序論

（近年のCTスキャナーの小型化，高精度化により，医療用途を中心にCTスキャナーの利用事例が豊富になった．その中で古文献の解読にCTスキャン技術を用いる研究が現れるようになった[]．それらは主に，経年劣化による構造的な脆さや学術的希少価値を理由に開くことができなくなった書物の内部を非侵襲的に測定し，可視化するといったものである.）

歴史的な文献は，忘れられた出来事の記述を長期間保存し，人々が後世になって歴史を紐解く大きな手がかりを与えてくれる．ただそのような古文献は経年劣化やその他の外的な浸食によって，開くことやページをめくることができないほど構造的に脆くなっている場合がある．そのような古文献の調査には非侵襲的な測定法によって内部を可視化する必要がある．

非侵襲的な測定が求められる古文献の実例としては，1750年，イタリア・カンパニア州において Karl Weber がヘラクラネウム付近にて発見した，紀元79年のヴェスヴィオ火山噴火によって炭化した1800枚を超えるパピルス紙の例が挙げられる[]．この例ではパピルス紙は筒状に巻かれており，損傷させずに開くことは困難であった．またこのような事例は文献の経年劣化や古代の災害による損傷に限らず，近年の洪水や火災によって保管されていた文献が開くことができなくなる場合もある．そのような例に，ドイツのDuchess Anna Amalia Library での2004年の火災が挙げられる．この火災では62000を超える古文献が，火災による高温と焼失，さらに消火の際の水によって重大な被害を被った[]．

日本においても，歴史的な文献が水損などによって固着する被害のほか，本の綴じ目近くの「のど」と呼ばれる箇所に綴じ込まれている箇所や，紙背文書と呼ばれる，ある文書が別の文書の紙の裏側に使用された結果，袋綴じ状態となった箇所の調査手法はいまだに確立されておらず，元の文書が修復される際に，副次的に調査することが可能になるという現状がある．

非侵襲的な測定手法として最も広く普及しているものに，3次元X線CTイメージングが挙げられる．この測定手法は近年，医療用途に限らず世界文化遺産などの調査にも用いられている[]．過去の研究は，通常のマイクロCTシステムであっても古文献の調査に有効であることを示している[]．インクと紙のセルロースの間に，X線の減衰率の差が少しでも存在すれば測定結果に一定のコントラストとして現れるので，十分な解像度を持った3次元X線CTスキャナーであれば，古文献を非侵襲的にデジタル化することが可能である．この手法の欠点としては，測定時に対象が放射線に曝露されるため，経年劣化がスピードを加速してしまうことが挙げられる．

測定対象に放射線を浴びせずにデジタル化する手法の一つに，3次元テラヘルツイメージングがある．この手法の欠点としては，テラヘルツ波の減衰が理由で測定深度が限られていることである．文献を対象とする場合数ページの測定に限られてしまう．現時点では，テラヘルツ波が実際の歴史的文献の調査に用いられた例はなく，インクに含まれる金属粒子とテラヘルツ波の相互作用も十分に明らかではない．またX線を用いる別の手法として，X線位相イメージングがあり，複数の古文献の調査へ利用されている[]．ただしこの手法を用いた測定装置は非常に大きく，持ち運びが困難であり，測定可能な施設も限られている．そのため図書館等に保管されている貴重な文献を測定する際に，研究施設に移動させる必要性があるためコストやセキュリティが測定による成果に見合わないという課題がある．マイクロCTスキャナーであれば，3次元テラヘルツイメージングに比較してより深度まで測定することが可能であり，X線位相イメージングの測定装置に比較して測定環境の準備が容易であるため，古文献の可視化に現時点でもっとも有効な手段と考えられる．

また測定手法を問わず，測定後得られた冊子体ボリュームデータからページを抽出するためには，ページ番号と大量のボクセルの対応付け（以降，アノテーションと呼ぶ．）を行う必要があり，文献の可視化の高速化，大規模化に対するボトルネックとなっている．

そこで本研究では，測定手法をマイクロCTスキャナーとして，ラプラス方程式を用いた冊子体ボリュームデータのアノテーションの補間を行うシステムを提案する．これは冊子体ボリュームデータを一種の積層コンデンサとみなすことで，アノテーション数を減らすことを可能にする手法である．

