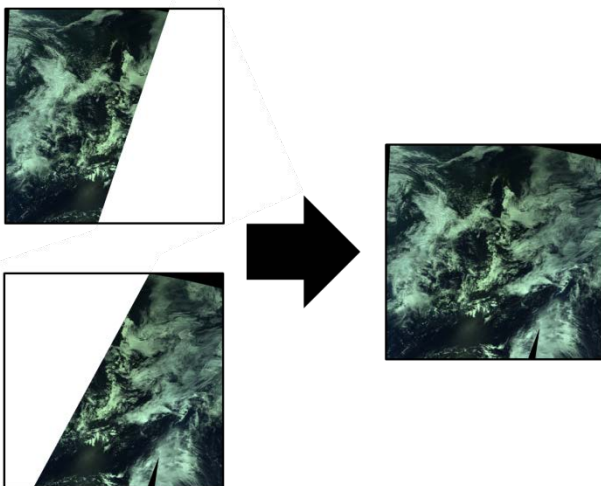


図3 モザイクの様子

3. Web GIS

SIDAS では数理情報解析による研究成果の配信と地理情報の収集を逐次的に行い、これらを可視化・分析する機能を持つ Web GIS(図4)を開発しており、ユーザ間で循環的に知識を集約し環境変動の観点や災害対策に利用可能なプラットフォーム構築を行なっている。Web GIS は下記の機能を持つ。

- 1) GoogleMap に代表されるズームを行うごとに解像度を大きくするリッチユーザインターフェースを搭載した地図アプリケーション機能
- 2) 衛星画像に対するメタ情報の付加・検索機能
- 3) 地図情報のレイヤ表示・編集
- 4) 時系列地図表示
- 5) オンデマンド型衛星画像配信

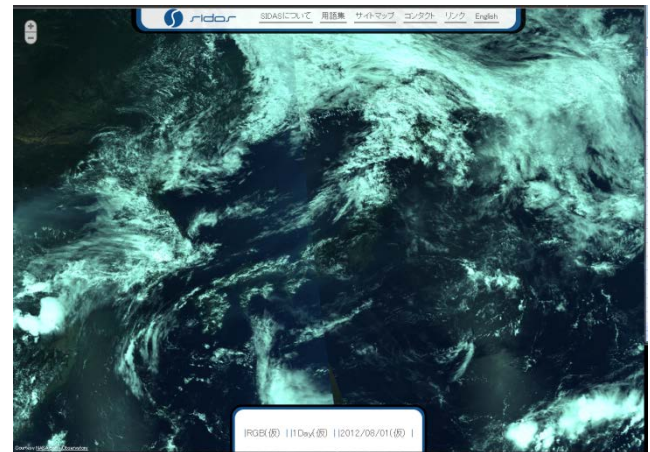


図4 Web GIS のインターフェース

3.1 オンデマンド型衛星画像データ配信

Web GIS ではオンデマンド型衛星画像データ配信により事前に用意していない画像でもユーザのリクエストに応じて衛星画像データの加工を行い提供している。その際、リクエスト内容によっては衛星画像データが大規模になることから通信トラフィックを考慮しながら DB システム・データストレージ及び AP システムの設計を行った。リクエストからのデータの配信までの手順が図5となる。

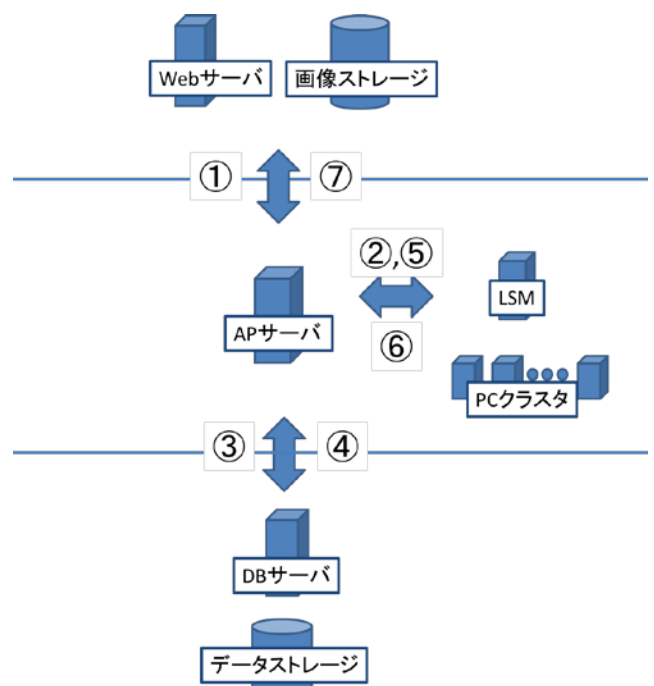


図5 リクエストからレスポンスまでの手順

LSM(Logical Storage Manager)はデータの再利用機能・スケジューリングなど PC クラスターの管理を行う。リクエストからレスポンスまでの手順は以下の通りである。①ではリクエスト情報を AP サーバに通知 ②ではリクエスト情報の必要データ情報を LSM に通達し

