

ビジュアル・ジョイン: 重ね合わせ操作による実空間コンテンツの融合

武田 十季[†] 鶴野 玲治[‡] 牛尼 剛聡[‡]

[†]九州大学大学院芸術工学府

[‡]九州大学大学院芸術工学研究院

E-mail: [†]toki@gospel.aid.design.kyushu-u.ac.jp, [‡]{tsuruno,ushiama}@design.kyushu-u.ac.jp

あらまし 人は日常空間に存在する様々な情報表現(コンテンツ)を利用して情報活動を行う。この際、複数のコンテンツに含まれる情報を統合して利用することが多い。我々は、携帯端末のカメラ機能を利用して複数の実空間コンテンツを重ね合わせる操作によって、複数の実空間コンテンツに含まれる情報をユーザが効率的かつ効果的に利用可能な機構を開発中である。ここでは、重ね合わせられた実空間コンテンツを、相互の意味内容に基づいて仮想的に融合し、拡張現実として携帯端末上のディスプレイに表示する。この際、存在する実空間コンテンツの種類には多様性があり、重ね合わせのパターンによって複数の融合形式が考えられるため、それらの融合を統一的に扱うモデルが必要である。本論文では、複数の実空間コンテンツの意味構造とレイアウト構造に基づいて、それらを融合した視覚表現を生成する一般的なモデルとして、ビジュアル・ジョインを提案する。また、提案手法に基づいたプロトタイプ・システムを用いた被験者実験により提案モデルの有効性を評価する。

Summary People utilize various presentations of information (contents) in their daily life for information activities. In such cases, people often integrate the information included in multiple contents. We have been working for development of a system which enables a user to browse information represented in multiple real world contents effectively and efficiently. In this system, a user can integrate virtually multiple real world contents on the basis of their semantic relationship on the display of a mobile device as augmented reality by overlapping operations. There are various types of real world contents and we can suppose some types of integration methods according to overlapping patterns, so a unified integration model is necessary. In this paper, we propose the visual join function, which is a model to integrate virtually real world contents based on their semantic structures and their layout structures. Then, we discuss the effectiveness of the proposed model based on the results of the user studies using our prototype system.

Visual Join: Actual Contents Integration by Overlapping Operation

Toki TAKEDA[†] Reiji TSURUNO[‡] and Taketoshi USHIAMA[‡]

[†] Graduate School of Design, Kyushu University

[‡] Faculty of Design, Kyushu University

1. はじめに

情報技術および情報化社会の発展に伴い、日常の活動の中で必要な情報を取得する要求が高まっている。それらの要求を満たすために、携帯端末を利用してユーザのニーズに合った情報を提示する研究が盛んに行われている。例えば、オンライン地図をはじめとする地図サービスでは、現在位置の地図の表示や目的地までのルート表示、乗り換え案内など、携帯端末に適した機能が開発され、利便性の向上に寄与している[1-3]。しかし、デジタル・コンテンツによる情報の閲覧が普及しているにもかかわらず、人はカレンダーや地図などの実空間上で実体を有する情報表現（以下、**実空間コンテンツ**と呼ぶ）を用いて必要な情報を取得することがある。

人が実空間コンテンツから情報を取得する際には、複数の実空間コンテンツを相互に参照して取得することがある。例えば、店舗紹介が掲載された情報誌では、店舗の一覧ページと地図のページが存在し、それらコンテンツを相互に比較することによって、店舗の場所を地図で確認することがある。また、道端のポスターで鑑賞したい映画を見つけた際には、ポスターや映画の上映スケジュールと自分のスケジュール帳を用いて、日時の確認を行うこともある。このように、私たちは日常の中で、複数の実空間コンテンツを相互に参照し、異なるコンテンツに含まれる情報を組み合わせ、必要な情報を獲得している。しかし、実空間コンテンツは、デジタル・コンテンツのように容易に情報の付加や融合ができないため、ユーザにとって、わかりにく

いことがある。本研究では、カメラ付き携帯電話を用いて、複数の実空間コンテンツに含まれている情報を融合し、ユーザへ提示することによって、ユーザがより効果的に情報閲覧可能とすることを旨とする。

2. 関連研究

近年、セカイカメラ[4]や実空間透視ケータイ[5]のように、カメラ付き携帯電話を用いた拡張現実[6]に関する研究や開発が盛んに行われている。その中でも実空間上のコンテンツを拡張することに着目した情報閲覧の研究も多く行われている。

HOTPAAPER[7]では、携帯電話を用いた紙媒体への、情報付加、取得を実現している。HOTPAAPERでは、携帯電話のカメラで紙面を覗いた場所を特定するために、文字列の空間的分布の特徴ベクトルを算出することによって、対象とする実空間コンテンツを認識としている。Rohs[8]らの研究では、実空間上の地図にあらかじめ、ドットを格子状に配置し、各格子内の32×32ピクセルの領域を全て画像として登録しておくことによって、携帯電話のカメラから地図を覗いた際に、画像のマッチングを行い、地図上のどの部分を覗いているか認識可能としている。

これらの研究では、任意の実空間上のコンテンツに付加情報を対応づけ、それらの情報を携帯電話のカメラから取得している。このため情報を取得する閲覧者にとっては、あらかじめ対応づけられた、コンテンツと付加情報の関係のみの閲覧に限られている。それに対して、本研究では、複数の実空間上のコンテンツを重ね合わせる組み合わせの違いによって情報が変化することで、個々のユーザに最適な情報を提供する手法を提案する。

