

令和 5 年 6 月 27 日現在

機関番号：14301

研究種目：挑戦的研究（開拓）

研究期間：2020～2022

課題番号：20K20632

研究課題名（和文）ラプラス方程式を使った冊子体三次元画像からのページデータ抽出

研究課題名（英文）Extraction of page datasets from 3-D CT data of booklets from Laplace equation

研究代表者

小山田 耕二（Koyamada, Koji）

京都大学・学術情報メディアセンター・教授

研究者番号：00305294

交付決定額（研究期間全体）：（直接経費） 10,600,000円

研究成果の概要（和文）：本研究では、3D CTスキャナーで取得した冊子体の3D画像から、ページデータを正確に抽出する手法を開発することが目的とされた。仮説として、各ページ面を等値面とするようなスカラー場を生成すれば、正しくページ情報を抽出できるとされた。具体的には、ラプラス方程式を使用してスカラー場を生成し、ページ面を等値面として抽出し、画像をマッピングする方法を提案した。さらに、教師あり学習を用いた画像分類器を構築し、提案手法を2つのデータセットで検証した。結果として、提案手法はテキストの重ね合わせがあっても、高い精度でページを抽出することができた。

研究成果の学術的意義や社会的意義

この研究では、3D CTスキャナーで取得された冊子体の3D画像からページデータを正確に抽出する手法を開発した。この手法は、スカラー場を生成するためにラプラス方程式を使用し、教師付き学習を使用して画像分類器を構築した。提案手法は、2つのデータセットでテストされ、高い精度でページデータを抽出することが示された。この研究の成果は、歴史的資料や図書館のコレクションのデジタル化に役立ち、文化遺産の保存に貢献することが期待される。

研究成果の概要（英文）：This research aimed to develop a method using the Laplace equation to effectively extract page data from 3D images of book blocks obtained using a 3D CT scanner. The hypothesis was that if a scalar field can be appropriately generated so that each page corresponds to an iso-surface, page information can be accurately extracted from 3D images. The proposed method was applied to two datasets consisting of 20 and 50 pages, respectively. The results showed that the proposed method can effectively extract page data even when there are overlapping texts or the pages are not completely separated. The method has the potential to be applied to the digitization and preservation of historical and cultural materials.

研究分野：可視化情報学

キーワード：可視化

1. 研究開始当初の背景

2017-2018 年度に実施された挑戦的萌芽研究により、冊子体のポリウムデータからページ領域を平面でなくても 1 枚のページであると認識する手法が開発された。具体的には、研究代表者が提唱した Active Grid 法を用い、与えられた三次元画像から冊子体に対応する空間領域を直方体格子に写像することが可能となった。Active Grid は、三次元格子モデルを前提としており、ポリウムデータ内部の三次元領域に合わせて各格子を変形させるものである。

また、金属元素を含むインクを用いて印字されたページから構成される冊子体を作成し、解像度 1024×1024×557 のポリウムデータを作成することで、Active Grid 法によるページ領域の空間セグメンテーション技術の有効性を評価した。その結果、少し撓んだ冊子体であっても直交体格子に写像し、冊子体を構成する 1 ページの情報を正確に抽出できることが確認された。ただし、場合によっては計算が収束しないことがあることが指摘されている。

さらに、深層学習技術を用いた冊子体を構成するページ文字認識技術も開発された。文字情報を含むページを画像データとして作成し、これを複数積み上げることにより、ポリウムデータを構成した。このポリウムデータを教師データとし、アフィン変換（回転、並進、拡大縮小）を適用して得られた疑似冊子体ポリウムデータを複数作成し、FCN (Fully Connected Network) ベースの深層学習を実施した。その結果、疑似冊子体ポリウムデータから元のポリウムデータが復元できることが明らかになった。ただし、ページに印字された文字情報に依存するモデルであるため、復元が十分に機能しない場合があることが報告されている。