データキャッシュのデータと参照する。

③では DB サーバに RAW データ及び範囲を到達しデータを請求する。

④では DB サーバが請求された RAW データにトリミングを行い請求 RAW データとして AP サーバに送信を行う。

⑤では LSM に RAW データを送ると共にリクエストされた処理を担当する PC クラスタを設定しデータ処理を行う。

⑥では処理の実行結果 RAW データを AP サーバに送信する

⑦では AP サーバに送られたデータをレスポンスデータとして Web サーバに通知し Web サーバは Web GIS 上への表示もしくはデータダウンロードを行う。

3.1.1 DB サーバによるデータトリミング

DB サーバは AP サーバよりデータ名と範囲を受け取りデータの選択・トリミングを行う。DB サーバでデータトリミングを行うのは DB サーバ・AP サーバ間のトラフィック量を減らすためである。

3.1.2 PC クラスタのデータ再利用機能

PC クラスタのデータの再利用機能は PC クラスタが保持し、複数の衛星画像データを処理に利用した際、新しい衛星画像データを PC クラスタ内に保持しておく。

この手法により利用度が高い衛星画像データは PC クラスタ内にあるため、データストレージから PC クラスタにデータを転送する必要がなくなりトラフィックの削減につながる。

3.1.3 データの結合

DB サーバでデータトリミングを行い PC クラスタのデータ再利用機能を利用する際、完全一致判定を行うとトリミング範囲が多少違うだけで一致が起きなくなる問題が発生する。この問題を防ぐため、PC クラスタのデータに無い範囲だけを DB サーバに請求しクラスタにて結合を行う。

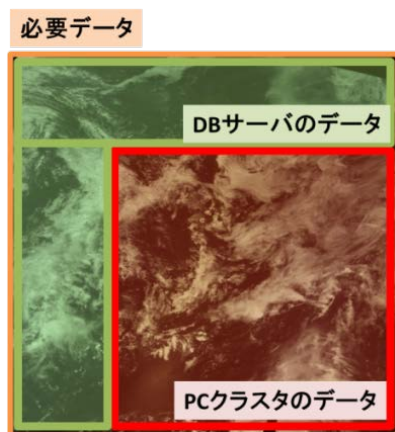


図 6 データの結合例

3.1.4 リクエストに対するデータ転送量

リクエストのデータ転送量 TB は下記の式 1.1 となる。

$$TB = TDC + TCW \quad (式 1.1)$$

TDC は DB サーバから PC クラスタへのデータ転送量であり、TCW はリクエスト AP 処理終了後の PC クラスタから Web サーバへのデータ転送量である。TDC, TCW は式 1.2, 式 1.3 となる。

$$TDC = (1 - h) \cdot (ID \cdot \frac{A'}{A}) \quad (式 1.2)$$

$$TCW = IO \quad (式 1.3)$$

式 1.2 の h はクラスタの再利用データのヒット率であり、ID は必要ファイルの総データ量、A は必要ファイルの総面積であり、A' はファイルの必要面積である。

式 1.3 の IO はアプリケーション終了後のレスポンスデータ量である。

3.1.5 リクエストに対するサーバ処理時間

リクエストのサーバ処理時間は式 1.4 となる。

$$TA = CST + DTT + DCT + AT + TI \quad (式 1.4)$$

TA はリクエストのサーバ処理時間である。CST は PC クラスタ内のデータサーチ時間であり、限りなく 0 に近い。DTT はデータトリミング時間、DCT はデータ結合時間であり AT はアプリケーション処理時間である。TI はデータ転送時間であり、式 1.5 で計算される。

$$TI = \frac{TB}{S} \quad (式 1.5)$$

TB は式 1.1 で求まるデータ転送量 TB であり、S はデータ転送速度である。

3.1.6 サーバ処理時間の検証

PC クラスタのデータ再利用機能と DB サーバによるデータトリミングについての検証を行った。

リクエスト内容

処理内容	コンポジット処理 (15Day)
期間	1ヶ月(2012年11月)
範囲	経度 120~180 緯度 50~80
データ ID	1,2,3,4,5,6,7,101 計 8 種類