異なる意味や機能を持ったオブジェクト同士を組み合わせることによって、新たな意味や機能を構築することに焦点を当てた研究としてはData Tiles[9]がある。Data TilesではRFIDを埋め込んだ透明な「タイル」が情報サービスを表しており、それらをディスプレイの上にレイアウトして、物理的な関係性を作ることによって、情報のカスタマイズが可能となる。本研究では、タイルが機能やコンテンツとしての役目を持つのではなく、コンテンツ自身を直接組み合わせる。また機能とコンテンツの組み合わせにより情報を編集、カスタマイズするのではなく、本研究では、実空間で興味のあるコンテンツを組み合わせることによって、よりユーザにとって最適な情報へ絞り込むという視点が大きく異なる点である。

3. Pick and Lapに基づく情報融合

3.1. Pick操作とLap操作

本研究の基盤となるPick and Lap操作に基づくコンテンツ融合の概要を述べる。



図1:コンテンツ・セグメント

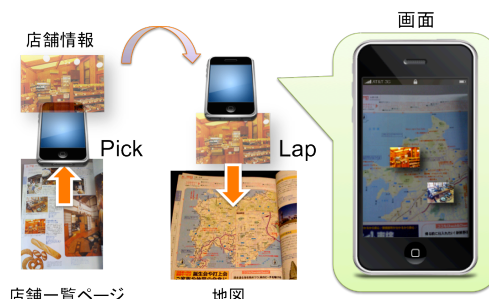


図2:Pick and Lapに基づく情報融合の実現例

Pick操作とは、カメラ付き携帯電話で実空間上のコンテンツを撮影することにより、対象とするコンテンツ全体、またはコンテンツの一部を選択する操作である。ここで、選択されたコンテンツ全体またはその一部をPickコンテンツと呼ぶ。

一方、Lap操作は、Pickコンテンツを、他の実空間上のコンテンツ全体、またはコンテンツの一部分へ重ね合わせる操作である。ここで、Pickコンテンツを重ね合わせる対象となるコンテンツをLapコンテンツと呼ぶ。

本研究では、対象とする実空間上のコンテンツの例として、旅行情報誌を利用する。旅行情報誌には、観光スポット、ホテル、店舗、土産に関する情報を紹介するページや、それらの場所を確認するための地図のページ等、様々な種類のコンテンツが含まれている。本研究では、観光スポット、レストラン、ホテル等に関する個々の紹介記事および個々の地図や時刻表等をPick操作およびLap操作の単位とする。ここで、Pick操作やLap操作の対象となるコンテンツの部分をコンテンツ・セグメントと呼ぶ。

本手法を実現するために、対象とするコンテンツの全体、またはコンテンツの一部を認識する必要がある。本研究では、コンテンツの認識対象として、2次元の矩形画像を利用する。すなわち、Pick操作やLap操作の対象に、地図やそれぞれの紹介記事に付加されている矩形画像を用いる(図1)。この画像を、コンテンツ・セグメント全体を代表して表す代表画像とする。一方、地図やカレンダーなどは、コンテンツ全体を画像として捉えることができ、コンテンツ・セグメント自身が代表画像と捉えることができる。

Pick and Lap操作が行われると、相互のコンテンツの意味的な関係性に基づいて、Lapコンテンツが持つ意

味的な空間が決定され、Pickコンテンツはその意味的
空間上での適切な位置関係に基づいて融合処理が行わ
れる。このとき融合処理は、代表画像の移動、拡大・
縮小によって行う。それぞれの具体的な動作例として
実空間コンテンツに旅行情報誌を用いた場合の融合結
果を図2に示す。店舗をPickして地図にLapすると携
帯電話画面上では、地図の上に店舗の代表画像が重畳
して表示される。つまり、PickコンテンツとLapコンテ
ンツに含まれる情報がAR技術を利用して融合されユー
ザに提示される。

旅行情報誌を対象にしたコンテンツの重ね合わせ
による具体的な表現は第4章の融合モデルで述べる。

3.2. 処理の流れ

提案手法の処理の流れの概要を図3に示す。Pick操
作、またはLap操作を行うと、実空間コンテンツに含
まれるコンテンツ・セグメントは携帯端末のカメラを
利用して矩形画像として認識される。システムはあら
かじめそれぞれのコンテンツ・セグメントに対応する
矩形画像とそれに対応するメタデータが登録されてお
り、画像の特徴量に基づいて対象とするコンテンツ・
セグメントが同定され、コンテンツ・セグメントに対
応するメタデータが取得される。

メタデータには店舗や地図などコンテンツ・セグメ
ントの種類が含まれ、さらに、それぞれのコンテンツ・
セグメントが表現する意味的なデータが含まれる。こ
こで、意味的なデータとは、たとえば、地図の場合
では対象領域、店舗の場合では店舗の位置を示す緯度経
度、開店時間、座席数などの情報が含まれる。

Pickされるコンテンツ・セグメントとLapされるコン
テンツ・セグメントの種類の組み合わせや順番に基づ
いて融合モデルが決定され、それぞれの意味的なデー
タと、Lapされるコンテンツの実空間上での矩形画像の
大きさや位置に基づいて、融合結果をLap対象コンテ
ンツ上に重畳して表現するためのパラメータが決定され
る。