このような問題を解決する研究構想として、三次元画像にアノテーションを行い、それに基づきページ情報が適切に表現されるスカラ場を生成することを提案した。

2. 研究の目的

近年、古文書の解読においては、そのもろさから開くことができない課題に直面していた。しかし、最新の研究により、CT スキャン装置を用いた手法が提案され、古文書の解読に光明が差していることが報告されている。特に、イタリアのポンペイで見つかった数千年前の古文書については、位相型 CT スキャンデータや「シンクロトロン」という粒子加速器を駆使したデジタル化技術が用いられ、驚くべき成果を生んでいる。こうしたデジタル技術により、未開封の冊子体を含めた古文書のデジタル化が進んでおり、その分析に関しても多くの期待が寄せられている。

そこで、本研究では、ラプラス方程式を駆使し、三次元 CT 装置で取得された冊子体三次元画像（図 1 参照）からページデータを効果的に抽出する方法を開発することを目的とする。学術的な問いである「冊子体から生成された三次元画像から正確にページ情報を抽出するにはどのような手法があるか？」に対して、本研究は、「冊子体の各ページ面を等値面とするようなスカラ場を生成することができれば、三次元画像から正確にページ情報を抽出できる」という仮説に着目し、その有用性を確認することを目指す。冊子体は、インクで印字された複数枚のページを綴じたものであり、ページを捲ることなく、冊子体を構成するページを見ることが目標である。



図 1 吸収型の X 線 CT を使った冊子体三次元画像の取得

本研究では、吸収型 X 線 CT を用いて、金属元素を含むインクが印字されたページから構成される冊子体を解析することを目的とする。しかしながら、現実的な問題として、紙とインクの線吸収率において、十分な差異が生じない場合、インク領域と紙領域の間に区別をつけることが困難になる可能性がある。このような場合、線吸収率に関するコントラストを増強するため、X 線の波長を変化させ、与えられた紙とインクのペアに対して最大のコントラストを示す波長を決定することが必要である。

研究開始に先立ち、市場で流通する印刷紙およびインクの線吸収率の波長依存特性を調査し、印刷紙・インクの組み合わせに対して、波長によるコントラスト最大化手法を開発した。本研究では、この手法を用いて、酸化鉄を 4-50% 含むインク (Canon Cartridge 039 H) で両面印刷された、厚さ 0.2 mm の普通紙郵便はがきから構成される冊子体試料を解析し、解像度 1024 × 1024 × 557 の三次元画像を作成する予定である。解析には、島津製作所製 inspeXio smxSMX-90CT puls を使用する。

3. 研究の方法

「冊子体の各ページ面を等値面とするようなスカラ場を適切に生成することができる」という仮説の前提に対応する研究計画は、次のように述べられる。具体的には、コンデンサ内部の電位を計算するためによく使用され、厳密解が知られているラプラス方程式を利用する。この方程式の未定係数は、十分な数の冊子体内部の点でページ番号と三次元座標を対話的に指定することによって同定する。

同定されたラプラス方程式を使用してスカラ場を生成し、その等値面を抽出し、三次元画像にマッピングすることで、冊子体の各ページを表現することができる。このプロセスを容易にするため、三次元グラフィックス技術を使用して対話的な環境を構築し、任意の断面を指定することができる。指定された点が特定のページ番号に対応するように、複数回の選択プロセスが行われる。こうして、ページ番号と三次元座標の組み合わせが複数特定される。(図2参照)

