

大規模災害時の要救助者検知のための ストリーム G 空間データ処理技術

小川祐一^{†1,a)} 菅谷奈津子^{†1,b)} 浅原彰規^{1,†1,c)} 林秀樹^{1,†1,d)} 富田仁志^{†1,d)}

大規模災害時には、現場の救助活動にあたる救助隊員にとって刻一刻と変化する被災状況をいち早く把握することは重要である。本研究では、スマートフォンなど移動しながらデータを送信するセンサや、雨量計などの特定の観測地点のデータを送信するセンサなど逐次的に収集される大規模なストリームデータに対して、2つのストリームデータの間の距離が近くなったときに、その事実を外部システムに通知するためのリアルタイム処理を可能とする技術を開発した。首都直下型地震で火災や倒壊の危険性が指摘されている人口100万人の地域を想定して検証した結果、100万台のセンサが5分に1回、位置情報を送信する環境において単一サーバでリアルタイムに処理できることを確認した。

1. はじめに

近年、世界的規模で自然災害が頻発している。また、日本においては、地理的、地形的、気象的諸条件から、台風、豪雨、豪雪等の自然災害が発生しやすい国土となっており¹⁾、今後南関東地域でM7クラスの地震が発生する確率は30年間で70パーセントと推定されている²⁾。首都直下地震が発生すると、建物倒壊や市街地延焼火災により、人的・物的被害が甚大となることが報告されており³⁾、このような大規模災害から人命や財産を保護することは、国の最重要課題である。

大規模災害時には、現場の救助活動にあたる救助隊員にとって、刻一刻と変化する被災状況をいち早く把握することは重要である。たとえば、現場の救助活動にあたる救助隊員に対しては、現場の火災状況、現場付近に滞在する人の分布や動きがリアルタイムに把握することができれば、救助の優先度を意思決定するための支援となる。さらに、現場付近に滞在する人が、火災に巻き込まれている事実や周囲の人から孤立している事実までリアルタイムに把握することができれば、救助活動への有効性は高いと考える。

現場付近に滞在する人が、火災に巻き込まれていないか、または周囲の人から孤立していないかをリアルタイムに外部システムへ通知するためには、2つの移動体データ(以下、ストリーム G 空間データと呼ぶ)の間の距離が近くなったこと調べる、すなわち定期的に2つのストリーム G 空間データ間の距離を調べることが必要となる。しかし、時間とともに移動するストリーム G 空間データの件数が多い場合、処理量が極端に多くなってしまいう問題がある。この処理では、図1に示すように、あるストリーム G 空間データに注目して(図中「注目移動 G 空間センサ位置」)周囲の他のストリーム G 空間データとの距離を計算して

距離の小さい組み合わせを検出し、次に別のストリーム G 空間データに注目して同様に距離を計算し、という手続きを繰り返し行うことになる(以下、この手法を従来手法と呼ぶ)。この従来手法は、距離を求めるべき対象が増えると急激に増加する。もちろん、十分多数のコンピュータを準備し、それらの間で分担して処理する方法も考えられるが、この膨大な処理量に見合う台数のコンピュータを準備することは現実的ではない。

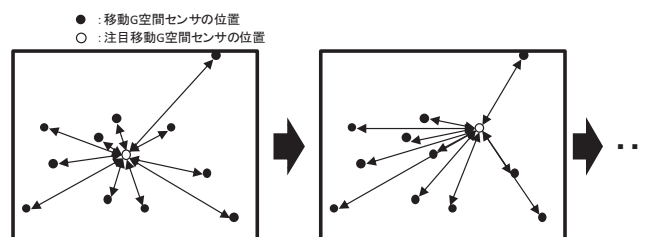


図1 従来手法による複数ストリーム G 空間データに対する相関の判定手順

そこで本研究では、現場付近に滞在する人の危険性を現場の救助隊員がいち早く特定できることを目的として、時間とともに移動するストリーム G 空間データ同士であっても、ストリーム G 空間データの位置に変化があるたびに行なわれる距離計算を高速に行ない、リアルタイムに結果を出力する技術を提案する。提案手法に求められる性能要件として、首都直下型地震で火災や倒壊の危険性が指摘されている⁴⁾人口100万人・人口密度10,000人/km²の地域を想定して、100万台のセンサが5分に1回(3,000件/秒)位置情報を送信する状況において、単一サーバでリアルタイムに処理できることを目標とする。