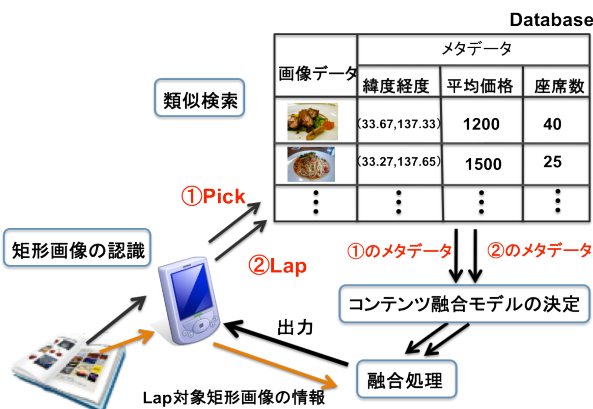


図3: 処理の流れ

4. 実空間コンテンツの融合

コンテンツ同士の融合を実空間上で視覚的に表現
するためにビジュアル・ジョインを提案する。

代表的なデータモデルであるリレーショナル・デー
タモデルでは、データを表として構造化し、複数の表
を組み合わせることで様々な目的のためにデータを効
果的に扱うことができる。ビジュアル・ジョインでは、
実空間コンテンツの代表画像において、画像上の物理
座標に対して意味座標を対応づけることによって、リ
レーショナル・データモデルにおいて、複数の表を結
合するのと同様なアプローチで、コンテンツを融合し、
コンテンツの意味内容に基づいて仮想的に融合するこ
とを可能とする。

本章では、ビジュアル・ジョインを実現するために、
コンテンツの融合の論理的な枠組みを示し、論理的な
関係を視覚的に表現する方法を示す。またこれらのコ
ンテンツ融合を表現レベルで形式化する手法として意
味的アフィン変換を提案する。

4.1. 論理モデル

コンテンツ・セグメントを代表して表す代表画像は
以下の2種類に分類できる。

- ・ **Map型画像**：意味的な広がりをもつ画像
- ・ **Icon型画像**：意味的な広がりをもたない画像

ここで意味的な広がりをもつ画像(Map型画像)とは、
画像の部分要素に対して、異なる概念を対応付けるこ
とができる画像のことである。画像の部分要素とは、
画像上の、点、線、領域である。この定義のもとで、
画像空間を部分要素の集合として離散化した場合、
個々の部分要素の基本単位は、画像が持つレイアウト
構造によって変化する。

例えば、地図は場所を示す点の集合と考えることが
できる。一方、カレンダーの場合は、目的に応じて異
なるレイアウト構造が用いられる。例えば、日にちを
分かりやすく示すブロック型や、1日のスケジュール
を管理しやすいようにタイムメモリが縦軸に並んだパ
ーティカル式が存在するように、利用の目的に合わせて
構造の種類が異なる。こうしたカレンダーは主に矩
形領域ごとに、日にちや時間といった意味的な情報と
対応づけられる。つまり、画像空間を、意味を持つ要
素に分解する単位は、コンテンツの構造に依存してお
り、点や線、領域の場合が考えられる。

意味的な広がりをもつ画像上における要素は、二つ
の座標系での意味を持つ。例えば、地図上の点は、紙
媒体上での位置を示す座標系と、現実空間における緯
度経度情報を示す座標系との二つの座標系での意味を
持つ。これらの座標を、前者を物理座標、後者を意味
座標と呼ぶ。それぞれの物理座標は意味座標と対応付
けされ、図4のように記述できる。このように、意味的

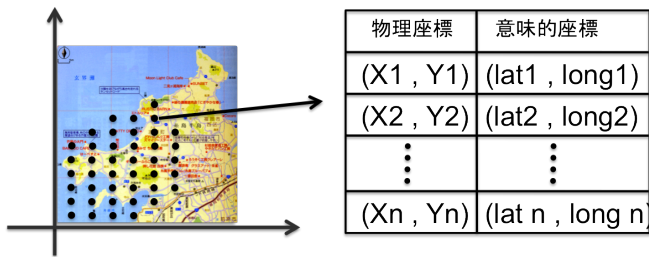


図4:地図の物理座標と意味的座標

な広がりを持つ画像では、コンテンツのレイアウト空間を構成する各要素と、意味的な情報とが対応づけることができるため、Map型画像と呼ぶ。

意味的な広がりを持たない画像 (Icon型画像) とは、情報誌に含まれる店舗画像のように、コンテンツ・セグメントを代表して表す画像のことを指す。Icon型画像は画像自身が別のコンテンツと対応づけられる1つの要素となり、アイコンとして捉えることができる。

いま、店舗画像をIcon型画像と捉え、図5のように意味的なデータとして位置情報や開店時間、平均価格を持っていた場合、地図の位置情報と共通の属性である店舗の位置情報とが結合され、店舗画像は、地図上の物理座標系における点と対応づけられる。つまりこれはリレーショナル・データベースのJoinと同様の振る舞いである。

次に、店舗のように意味的な広がりを持たない画像 (Icon型画像) 同士を重ね合わせた場合についての対応付けについて説明する。

Icon型画像は、画像自身が異なる概念と対応付けることのできる点であり、重ね合わせられるコンテンツがどちらもIcon型の場合は、相互のIcon型画像が対応付けられる意味的な広がりを持った空間を動的に構成し、仮想的な空間として提示に利用する。このとき仮想的に作り出される空間のことを仮想Mapと呼ぶ。仮想MapはMap型コンテンツと同様に、物理座標と意味座標の二つの座標系を持ったMapである。このときの物理座標と意味座標の対応付けはLapコンテンツを中心とした関係性の表現方法によって定められる。この表現方法は、PickコンテンツとLapコンテンツの共通属性によって決定される。