前提条件

PC クラスタ数	32 台
再利用データ重複	なし
データサーチ時間	1 秒
データトリミング時間	1 データ 0.5 秒
AP 処理時間	1 コンポジット 5 秒
データ転送速度	1Gbps

以上の条件のもと、PC クラスタのデータ再利用機能及び DB によるデータトリミング機能についてサーバ処理時間との関係が図7となる。

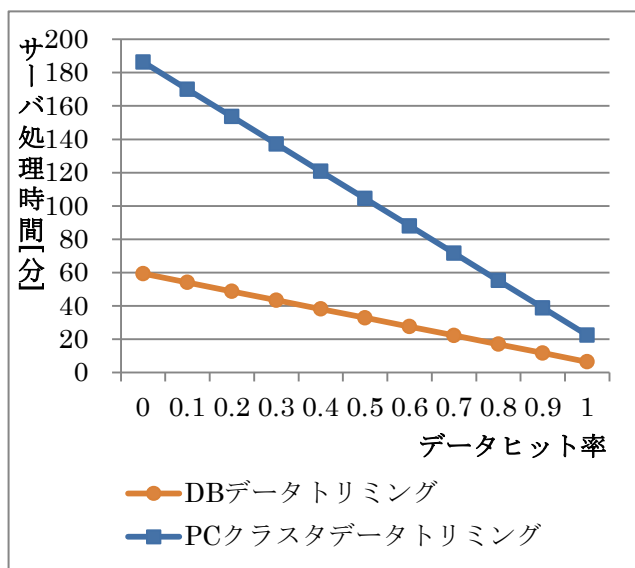


図7 サーバ処理時間とヒット率・トリミングの関係

図7からPCクラスタのデータ再利用ヒット率がデータ処理時間に強い影響を与えることが分かる。またDBサーバでのデータトリミングはPCクラスタでのデータトリミングに比べPCクラスタのデータ再利用ヒット率が低いほど効果があることが分かる。これはPCクラスタでデータトリミングを行うと、DBサーバからPCクラスタまでデータ利用のたびに1ファイルを全て転送する必要があるためである。

4. おわりに

本報告では、Web GISの開発を主軸とし大規模な衛星データの地理情報の視覚化及びデータ提供を行う際のデータストレージ・大規模データ処理について衛星データの特徴を考慮しながら検討を行った。

オンデマンド衛星画像データ配信に対して、サーバ処理時間の削減のためPCクラスタの構築を行うと共に、DBサーバでのデータトリミング機能、PCクラスタのデータ再利用機能を実装・検証を行った。

今後は、以下の2点を中心にWeb GISの拡張及びオンデマンド衛星画像データ配信におけるリクエスト処理の高速化の検討を行う。

(1) 提供アプリケーションの増加

林野火災探索・類似画像探索・土地被覆変化解析や気象変動解析などの衛星画像解析アプリケーションの提供

(2) 大規模データ分散システムの検討

大規模データ分散フレームワーク hadoop による分散データベースの検討

謝辞

本研究は、私立大学戦略的研究基盤形成支援採択事業東京情報大学 国際環境研究プロジェクト「アジア東岸域の環境圏と経済・社会圏の持続的発展のための総合研究」の補助を受けた。

参考文献

- 1) Nakamura, A et al.: Development and Evaluation of Satellite Image Data Analysis Infrastructure, Journal of Artificial Life and Robotics (2012).
- 2) Mori, K et al.: Application of neural network swarm optimization for paddy-field classification from remote sensing data, Journal of Artificial Life and Robotics, vol.16 no. 4 pp.497-501(2012)
- 3) Park, J et al.: Determination of inundation area based on flood hazard for a global water risk assessment, Risk in water Resources Management IAHS 2011, pp.61 - 64.
- 4) Nunohiro, E et al.: Development of satellite data analysis system, Proceedings of the 5th International Conference on Information, pp.143-146(2009)
- 5) Igarashi, M et al.: A method to fit a Nonlinear Curve to NDVI, SST and LST, International Journal of Geology, vol. 3 no.4 pp97-100.
- 6) Nunohiro, E et al.: Forest and Field Fire Search System using MODIS Data, Journal of Advanced Computational Intelligence and Intelligent Informatics, vol.11 no.8 pp1043-1048 (2007)
- 7) Kenneth J.M et al.: Land and Surface Cover Classification by Soft Computing Methods using MODIS Satellite Data, Journal of International Information Institute, INFORMATION vol.13 no.3 pp.1013-1018(2010).
- 8) NASA.: National Aeronautics and Space Administration MODIS Web , <http://modis.gsfc.nasa.gov/>