以下では、2章では関連研究について述べ、本研究の位置付けを明確にする。3章では提案手法および提案手法に基づいて、周囲の人から孤立している人がいることをリアルタイムに検知するシステムの実装方法について説明する。4章では性能評価の結果を示す。最後に5章で本論文のまとめを述べる。

1 情報処理学会
 IPSJ, Chiyoda, Tokyo, 101-0062, Japan.
 †1 現在、株式会社日立製作所
 Presently with Hitachi, Ltd.
 a) yuichi.ogawa.nf@hitachi.com
 b) natsuko.sugaya.tr@hitachi.com
 c) akinori.asahara.bq@hitachi.com
 d) hideki.hayashi.xu@hitachi.com
 e) hitoshi.tomita.jg@hitachi.com

2. 関連研究

ストリーム G 空間データ間の距離計算を伴う本研究に関連した関連研究⁵⁾として、ストリーム G 空間データが継続的に発生する一方で、ユーザの問合せも継続的に登録される環境において、リアルタイムにユーザの問合せに対する結果を返す時空間データストリーム技術が提案されている。本技術により、例えば、ある地点を中心とする半径 10 メートル以内に存在する移動体を返す問合せ（範囲検索）やある地点から距離の近い上位 3 件の移動体を返す問合せ（k 最近傍検索）を実現できる。しかし、本研究の課題のように複数の移動体同士の距離が閾値以下になることを検出するのは領域が変化する場合に相当するため、そのままでは適用困難である。

以上により、2つのストリーム G 空間データの間の距離が近くなったときに、その事実を外部システムに通知するためには、定期的に2つのストリーム G 空間データ間の距離を調べることが必要となる。しかし関連研究および従来手法では、時間とともに移動する2つのストリーム G 空間データの件数がそれぞれ多い場合、処理量が極端に多くなってしまいう問題がある。そこで、本研究では、時間とともに移動するストリーム G 空間データ同士であっても、ストリーム G 空間データの位置に変化があるたびに行なわれる距離計算を高速に行ない、リアルタイムに結果を出力していくことを課題とする。

3. 提案手法

以下、大規模なストリーム G 空間データ群に対して、それらの間の距離が閾値以下であることを高速に判定する手法について詳細を述べる。

3.1 基本的な考え方

大規模なストリーム G 空間データ群に対してそれらの間の距離が閾値以下であることを高速に判定する手法として、空間を格子状の領域に分割して管理する方法を適用する。提案手法のアプローチについて図 2 を用いて説明する。まず、提案手法では、事前に空間を格子状の区画に分割し、それぞれに識別番号を割り振っておく。G 空間データベースサーバはストリーム G 空間データを取得するとすぐにその位置がどの区画に含まれるかを判定して識別番号を付与する。距離計算を行う際には、注目するストリーム G 空間データの周囲すべてについて距離を計算するのではなく、距離の近い区画(図中破線の円と重なる区画)の識別番号をもつストリーム G 空間データのみを距離計算の対象とすることにより、距離計算の必要なストリーム G 空間データの組み合わせを大幅に削減できる。

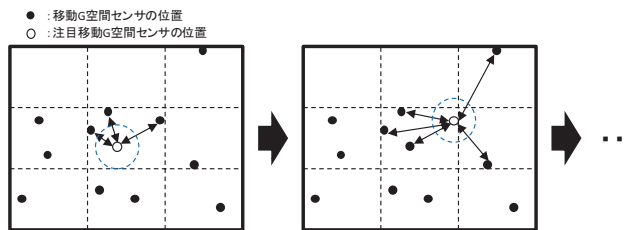


図 2 大規模ストリーム G 空間データ処理技術 (提案手法)