つまり、Icon型画像の重ね合わせでは、相互の共通の属性値と意味的に対応する仮想Mapが呼び出され、この仮想Mapと、Pickコンテンツ及びLapコンテンツと結合することで、相互の関係性が可視化される。例えば、店舗Aと店舗Bが共通属性に緯度経度を持っていた場合、二つの店舗の位置関係を示す仮想Mapが呼び出される。図6は、仮想Mapと店舗A(Pickコンテンツ)との結合を示したものである。このときの仮想Mapの物理座標は、店舗B(Lapコンテンツ)を中心とした店舗A(Pickコンテンツ)の方角を表す円周上の点の集合から成り、各点に対応する意味座標は、その点が示す方角に含まれる緯

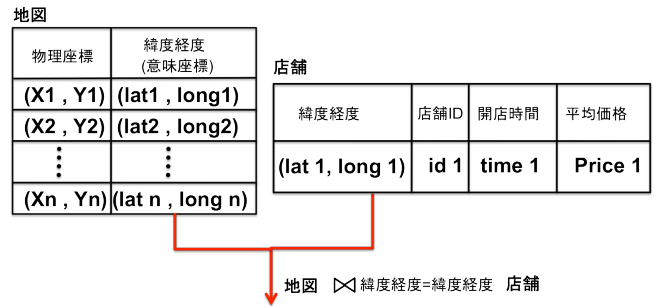


図5:地図の店舗の緯度経度による結合

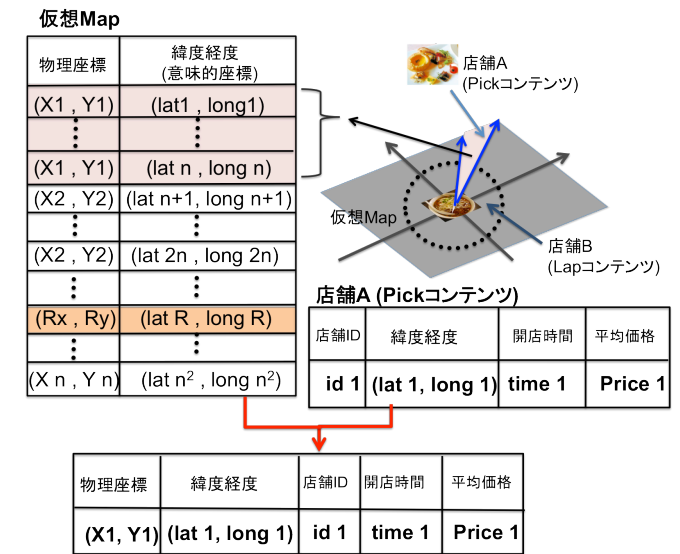


図6:Icon型画像同士の重ね合わせ
平均価格の仮想Map

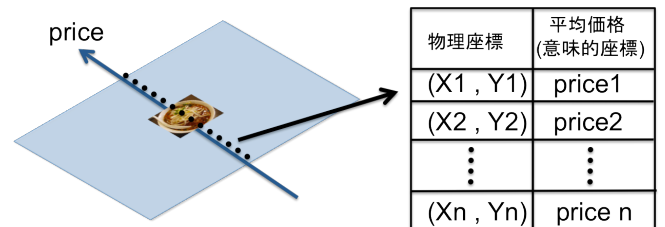


図7:平均価格を示す仮想Map

度経度の点の集合である。図6における仮想Mapの物理座標 (X1, Y1), (X2, Y2), ..., (Xn, Yn) は円周上のn個の点である。それぞれの点は、円周の中心を結んだ領域に含まれる緯度経度の点集合を円周上の点として表したものである。例えば (X1, Y1) は、円周の中心座標 (Rx, Ry) から (X1, Y1) へ向かって広がる領域の緯度経度 (lat1, long1), (lat2, long2), ..., (lat n, long n) と対応付けられる (図のピンク部分)。図6に示す結合では、店舗Aの緯度経度は (lat1, long1) であることから、仮想Mapの意味座標である緯度経度と結びつき、図の赤い矢印の先に示すように、店舗Aは仮想Map上で物理座標 (X1, Y1) に対応付けられる。また、店舗Bの緯度経度 (lat R, long R) が、仮想Map上の物理座標における円周の中

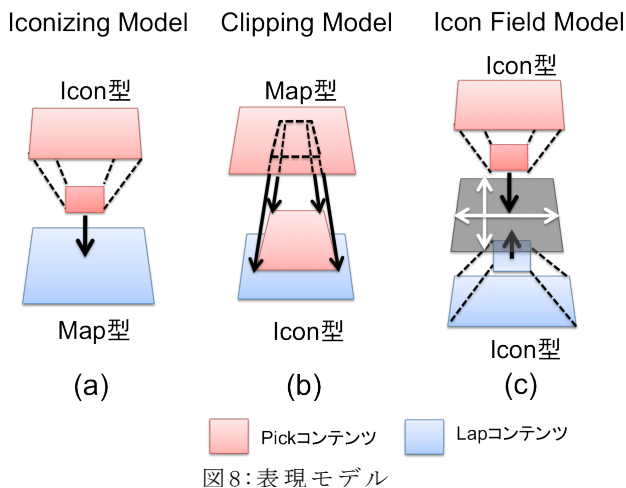


図8:表現モデル

心座標 (Rx, Ry) と対応付けられることによって, 仮想 Map は Lap コンテンツを中心に作られる.