本稿では，実際のCTスキャンによって得られた冊子状ボリュームデータに対する，本手法の有効性を定量的に分析する実験を行う．

本手法を用いて冊子体ボリュームデータからページ抽出を行う際に十分なページ抽出精度を保ちながらアノテーション数を削減することを可能にする条件を明らかにするために，検証実験をデザインする．

具体的には，鉄分を含むインクを用いて作成した擬似的な古代文献（以降，擬似試料[要修正]と呼ぶ．）を用意し，この擬似試料に対してCTスキャンを行うことで得られた冊子体ボリュームデータへアノテーションを施し，一方で本手法を様々な条件下で施した．人によるアノテーションデータを正解データとして，本手法を適用することで抽出されたデータとの比較によって本手法のページ抽出精度を定量的に評価した．

本研究において，仮説は以下のように設定する．

H1 冊子体ボリュームデータからのページ抽出課題において，ページ番号を推定する関数がラプラス方程式を満たすと仮定し，関数のパラメータを最適化すれば，ページ抽出の精度を保ちながら，必要なアノテーション数を減らすことを可能にする．

H1.1 冊子体ボリュームデータからのページ抽出課題において，ページ番号を推定する関数がラプラス方程式を満たすと仮定し，関数のパラメータを最適化すれば，ページ抽出の精度を保ちながら，1ページあたりに必要なアノテーション数を減らすことを可能にする．

H1.2 冊子体ボリュームデータからのページ抽出課題において，ページ番号を推定する関数がラプラス方程式を満たすと仮定し，関数のパラメータを最適化すれば，ページ抽出の精度を保ちながら，アノテーションが必要なページ数を減らすことを可能にする．

H1.3 冊子体ボリュームデータからのページ抽出課題において，ページ番号を推定する関数がラプラス方程式を満たすと仮定し，関数のパラメータを最適化すれば，ページ抽出の精度を保ちながら，1ページあたりに必要なアノテーション数とアノテーションが必要なページ数を同時に減らすことを可能にする．

H2 本手法で用いるラプラス方程式に適切な補正項を加えることで，冊子体ボリュームデータのページ分布をより精度よく推定できて，アノテーション数をさらに削減する

[間に合うか不明]

本手法ではラプラス方程式のパラメータ推定に非線形最小二乗法を用いる．非線形最小二乗法のアルゴリズムはLevenerg-Marquardt 法を用いる．

非線形最小二乗法は，最小二乗法を非線形関数のパラメータフィッティングへ拡張したもので，目的変数と説明変数の間の関係を表す式をサンプルデータとモデル関数の残差平方和の最小化によって得る方法である．本研究では，数値解析ソフトウエア SciPy に実装されている非線形最小二乗法の関数を用いる．

本稿の以降の構成は次のとおりである. まず第 2 章では関連研究について説明する.続いて第 3 章で冊子体ボリュームデータについて述べ，第 4 章では今回の実験で用いる手法について、第 5 章では仮説と仮説の検証方法を述べる.第 6 章では実験結果について述べ，第 7 章では実験結果の考察をする．最後に第 8 章で結論を述べる.

関連研究１

本章では関連研究について記述する．

2.1節では

CTスキャナーによるボリュームデータの取得について

ボリュームデータからの平面抽出

について説明する．

2.2節では

等値平面（以降，ISO Surfaceと呼ぶ）の可視化について述べる．

2.3節では

非線形方程式のパラメータ推定について記述する．

2.4節では

OCRにおける文字認識について述べる．

ボリュームデータからの平面抽出とその可視化について

CTスキャナーなど３次元測定装置によって得られたボリュームデータから平面抽出する方法には複数の手法が存在する．

・人間のアノテーション

・ISO Surface

CTスキャナーについて

非線形関数のパラメータ推定について

まず線形関数のパラメータ推定についてであるが，もっとも古典的な手法として最小二乗法があげられる．想定する関数をfとして，既知の関数

時計 が含まれている画像

自動的に生成された説明

・ボリュームデータからのページ抽出

・ISO Surface の可視化について

・非線形方程式のパラメータ推定について

・OCRについて

・CTスキャナーによる紙のスキャンについて