3.2 実装方法

以下、表 1 にスマートフォンなどの移動体端末から生成されるデータを想定した、本研究で扱うストリーム G 空間データの構成要素を示す。

表 1 ストリーム G 空間データの構成要素

	データ名	データ
1.	生成時刻	データ生成時の時刻
2.	ID	データを生成したセンサや人等に紐付けられた ID
3.	緯度	データ生成時の位置情報(経度)
4.	経度	データ生成時の位置情報(緯度)
5.	投影座標系の X 座標	経度(#3)・緯度(#4)を投影座標系 (IX 系) に変換した X 座標
6.	投影座標系の Y 座標	経度(#3)・緯度(#4)を投影座標系 (IX 系) に変換した Y 座標

以下、図 3 に提案手法の処理手順および各(1)~(4)の処理内容を示す。(1)~(4)は、直近で受信したストリーム G 空間データ(以下、受信データと呼ぶ)を 1 件 1 件順に処理していく。

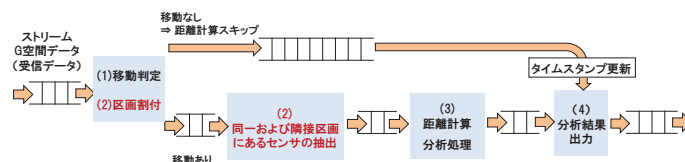


図 3 提案手法の処理手順

(1) 移動判定

受信データと前回受信した同一 ID のデータとの位置(座標)を比較し、受信データの ID が移動しているか否かを判定する。その結果、前回受信時から移動していなければ、以降(2)(3)の周囲との距離計算処理をスキップする。これは、受信したストリーム G 空間データが移動していなければ周囲の変化が小さいと想定するためである。

一方、前回受信時から移動していた場合は、周囲の変化が大きいと想定し、以降(2)(3)の周囲との距離計算処理を行なう。

(2) 区画割付および同一・隣接区画にあるデータの抽出

受信データの位置情報(投影座標系の座標)から、その受信データが投影座標系上に定義された格子状のどの区画に属するかを判定し、その区画 ID を付与する。そして、受信データと同一区画および隣接区画(図 2 の破線の円と重なる区画)の区画 ID をもつストリーム G 空間データを、受信データと距離計算をする対象ストリーム G 空間データとして抽出する。

(3) 距離計算および分析処理

(2)で抽出された距離計算対象のストリーム G 空間データに対して、受信データとの距離を計算する。距離計算は、ユークリッド距離に基づいた二乗距離で計算する。

その距離計算結果に対して、本実装では、「周囲の人から孤立している人」を特定するケースに着目し、分析シナリオとして、距離計算結果から受信(着目)データの周囲に存在するストリーム G 空間データの数、すなわち、人の位置データが更新されるたびに、その人の周囲にいる人数をカウントしていくことで、周囲から孤立している人を検知していく。

具体例として、ある人の周囲に滞在する人数をカウントする際の距離(以下、周囲距離と呼ぶ)を 200m、周囲に滞在する人数が 0 人のときに孤立している可能性があるともみなす場合の分析イメージを図 4 に示す。図 4 の例では、受信(着目)データが A の位置へ更新された場合、A の周囲 200m には 6 人滞在しているとカウントされるため A の位置にいる受信データは孤立しているとはみなされない。一方、受信(着目)データが B の位置へ更新された場合、B の周囲 200m には 1 人も滞在していないため B の位置にいる受信データは孤立している可能性があるともみなされる。

(4) 分析結果出力

(3)の分析および検知した結果を、一時領域で保持する。保持された結果の利用例の 1 つとしては、地図情報システムなどのアプリケーションからのリクエストによって逐次参照されることを想定している。