共通属性に平均価格を持っていた場合も同様に, Lap コンテンツを中心に縦軸で表現した物理座標と, その各点が平均価格を示す点の集合で構成された意味座標を持った仮想 Map が呼び出され, Pick コンテンツと Lap コンテンツの関係性を可視化する (図7).

ここで共通属性が複数存在する場合, どの属性に対する仮想 Map を呼び出すかという問題がある. コンテンツの重ね合わせる順番や組み合わせから文脈を考慮することで, ユーザの意図に合った Map を提示することは課題の一つである. また提示の複数切り替えのようなインタフェースの設計を行うことによって解決することも可能である.

4.2. 表現モデル

複数のコンテンツ・セグメントを重ね合わせる組み合わせと順番には, 多数のパターンが存在する. 前章ではコンテンツの代表画像を大きく二つのタイプに分けて定義したが, これらのタイプの組み合わせの違いによって以下の三つの表現モデルを定義する.

A) Iconizing Model

このモデルは Pick コンテンツが Icon 型コンテンツで, Lap コンテンツが Map 型コンテンツの時に適用される. Pick した画像はアイコンとして見なされ, Lap コンテンツの代表画像上で対応する点に重ね合わせて表示される (図8:a)

B) Clipping Model

このモデルは Pick コンテンツが Map 型で, Lap コンテンツが Icon 型の時に適用される. Pick コンテンツ内で Lap コンテンツに対応する領域をクリッピングして Lap 対象コンテンツの矩形画像領域内に表示する (図8:b). つまり Pick コンテンツは Lap 対象によって情報が絞り込まれ限定される.

C) Icon Field Model

このモデルは, Pick コンテンツと Lap コンテンツがどちらも Icon 型画像の時に適用される. このとき相互の

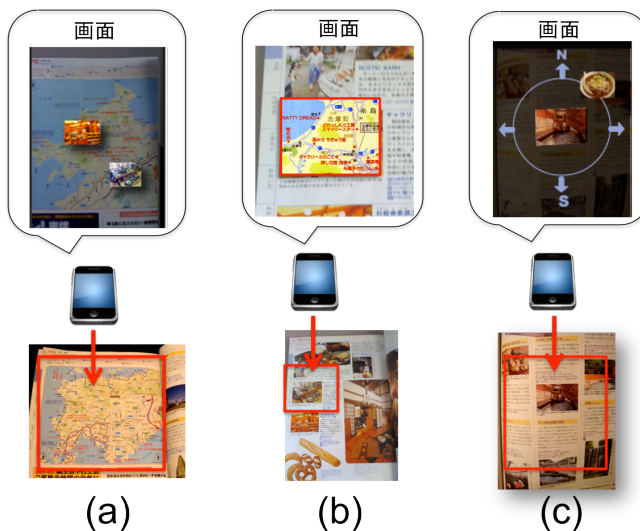


図9:情報誌への表現モデルの適用

共通属性である意味的情報を対応付けることが可能な Map を呼び出し, Lap コンテンツを中心に仮想的に設定することによって, 相互の意味的関係性を視覚的に表現する. (図8:c)

4.3. 旅行情報誌を利用した融合モデルの表現

前節の3つの表現モデルを, 旅行情報誌に適用した例を用いて説明する. 旅行情報誌には様々なコンテンツ・セグメントが含まれているが, ここでは店舗紹介記事(店舗)と地図, 観光スポットの3種類のコンテンツ・セグメントを対象にして以下の3つの重ね合わせのパターンに対する表現モデルの適用と表現方法を述べる.

1. 「店舗」を「地図」へ重ね合わせる
2. 「地図」を「店舗」へ重ね合わせる
3. 「店舗」を「観光スポット」へ重ね合わせる

「店舗」を「地図」へ重ね合わせた場合, Iconizing Model を適用し, 携帯端末の画面上では「店舗」の矩形画像を地図上で「店舗」が存在する位置に合わせて表示する (図9:a). これによって, ユーザは興味を持った「店舗」を, 閲覧したい「地図」上へ重ね合わせることで, 店舗の位置を確認することができる. また, 地図には同じ地域を示す地図でも略地図などのように種類の異なる地図が存在するが, ユーザは重ね合わせる操作によって, どの地図上でも店舗の位置を確認することが可能となる. さらに複数の店舗を Pick して地図へ Lap すると, 一度にそれぞれの店舗の場所を確認することも可能となり, 従来の旅行情報誌で閲覧するときのように店舗一覧ページと地図を交互に複数回参照する必要がなくなる.

