

ビジュアル・ジョイン：実空間コンテンツの仮想融合モデル

武田 十季^{1,†1,a)} 鶴野 玲治² 牛尼 剛聡²

受付日 2011年9月20日, 採録日 2012年1月6日

概要：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・ジョイン演算を提案する。ビジュアル・ジョイン演算では、複数の実空間コンテンツに含まれる画像を、その意味的な関係に基づいて融合可能である。ここで、画像の融合はアフィン変換に基づいて行われるが、そのパラメータは画像のメタデータに基づいて自動的に決定される。さらに、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有効性を評価する。

キーワード：拡張現実, モバイルインタフェース, コンテンツ融合, ビジュアルジョイン, Pick and Lap

Visual Join: Virtual Integration Model for Real World Contents

TOKI TAKEDA^{1,†1,a)} REIJI TSURUNO² TAKETOSHI USHIAMA²

Received: September 20, 2011, Accepted: January 6, 2012

Abstract: 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.

Keywords: augmented reality, mobile interface, contents integration, visual join, Pick and Lap

1. はじめに

近年、携帯端末上で様々なデジタル・コンテンツを利用可能となり、携帯端末を利用してユーザの要求する情報を提供する研究がさかんに行われている。たとえば、携帯電話上の地図サービスでは、現在位置の地図の表示や目的地までのルート表示、乗り換え案内等、携帯端末に適した機

¹ 九州大学大学院芸術工学府芸術工学専攻
Kyushu University, Fukuoka 815-8540, Japan

² 九州大学大学院芸術工学研究院
Kyushu University, Fukuoka 815-8540, Japan

^{†1} 現在, NTT サイバーソリューション研究所
Presently with NTT Cyber Solutions Laboratories

^{a)} tokitokibu@gmail.com

能が開発され、利便性の向上に寄付している [1], [2], [3]. しかし、すべての情報活動がデジタル・コンテンツを利用して行われているわけではない。人は雑誌、ピラ、看板等の、実空間上で物理的な媒体上の情報表現（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。

実空間コンテンツには、デジタル・コンテンツにはない利点が存在する。たとえば、雑誌では、携帯端末に比べ高精度で広い領域を用いた情報表現を安価に提供することができる。このような利点があるため、デジタル・コンテンツが普及している現在でも、実空間コンテンツは広く利用されている。しかし、実空間コンテンツはデジタル・コンテンツと比較していくつかの欠点がある。その1つに、ユーザの目的に応じたコンテンツ間の相互参照と融合が困難であるという点があげられる。人がコンテンツを利用する際には、目的に応じて、複数の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て、必要な情報を獲得することが多い。たとえば、旅行情報を確認する際には、店舗情報と地図を相互に参照することによって、店舗の位置を地図で確認することが行われる。デジタル・コンテンツでは、このような複数のコンテンツ間の参照や融合をハイパーリンク等の機構を利用して簡単に実現できる。しかし、実空間コンテンツに関しては、そのような相互参照や融合は、複数の実空間コンテンツの内容をユーザが記憶し、ユーザの頭の中でそれらを融合する必要がある。しかし、人間は多くの情報を記憶しておくことは困難であるため、複数の実空間コンテンツを利用して、相互の参照と融合が必要となる情報を取得するのは、困難である。

我々は、この問題を解決するために、携帯端末を利用して、利用者の目的に応じて実空間コンテンツの相互参照と融合を拡張現実 [4] として仮想的に行う手法を開発している。我々は、人が相互参照をする行為を、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置き換え、重ね合わせられた実空間コンテンツどうしを仮想的に融合するための操作体系として、Pick and Lap 手法を提案している [5]。Pick 操作は、カメラ付き携帯端末で実空間コンテンツを撮影することにより、コンテンツを選択する操作である。Lap 操作は、他の実空間コンテンツを撮影することで、Pick したコンテンツを重ね合わせる操作である。これらの操作が行われると、重ね合わせられたコンテンツを仮想的に融合し、結果を、Lap した対象の実空間コンテンツに対する拡張現実として携帯端末上のディスプレイに表示する。

図 1 は、Pick and Lap 手法に基づいて重ね合わせが行われた際の、実空間コンテンツの仮想的な融合の例を示す。ここでは、情報誌に記載された店舗一覧のうち、1 件の店舗情報が Pick され、Lap 操作によって地図に重ね合わせた



図 1 実空間コンテンツの仮想的な融合表現の例

Fig. 1 An example of virtual integration expression of actual contents.

場合、店舗を表す画像が、地図上で店舗が存在する位置に重畳して合成されたように、携帯ディスプレイ上の画面に拡張現実として表示する。これにより、ユーザは Pick 操作によって選択した店舗の地図上での位置を直感的に分かりやすく確認可能である。ここで、ユーザが重ね合わせる実空間コンテンツの組合せや順番の違いによって、融合形式を変化させることで、必要な情報を効率的かつ効果的に利用可能とすることが考えられる。

Pick and Lap 手法による実空間コンテンツの仮想的融合による情報システムを実現するためには、主に以下に示す技術的課題が存在する。

- (1) 携帯端末のカメラによる実空間コンテンツの認識方法
- (2) 認識された実空間コンテンツの意味内容（メタデータ）の取得方法
- (3) 実空間コンテンツの融合方法
- (4) 携帯端末上のアプリケーション・システムとしての実装方法

本論文では、上記の技術的課題のうち、(3)の「実空間コンテンツの融合方法」に対象を限定し、実空間コンテンツの融合モデルを提案する。すなわち、携帯電話のカメラによる認識が正確に行われ、抽出された実空間コンテンツに対してあらかじめメタデータが与えられていることを前提とする。また、提案手法を、処理能力と記憶容量に制限がある携帯端末で実行するための効果的な実装方法については対象としない。