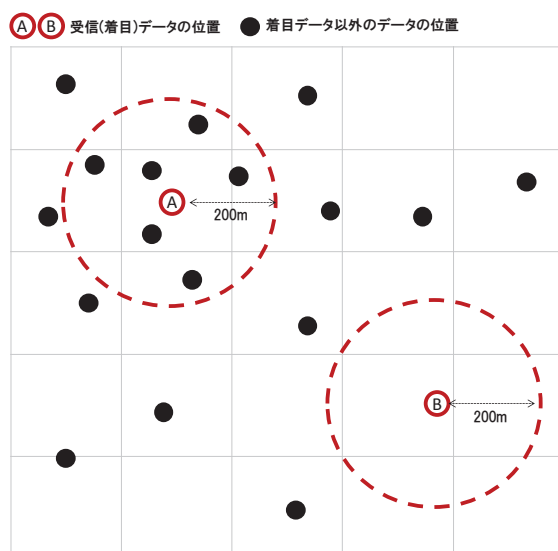


図 4 孤立検知の分析イメージ

4. 性能評価

4.1 評価目的

提案手法の有効性の確認、および次節に示す本評価における条件下(表 2)で、100 万台のセンサが 5 分に 1 回送信するデータを単一サーバでリアルタイムに処理できること、すなわちスループット 3,000 件/秒を満たすことを目標とし、その達成度合いを確認することを目的とする。

4.2 環境および条件

以下、本評価における測定環境、評価方法および疑似データの生成について詳細を示す。本評価では、ストリーム G 空間データとして、東京大学「人の流れデータ」⁶⁾から生成した疑似データを用いる。人の流れデータは、国や地方自治体等で調査を行っているパーソントリップ調査データから作成した人の流れのデータであり、防災分野や交通分野などの様々な研究で利用されている実績をもつ⁷⁾。

(1) 測定環境

図 5 に、提案手法の評価の測定環境の構成を示す。本評価では、疑似データの生成も含めて、CPU : 1.80GHz × 4 コア、メモリ : 16GB のスペックを持つ単一サーバ内に図 5 に示す処理を実装し、測定を実施した。

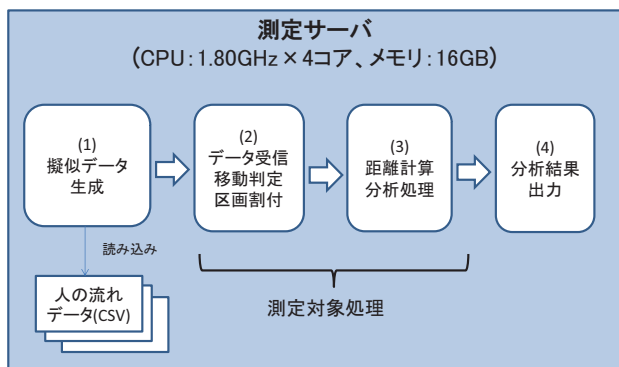


図 5 測定環境

(2) 評価方法

本評価では、図 5 に示す(2)および(3)の処理を測定対象とし、評価指標として 1 秒あたりに処理できる処理件数(=スループット)とした。具体的には、図 5 に示す(1)から受信データが 1 件転送される毎に実行される(2)+(3)の処理を 1 件あたりの処理とし、1 秒間あたりに処理可能な処理件数とした(処理件数は、図 3 に示す距離計算/分析処理スキップ分も含む)。スループットの求め方としては、擬似データ生成プログラムにより、転送間隔を 1 分間隔から n 分間隔へ順に間引いていくことで段階的に転送レートを落としていき、キューがほぼ溜まらない状態で処理されたときの転送レートを提案手法のシステムのスループットと見なす(1 分間隔と 2 分間隔の間など n が小さい場合は、転送レートに大きな差が出てしまうので、厳密には正確な最大スループットは求められず、その場合には推定値で補う)。