「地図」を「店舗」へ重ね合わせた場合, Clipping Model を適用する. これはユーザが Pick した「地図」上で「店舗」の場所がどこに存在するかを確認できるように, 「店舗」の矩形画像領域内に「店舗」の場所を中心とした「地図」を表示する (図9:b). つまりユーザ

は、ある特定の地域を示した地図や、縮尺の異なる地図など、様々な地図の中から閲覧したい地図を選んでPickし、興味のある店舗へLapすることで、Pickした対象の地図上においてそれぞれの店舗がどのあたりに存在するかを確認可能となる。

「店舗」と「観光スポット」を重ね合わせた場合、Icon Field Modelを適用し、「店舗」と「観光スポット」の位置関係の比較を行えるように提示する。例えば、ユーザが特定の観光スポットに行こうと思った際に、その観光スポットから近い飲食店で食事したいと考えたとする。このような場合に、ユーザは興味を持った複数の「店舗」を選んでPickし、「観光スポット」へLapすると、それぞれの店舗が観光スポットから、どのくらい離れ、どの方向に位置するかということを確認できる。

位置関係を示すために、図9:cのように、「観光スポット」を中心にした方向を示す軸を仮想的に作ることで「観光スポット」から見た「店舗」の方向を示す。このとき「観光スポット」を中心にした円周上に「店舗」を表示することで、観光スポットから見た店舗の方向を示し、観光スポットから距離が近い店舗ほど大きく、遠い店舗ほど小さく表示することで距離を示す。これにより、複数の「店舗」をPickし、「観光スポット」へLapすると、「観光スポット」を中心とした円周上に複数の「店舗」が一度に表示され、相互の位置関係を容易に認識可能となる。

4.4. 意味的アフィン変換

前章では、コンテンツの融合の論理的な側面を、リレーショナル・データモデルとの類似という視点で説明した。本章では、コンテンツの融合を表現的な側面から形式化する、幾何意味的アフィン変換について説明する。Pick and Lap操作によるコンテンツの融合は、重ね合わせられる複数のコンテンツ・セグメントの代表画像に対して、それらの意味的な関係に基づいて、拡大・縮小等の幾何学的変換を行い、適切な配置で、適切な順番で重ね合わせるによって実現する。

2次元アフィン変換は2次元空間上の代表的な幾何変換である。2次元アフィン変換は、式(1)のような行列演算として表現できる。ここで、 (x, y) は変換前の座標であり、 (x', y') は変換後の座標である。

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1)$$

いま、2次元アフィン変換の変換行列をTとし、重ね合わせの対象となる2次元コンテンツをA, Bとするとき、Pick and Lap操作に基づく情報の融合は次式のように形式化できる。

$$Lap(Pick(A), B) = Over(Conv(T, A), B)$$

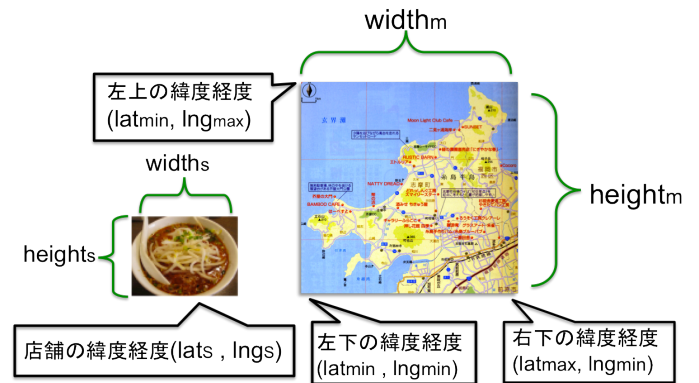


図10: 意味情報の登録

ここで、 $Pick(A)$ はAをPickすることによって得られるコンテンツ・セグメントを表し、 $Lap(C, B)$ はコンテンツ・セグメントCをコンテンツ・セグメントB上にLapすることを表す。 $Over$ は画像を重ね合わせる関数であり、 $Conv(T, A)$ はコンテンツAを2次元アフィン変換行列Tに基づいて幾何変換を行うことを表す。

意味的アフィン変換では、対象となるコンテンツの意味的な関係に基づいてTを決定する。なお、簡単化のために、幾何変換として、縦横比を維持した拡大・縮小と平行移動のみを考える。このとき、2次元アフィン変換は以下の式で表現できる。ここで、 t は拡大・縮小率であり、 α はX方向の移動距離、 β はY方向の移動距離を表す。

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} t & 0 & \alpha \\ 0 & t & \beta \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2)$$

変換行列Tは重ね合わせられるコンテンツ・セグメントの組み合わせによって変化する。店舗から地図へ重ね合わせた場合の例を用いて式(2)の変換行列Tを求める式を以下に示す。

図10のように、店舗のメタデータは $(lats, lngs, widths, heights, imgs)$ であり、地図のメタデータ $(latmin, lngmin, latmax, lngmax, widthm, heightm, imgm)$ であるとする。ここで、 $lats$ は店舗の緯度、 $lngs$ は店舗の経度を表し、 $widths$ と $heights$ は店舗画像の幅と高さを表し、 $imgs$ は店舗画像の画像を表す。また、 $latmax$ と $latmin$ は地図に表現された緯度の最大値と最小値を表し、 $lngmax$ と $lngmin$ は地図が表す緯度の最大値と最小値を表し、 $widthm$ と $heightm$ は地図画像の幅と高さを表し、 $imgm$ は地図画像を表す。

「店舗」をPickして「地図」へLapした場合、ディスプレイ上には図のようにカメラを通して見た地図上の適切な位置に店舗が表示される。つまり、Pickした店舗画像をLapした対象の地図上の適切な位置 (α, β) に表示するとしたとき、この (α, β) は、店舗の緯度・経度 $(lats, lngs)$ を地図画像の左下端を基準にして、携帯端末のディスプレイ上での地図画像の矩形の幅 $(widthd, heightd)$ を利用して、以下の式により求める。

$$\alpha = \frac{lat_s - lat_{min}}{lat_{max} - lat_{min}} \times width_d \quad \beta = \frac{lng_s - lng_{min}}{lng_{max} - lng_{min}} \times height_d$$