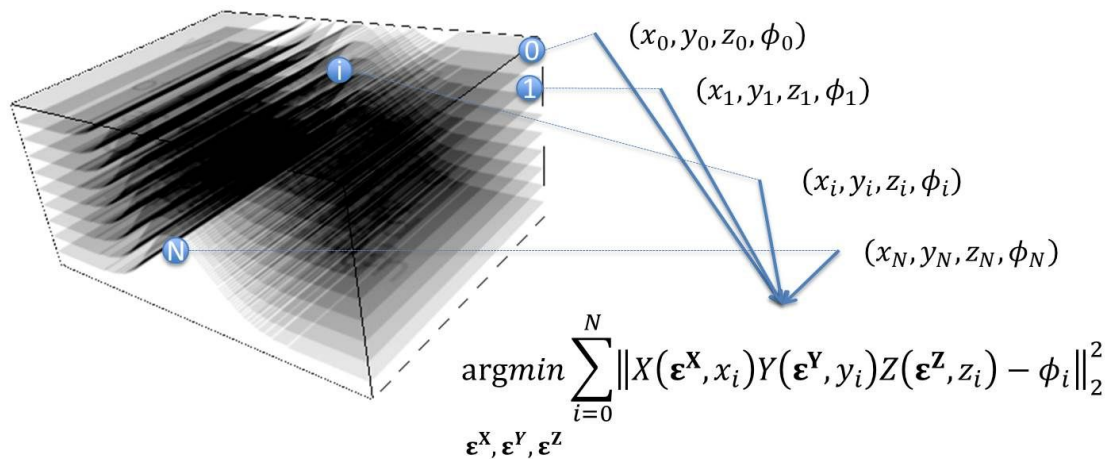


図2 ラプラス方程式を使った、冊子体の各ページ面を等値面とするスカラ場生成

ただし、冊子体は印刷されたページから構成されているが、各ページが平面であるとは限らず、曲面である可能性があることに留意する必要がある。ページが交差しないことを仮定し、各ページを曲面として扱うことが適切である。また、冊子体の表紙と裏表紙をコンデンサの電極として扱い、その内部の電位を生成することで、ページが等電位面に対応するようにすることができる。

対話環境を利用したスカラ場の構築においては、事前に決められたラプラス方程式の厳密解を使用し、未知の係数との差を定式化することにより、より適切なスカラ場が生成されたかどうかを判断する。内部点の数を増やすことで、生成されたスカラ場の誤差を減らすことができるかどうかを調べ、必要な内部点の数を決定する。

仮説の帰結である「三次元画像から正しくページ情報を抽出することができる」という問題に対して、複数の画像データを用いて、等値面が適切にページ情報を表現しているかどうかを目視で判断し、教師あり学習を用いて画像分類器を構築する。この分類器を用いて、各ページのマッピング適切さを判定し、全ページで適切なマッピングを行うように未知の係数を最適化することで、仮説が検証される。具体的には、マッピング適切さを最大化するような未知の係数を探索する。図3に示されているように、左の画像ほどマッピング適切さが高く、ページ情報を正確に抽出できることが期待される。