なお、本評価における前提条件および設定根拠を表 2 に示す。

(3) 擬似データの生成

本評価では、連続的に擬似的な位置データを送信するプログラムを作成した。擬似データについては、東京大学保有「人の流れデータ 2008 年東京都市圏(空間配分版)」⁹⁾に基づき、首都直下型地震において火災や倒壊の危険性が指摘されている地域⁴⁾の中から、人口約 100 万人、人口密度約 10,000 人/km² の地域を想定して生成した。また、データの転送量については、全 ID(表 1 の ID に相当)のデータを 1 分間隔で転送した場合、最大約 25,000 件/秒のレートで転送することができ、パラメータで転送間隔を分単位で間引くことで転送レートを調整できるようにした。

以下、表 3 に本評価で用いた擬似データ生成プログラムおよび擬似データについて、各時間帯および転送間隔における実際の転送レートを示す。表 3 によると評価対象地域には、8 時に 1,395,382 人滞在しており、それらすべての人のデータを 1 分間隔で送信した場合、1 秒あたりの転送レートに換算すると 23,256 件/秒で本測定環境へ転送することになる。

表 2 本評価における前提条件および設定根拠

#	項目	前提条件	設定根拠
1	対象地域	東京 23 区北部	首都直下型地震で、火災や倒壊の危険性が指摘されている地域 ⁴⁾ を想定。
2	人口	約 100 万人	
3	人口密度	約 10,000 人/km ²	
4	周囲の人数をカウントする際の距離(周囲距離)	200m	孤立状態として、他の人から目視や声などで確認できない範囲を想定。

表 3 各時間帯および転送間隔における転送レート

時間	人口[人] (人口密度[人/km ²])	転送レート(件/秒)		
		1 分 間隔	5 分 間隔	15 分 間隔
8:00	1,395,382 (15,461)	23,256	4,651	1,550
14:00	1,559,764 (17,283)	25,996	5,199	1,733
18:00	1,471,472 (16,320)	24,525	4,905	1,635
21:00	1,228,724 (13,615)	20,479	4,096	1,365

4.3 結果および考察

本節では、従来手法に対する提案手法の有効性の結果、考察を示す。また、本研究の目標に対する結果、考察および改善案を示す。

4.3.1 従来手法と提案手法の比較評価

提案手法の有効性を確認するために、本評価における条件下(表 2)で、従来手法との比較を行なった結果を図 6 に示す。

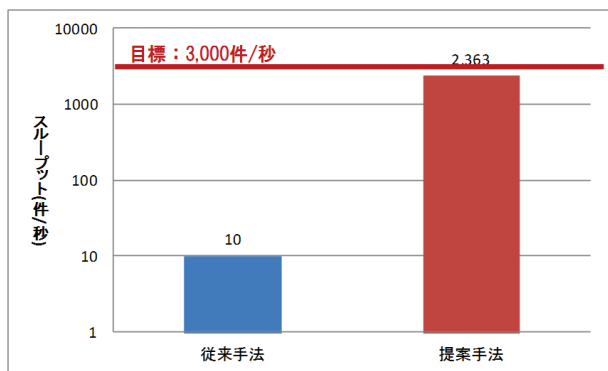


図 6 従来手法と提案手法の測定結果

本結果により、提案手法は従来手法より高いスループットを示し、提案手法が有効であることが分かった。ただし、提案手法のスループットは、目標スループット 3,000 件/秒を満たすことはできなかった。

4.3.2 提案手法による並列化(改善案)

4.3.1 節の評価の結果、表 2 に示す条件下では、目標スループット 3,000 件/秒を満たすことはできなかった。そこで、測定環境の CPU およびメモリの利用状況を調査した結果、スループットが上限に達しているにも関わらず、測定環境の CPU 利用率およびメモリ利用率ともにまだ余力を残していることが分かった。特にメモリに空きがある状況下で CPU 利用率は平均 25% 程度であったことから CPU を効率よく利用することで性能が向上できると仮定して、単一サーバで提案手法による並列化を検討した。基本的な考え方としては、対象地域を複数のエリアに分割してそれらを並列に処理できる仕組みを検討した。