また、店舗画像の大きさは、携帯電話の画面上での地図の大きさによって変化する。携帯電話の画面上での地図の大きさに対する店舗画像の大きさの割合を s としたとき、店舗画像の拡大縮小率は $t = (width_d) / width_s$ と表せられる。以上の α 、 β 、 t は式(2)のように変換行列のパラメータとして表すことができる。

他の表現モデルの場合も同様に、コンテンツが持つ意味的データを利用して相互の関係性を表すパラメータを求めることができる。

5. プロトタイプ・システム

提案手法に基づいたプロトタイプ・システムをパーソナルコンピュータとUSBカメラを用いて実装した。

今回の実験の目的は、実空間コンテンツの融合結果に対する、ユーザの理解度や印象に関する評価を得ることであるため、実装において、携帯電話よりも実現しやすいパーソナルコンピュータを用いた。また、矩形画像を高精度で認識するために、用いる旅行情報誌は、認識対象の矩形画像以外の部分を白黒としている。

本プロトタイプ・システムで、3種類の表現モデルを適用した際の実行例を示す。図11に示すように、画面下部にあるPickボタンをマウスでクリックすることにより、コンテンツ・セグメントをPickできる。また、Lapボタンを押すと、PickコンテンツをLapできる。画面右には、PickコンテンツもしくはLapコンテンツが表示される。カメラが矩形を認識している間は、矩形に対して黄緑色の枠が表示される。図11(a)、(b)、(c)はそれぞれ、店舗を地図に重ね合わせた場合、地図を店舗に重ね合わせた場合、店舗を観光スポットへ重ね合わせた場合の結果画面を示している。

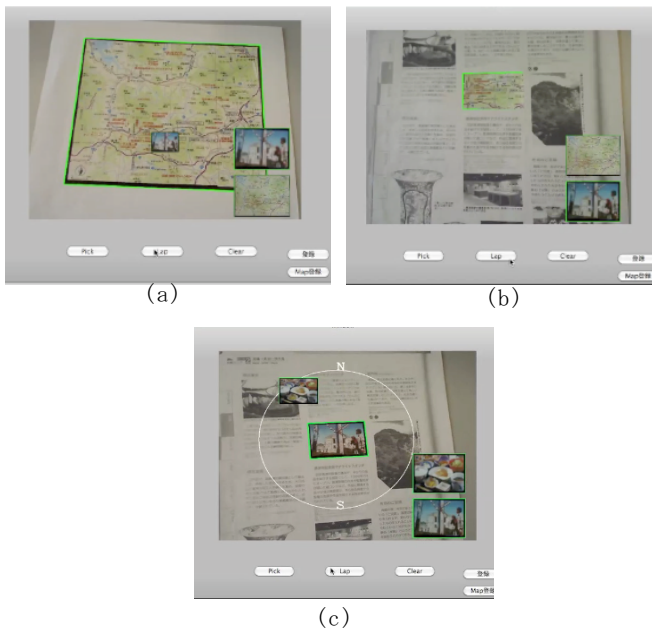


図11: 実行画面

5.1. 実験

プロトタイプ・システムを用いて、本手法によるコンテンツの融合の有効性を評価するために、被験者実験を行った。被験者数は10名(男 7:女 3)であり、年齢は21歳~25歳である。実験の方法としては、被験者に本システムを操作してもらい、質問紙により主観的評価を回答してもらった。被験者にはあらかじめ、本手法の概要およびPick操作とLap操作の説明をしておいた。各被験者は、「店舗を地図へ重ね合わせる」、「地図を店舗へ重ね合わせる」、「店舗から観光スポットへ重ね合わせる」の3つのパターンの操作をそれぞれ行ってもらい、各操作で得られた結果に対して、表現の自然さに関して5段階で評価してもらった。さらに、全てのパターンを実行した後に、重ね合わせるという操作が理解しやすかったかを5段階で評価してもらった。