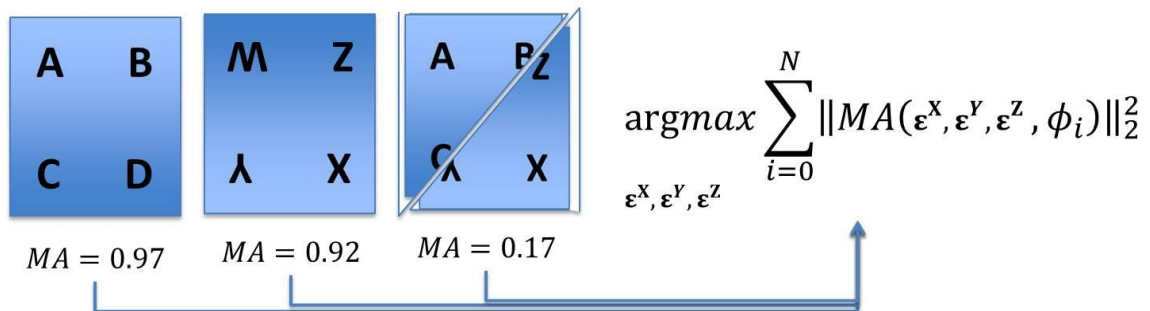


図3 すべてのページにおいてマッピング適切さの総和を最大化

4. 研究成果

WebGL を使ったアノテーション

WebGL (Web Graphics Library) は、複雑な 3D グラフィックのレンダリングとユーザーとのインタラクションを必要とするさまざまなプロジェクトのカスタマイズされた作業環境を作成することができる、Web 3D グラフィックスプロトコルの新しい標準の 1 つとなっている。このツールは、JavaScript と OpenGL ES 2.0 (組み込みシステム向けの OpenGL) の利点を組み合わせている。私たちは、HTML5 (Hyper Text Markup Language) キャンパスでシーンをレンダリングするスクリプトを開発し、OpenGL によるハードウェアレンダリングの加速を組み合わせている。HTML スクリプトはブラウザ自体以外の追加プラグインを必要としないため、プログラムをインターネット上で公開利用するのに有利であり、OpenGL モクロスプラットフォームの標準である。この技術を使って、CT 機でスキャンされた冊子情報を含む全体の 3D ポリウムデータを可視化する作業プラットフォームを構築した。これは、透視の原理に従って 1 つの 3D 空間を提供し、主要なオブジェクト (ポリウムデータ) を空間の中心に配置し、カメラを初期状態ではその上に配置することを意味する。ユーザーは、左ボタンをドラッグしてオブジェクトを回転させたり、マウスホイールをスクロールしてカメラとオブジェクトの距離を調整したり、右ボタンでオブジェクトをドラッグして移動したりすることで、いくつかの基本的な操作を行うことができる。これらのアクションにより、ユーザーはオブジェクトのすべての詳細を明確に見ることができる。内部構造を観察するために、私たちは互いに垂直な 3 つのスライスを作成し、座標を変更できるようにした。また、ページの境界を強調して、ページ上の特徴点のサンプルを取得しやすくするために、グレイ値の範囲を指定して、グレイ値がその範囲外のポイントを非表示にする機能を追加した。

当初のプログラムでは、ユーザーは実際のページ上の点をダブルクリックすることで、スライス上の点を選択し、3D 座標とページ番号を CSV ファイルに記録した。これはブラウザのコンソールからユーザーが確認できる。しかし、このサンプリング方法は効率が低いことが判明した。そこで、ポイントに従って可変半径の球体を作成することで再設計を行った。ユーザーは適切な半径を設定し、選択したページ上の点から次のページや他の可視点のどの点よりも距離が短くなる。その後、この球体内のすべての点は、探している正確なページ上にある必要がある。そして、ユーザーがその球体をダブルクリックすると、そのグループの点に関する座標情報が CSV ファイルに記録される。シングルクリックによる回転機能とダブルクリックによるサンプリングの間に競合があるため、回転機能のオン・オフのオプションがコントロールバーに追加された。

スカラー場の生成

冊子の CT データを 3D アノテーションすることにより、大量の 3D 座標とそれらに対応するページ番号を取得できる。私たちは、これが実際のページの近似位置に対応しているかどうかを 3D 可視化を使用して確認できる。実際、これは全体のごく一部であり、例えばページに約 200 のラベルがあれば、おそらく 10 万以上の実際のページボクセルポイントが存在する。

このプログラムの目的は、得られた 3D 座標データを使用して、全体の 3D 空間の構造を計算することである。例えば、キャパシタとして、空間内のすべての点のポテンシャルを計算する。私たちが採用した手法は、キャパシタフィールドが実際にラプラス方程式と互換性があるため、ラプラス方程式である。ページ番号が既に 3D 座標に代入された場合、ラプラス方程式を一意に決定するための、一連のパラメータが得られる。このパラメータセットは、ターゲットデータの空間構造モデルを表している。このパラメータセットを使用して新しいスカラー場を計算し、それを用いて全空間内のすべての点のページ番号を求めることができる。

適用事例

提案手法は、2 種類のデータセットに適用された。

第 1 のデータセットは、432 x 275 x 438 の解像度と 0.0538mm のボクセルサイズを持つ 20 ページから構成された。抽出されたページは完全であり、その上のテキストは明確に見える。