以下、並列化の実装方法として、エリアの分割方法および並列化のアーキテクチャについて示す。

(1) 並列化のためのエリアの分割方法

並列化のためのエリアの分割方法については、エリアの境界付近のデータに対する考慮が必要となる。以下、エリアを 4 分割した例を図 7 を用いて説明する。

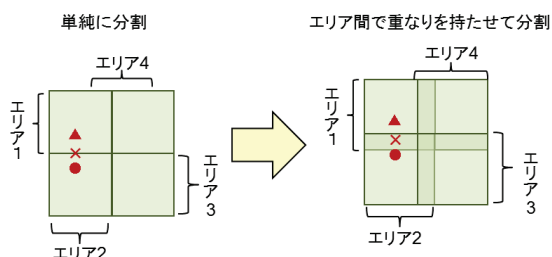


図 7 並列化のためのエリア分割方法

図 7 の左図のように、単純にエリアを分割してしまうと、距離計算対象となるべき×(エリア 1)と●(エリア 2)は、エリアが異なるため計算対象にならない。そこで、本並列化

手法では図 7 の右図のように、エリア間で重なりを持たせて分割することで、×はエリア 1 とエリア 2 の両方でそれぞれ処理されるようにし、エリア 1 とエリア 2 の結果を集約して最終的な×の結果となるように実装した。これにより、エリアの境界付近の距離計算漏れを解消することができる。

(2) 並列化のアーキテクチャ

以下、図 8 に並列化のアーキテクチャおよび処理の流れを示す。

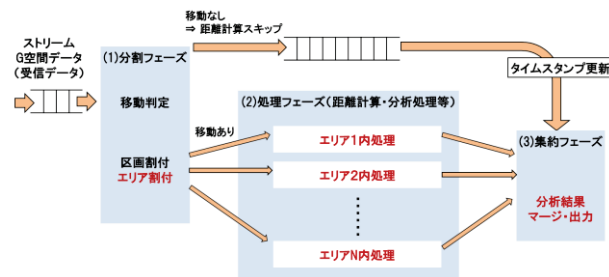


図 8 並列化のアーキテクチャ

<分割フェーズ>

受信データがあらかじめ分割されたエリアのどのエリアに属するかを判定し、そのエリア ID を受信データに付与する。このとき、エリアの境界付近の受信データについては、複数のエリア ID が付与されることもある。

<処理フェーズ>

分割フェーズで付与されたエリア ID に基づいて、そのエリア内において距離計算および分析処理を行ない、各エリアにおいて受信データの周囲 X m に何人の人が滞在しているかをカウントする。このとき、エリアの境界付近の受信データについては複数のエリアで処理されることもある。

<集約フェーズ>

処理フェーズにおいては、エリアの境界付近の受信データは複数のエリアで処理されることがあるため、それらの処理結果を 1 つにマージする処理を行なう。例えば、ある受信データについて、エリア 1 における周囲人数が 3 人で、エリア 2 における周囲人数が 5 人であった場合は、マージ処理によって受信データの周囲人数は 8 人とカウントされ、このマージ処理の結果に対して、最終的に孤立検知を行なう。

4.3.3 提案手法による並列化の結果と考察

提案手法における並列化手法の有効性を確認するために、本評価における条件下(表 2)で、単一処理時と 4 並列処理時の比較を行なった結果を図 9 に示す。

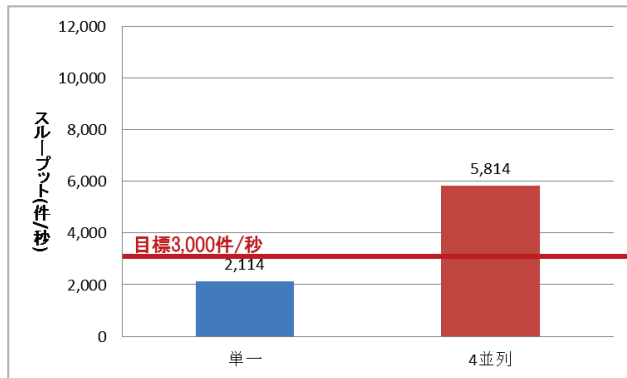


図 9 単一処理時と 4 並列処理時の測定結果

本結果により、4 並列処理は、単一処理より高いスループットを示し、提案手法による並列化が有効であることが分かった。さらに、本評価における条件下(表 2)で、目標スループット 3,000 件/秒を満たすことができた。