融合結果は自然に感じられたか

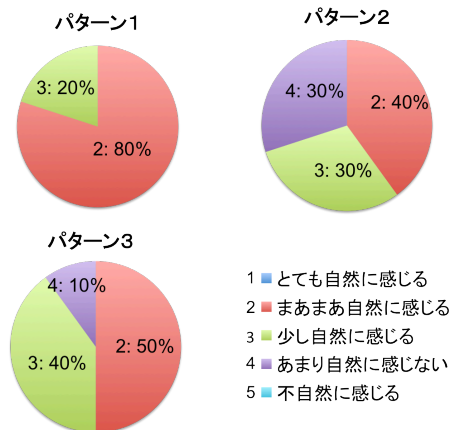


図12: 各パターンで得られた結果に対する印象の結果

重ね合わせる操作が自然であったか

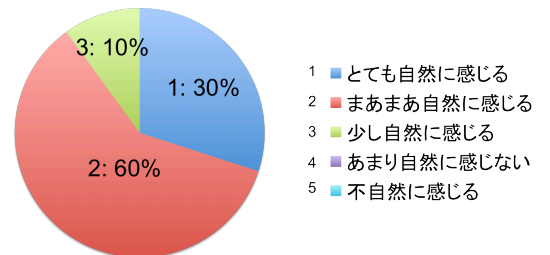


図13: Pick and Lap操作の理解しやすさに関する結果

操作は理解しやすかったか

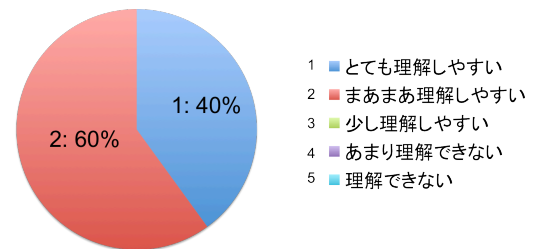


図14: コンテンツの重ね合わせ操作への自然さに関する結果

5.2. 結果と考察

結果を図12, 13, 14に示す. ここで, 「パターン1」は「店舗を地図へ重ね合わせる」操作であり, 「パターン2」は「地図を店舗へ重ね合わせる」操作であり, 「パターン3」は「店舗から観光スポットへ重ね合わせる」操作である.

結果から, 各パターンの融合結果に対する感じ方は, 肯定的な評価の中でも, 「とても自然に感じる」と答えた被験者はいなかった. 評価後にインタビューを行ったところ, 一度見たら分かるけれど, 初めて見た時は意外に感じたからという理由が多かった.

パターン1の場合は, ほとんどの被験者が肯定的であったが, 一方でパターン2とパターン3は「あまり自然に感じない」という否定的な評価も得られている. インタビューを行ったところ, パターン2に対して「クリッピングされた地図の画像が店舗の上に重畳されて, 店舗画像が見えなくなってしまう」といった回答が得られた. また, パターン3に対しては「表示される画像が増えると重なってしまうのではないか」「もう少し情報量が欲しい」という回答が得られた. こうした表示に関する改善は, 本研究の課題の1つと言える.

また全体的な操作性の理解度に関する評価では, ほとんどの被験者が一回で覚えられ, 分かりやすいという肯定的な意見が得られた.

6. まとめ

本研究では, 実空間においてコンテンツを相互に参照し, 情報を統合して利用していることに着目し, 実空間のコンテンツを仮想的に融合する手法について述べた. 携帯電話のカメラを利用して実空間コンテンツを重ね合わせることで, コンテンツの融合を行う操作を提案し, 融合結果をコンテンツの物理的なレイアウトを考慮して表現するためにPick and Lap手法を提案した. Pick and Lap手法は論理モデルと表現モデルから構成される. 論理モデルとしてビジュアル・ビジョンを提案し, また表現モデルの形式化として意味的アフィン変換を提案した. 意味的アフィン変換では, コンテンツのメタデータに含まれる意味的データを用いて, 変換行列のパラメータを導出する. 実験では, 以上の手法に基づいて旅行情報誌を使った融合モデルのパターンを実装し, 本手法への理解度, 直感性に関する評価実験を行い, 一定の評価が得られた.

本研究では, 旅行情報誌を対象に論じてきたが, コンテンツの融合をモデル化し, 表現のための処理を意味的アフィン変換によって形式化することによって, 様々な実空間コンテンツに応用することが可能である.

例えば, Map型コンテンツには地図以外にも, コンテンツの空間的広がりや時間的な広がりや, 時間の流れを表す時計などがあげられる. 別の

例としては, 道端の映画のポスターをPickし, 時計にLapした場合, 映画のメタデータに上映時間が含まれていたなら, 時計との共通属性は時間となり, 二つは対応付けが可能となる.

また, 本研究のプロトタイプでは, Pick操作からLap操作へと移り変わる時間間隔は短く, その場限りの重ね合わせになっていたが, Pickしたコンテンツを携帯電話内に蓄積し, Lapしたい場所や状況でPickしておいたコンテンツを選択するシステムにすることによって, より重ね合わせる対象の組み合わせに幅が広がり, 利用シーンの可能性も広がる.

今後の課題としては, 本手法を旅行情報誌以外の実空間コンテンツに適用したとき, 様々な組み合わせが考え得る中, 相互の実空間コンテンツに共通属性が複数存在した場合, どの属性を基準にして融合するかという問題が挙げられる. この場合, コンテンツを重ね合わせる前後の文脈から, ユーザの意図を考慮して情報提示を行うことが必要となると考えられる. また, 日常的な状況下による長期運用実験を実施した上での議論が必要なことが挙げられる.

参考文献

- [1] GMAP, available at <http://gmap.jp/gmap/>
- [2] ALPSLAB, available at <http://www.alpslab.jp/>
- [3] ゼンリンいつもNAVI, available at <http://www.zenrin-datacom.net/mobile/>
- [4] "セカイカメラ", available at <http://sekaicamera.com/>
- [5] "実空間透視ケータイ", available at <http://kazasu.mobi/>
- [6] R. T. Azuma, "A Survey of Augmented Reality", Presence, Vol. 6 No. 4, pp. 355-385, 1997.
- [7] B. Erol, J. Graham, E. Antunez, J. J. Hull, "HOT PAPER demonstration: multimedia interaction with paper using mobile phones", M. M2008, pp. 983-984, 2008.
- [8] M. Rohs, J. Schoning, A. Kruger, and B. Hecht, "Towards real-time markerless tracking of magic lenses on paper maps", In Adjunct Proceedings of the 5th Intl. Conference on Pervasive Computing Pervasive, pp. 69-72, 2007.
- [9] Jun Rekimoto, Brygg Ullmer, and Haro Oba, "DataFiles: A Modular Platform for Mixed Physical and Graphical Interactions", CHI2001, pp.269-276, 2001.