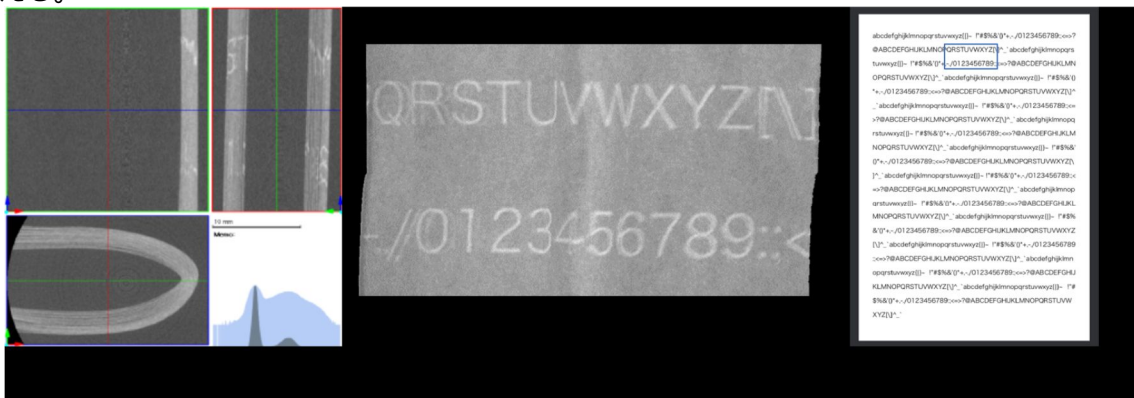


図 4 20 ページから構成された文書から作成されたデータセット

第2のデータセットは、970 x 489 x 976の解像度と0.0499mmのボクセルサイズを持つ50ページから構成されていた。A4用紙の厚さが0.1mmに達しても、図から各ページの位置を特定することは困難である。ページの抽出結果(二値化処理済み)には、識別が困難な重ね合わせられたテキスト情報が含まれている。

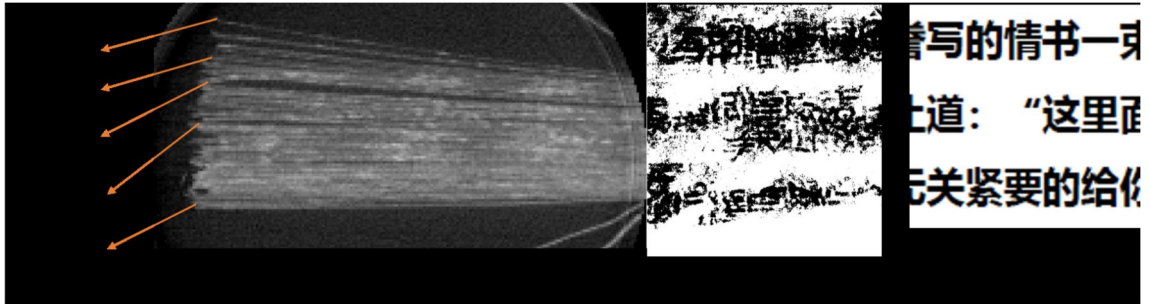


図5 50ページから構成された文書から作成されたデータセット

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Ou Jiarui、Han Zhongjiang、Koyamada Koji	4. 巻 8
2. 論文標題 Three-dimensional book data page segmentation and extraction method using Laplace equation	5. 発行年 2021年
3. 雑誌名 Journal of Advanced Simulation in Science and Engineering	6. 最初と最後の頁 223 ~ 236
掲載論文のDOI（デジタルオブジェクト識別子） 10.15748/jasse.8.223	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Koyamada K.、Long Y.、Kawamura T.、Konishi K.	4. 巻 12
2. 論文標題 Data-driven Derivation of Partial Differential Equations using Neural Network Model	5. 発行年 2021年
3. 雑誌名 International Journal of Modeling, Simulation, and Scientific Computing	6. 最初と最後の頁 19-29
掲載論文のDOI（デジタルオブジェクト識別子） 10.1142/S1793962321400018	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計2件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 韓 忠江, 欧 家瑞, 小山田 耕二
2. 発表標題 三次元CTで取得した冊子体データからのページ抽出の高精度化
3. 学会等名 第48回可視化情報シンポジウム
4. 発表年 2021年

1. 発表者名 OHAN ZHONGJIANG, OU JIARUI, 小山田 耕二
2. 発表標題 三次元CTで取得した冊子体データからの ページ抽出の高精度化
3. 学会等名 可視化情報シンポジウム
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------