なお、単一処理時に比べて、4 並列処理時の性能向上が 3 倍弱(単一処理時のスループット 2,114 件/秒に対して、4 並列時の 1 スレッドあたりの平均スループット 1,454 件/秒)に留まった要因としては、分割された各エリアの人口密度に偏りがあったために、実際には 4 つのエリアに均等にデータが分散されなかったことが大きな要因であることがわかった。そのため、データが均等に分散されるように、人口密度に応じて分割するエリアの大きさを変えることで、さらなる性能向上が図れる可能性はあると考える。

4.3.4 性能評価まとめ

以下に、性能評価全体をまとめる。

- ・従来手法に比べて、スループットを大幅に向上でき、提案手法の有効性を確認した。ただし、単一処理時では、目標スループット 3,000 件/秒は満たせなかった。
- ・対象エリアを 4 分割し、処理の並列化を行った結果、単一処理時と比べてスループットの約 3 倍向上し、目標スループット 3,000 件/秒を満たすことができた。

5. まとめ

本研究では、大規模なストリーム G 空間データ群に対して、それらの間の距離が閾値以下であることを高速に判定する手法を提案した。また、首都直下型地震で火災や倒壊の危険性が指摘されている地域を想定して性能評価した結果、100 万台のセンサが 5 分に 1 回(3,000 件/秒)位置情報を送信する状況において、単一サーバでリアルタイムに処理できることを確認した。

本研究では、疑似データを利用して評価を行ったが、実

際の移動体データの動きはそのときの状況によって様々であると考えられる。今後の課題としては、並列処理に伴う転送データ数の均一化など、様々な状況を考慮した並列処理の効率化が必要であると考えられる。また、本実装ではストリーム G 空間データ(受信データ)が前回受信時から移動していなければ、距離計算処理をスキップする実装としたが、比較対象である他方のストリーム G 空間データも時間の経過とともに移動するため、利用目的によっては、受信したストリーム G 空間データが移動していなくても、数十分に 1 回は距離計算処理を行う必要があると考え、今後あらゆるケースを想定して検証していく必要がある。

謝辞 本技術は、総務省の『G 空間プラットフォームにおけるリアルタイム情報の利活用技術に関する研究開発』による委託を受けて実施した研究開発による成果です。

参考文献

- 1) 内閣府：日本の災害対策，入手先(http://www.bousai.go.jp/1info/pdf/saigaipamphlet_je.pdf) (2015).
- 2) 地震調査研究推進本部地震調査委員会：相模トラフ沿いの地震活動の長期評価（第二版）について，入手先(http://www.jishin.go.jp/main/chousa/14apr_sagami/index.htm) (2014).,
- 3) 内閣府中央防災会議首都直下地震対策検討ワーキンググループ：首都直下地震の被害想定と対策について（最終報告），入手先(http://www.bousai.go.jp/jishin/syuto/taisaku_wg/pdf/syuto_wg_report.pdf) (2013).
- 4) 東京都都市整備局 地震に関する地域危険度測定調査（第 7 回）(平成 25 年 9 月公表)
- 5) Mohamed F. Mokbel, Walid G. Aref: SOLE: scalable on-line execution of continuous queries on spatio-temporal data streams. VLDB J. 17(5): 971-995 (2008)
- 6) 東京大学 空間情報科学研究センター 人の流れプロジェクト，“人の流れデータとは，”
http://pflow.csis.u-tokyo.ac.jp/?page_id=10
- 7) 東京大学 空間情報科学研究センター 人の流れプロジェクト，“データを活用した共同研究，”
http://pflow.csis.u-tokyo.ac.jp/?page_id=12