本論文では、実空間コンテンツに対する Pick and Lap 操作によって、実現される実空間コンテンツの仮想融合モデルを提案する。仮想融合モデルは、融合表現モデルと、融合表現を実現するための演算から構成される。本論文では、2つの実空間コンテンツを重ね合わせた際に、どのような融合表現を行うかを決定するための形式として、アイコン化モデル、クリッピング・モデル、アイコン・フィールド・モデルの3種類の融合表現モデルを提案する。さらに、上記の融合表現モデルに基づいた融合処理の論理モデルとして、ビジュアル・ジョイン演算を提案する。

ビジュアル・ジョイン演算は、メタデータが付与された2次元画像を対象に、複数の2次元画像から、それらの意味的な関係性に基づいた融合を行い、結果をメタデータ付

きの2次元画像として導出する。ビジュアル・ジョイン演算の中で、画像の融合は、アフィン変換を用いて定義され、変換に必要なパラメータは、画像のメタデータから導出された意味的な関係に基づいて自動的に決定される。

これまで、メタデータ付きの画像を対象として、それらの意味内容に基づいて合成するための一般的なモデルは提案されていない。文字、数値データに関しては、代表的な論理データモデルであるリレーショナル・データモデルが提供する、結合演算、選択演算、射影演算等の演算は、1つまたは複数の表から別の表を生成する演算であるために、演算を組み合わせることによって、ユーザの様々な要求に応じた表を生成するための一般的な枠組みを提供している。本論文で提案するビジュアル・ジョイン演算は、リレーショナル・データモデルの結合演算が複数の表を合成して新しい表を構成するのと同様に、メタデータが付与された複数の2次元画像を、メタデータが表す意味的な内容に基づいて合成して新しいメタデータ付きの2次元画像を構成可能である。本論文で提案する実空間コンテンツの融合表現モデルとそれに基づくビジュアル・ジョイン演算は、ユーザの目的に応じた意味内容に基づいた様々な種類の実空間コンテンツを仮想的に合成するための一般的な枠組みを提供する。

本論文の構成は以下のとおりである。2章では関連研究を示す。3章ではPick and Lap手法の概要と、我々が開発した実空間コンテンツの融合表現モデルについて説明する。4章では我々の融合表現モデルに基づく融合を実現するための論理モデルとして、ビジュアル・ジョイン演算を形式的に定義する。5章では、3種類の融合表現モデルを実装したプロトタイプ・システムを用いた被験者実験の結果を示し、提案手法の有効性を評価する。6章では、まとめと今後の課題について述べる。

2. 関連研究

近年、カメラ付き携帯端末上での拡張現実を利用した情報提供に関する研究や開発がさかんに行われている。セカイカメラ[6]では、携帯端末のカメラで現実世界を撮影すると、ディスプレイ上でエアタグと呼ばれる付加情報が、眼前の光景に対して重畳して描画される。一方、実空間透視ケータイ[7]は、現実世界を携帯端末のカメラを通して覗くと、ディスプレイ上に、眼前にどんな建物があるかが表示され、遠方にある建物等を擬似的に透視することができる。これらは、ユーザが携帯端末を利用している位置に基づいて、ユーザが見る風景にデジタル・コンテンツを重ねさせている。これらのシステムでは、GPS等によりユーザの現在位置を取得し、携帯端末のコンパス等でユーザの視線方向を取得する。これにより、ユーザの視界内に存在する空間に関して関連する情報を提供する。

一方、近年では、ユーザの視界というマクロなレベルで

はなく、ユーザが興味を持つオブジェクトの単位を単位として、拡張現実を利用した情報提示を行う研究が行われている。その中でも特に、実空間コンテンツを拡張することに着目した情報閲覧の研究が注目されている。携帯端末を利用して実空間コンテンツを拡張するための代表的なアプローチは、「のぞき穴 (peephole) ディスプレイ」[8]である。のぞき穴ディスプレイでは、携帯端末を実空間コンテンツに重畳した際に、重なった部分に関するデジタル・コンテンツを携帯端末上のディスプレイに表示する手法である。HOTPAPER[9]では、携帯端末をのぞき穴ディスプレイとして、紙媒体に印刷されたコンテンツへの、情報付加、取得を実現している。HOTPAPERでは、携帯端末のカメラで紙面を撮影した場所にユーザのコメントやビデオ等デジタル・コンテンツを付与することができる。すでに情報が付加されている部分を携帯電話で撮影すると、付与されたコンテンツが表示される。ここで、ユーザが携帯端末のカメラで撮影した部分を特定するために、実空間コンテンツの文字列の空間的分布の特徴を利用している。Rohsら[10]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、各格子内の領域をすべて画像として登録しておくことによって、携帯端末のカメラで撮影した地図の部分を同定可能としている。これを応用し、Wikeye[11]では、携帯端末のカメラを通して地図を覗いた際に、覗いた部分のランドマークに関する情報が記載されたWikipediaのページを提示する手法を提案している。さらに、Real World[12]では、ペーパークラフトによって作られた地球型の3Dオブジェクトを、携帯端末で覗くと、覗いた場所の地名や山脈の情報が提示されるシステムを提案している。これらの研究では、実空間上のコンテンツにデジタル・コンテンツを対応付け、それらを携帯電話のディスプレイで提示している。このため、あらかじめ対応付けられた、デジタル・コンテンツの閲覧に限定されている。一方、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合せの違いによって情報を変化させることで、個々のユーザに最適な情報を提供する手法を提案する。

3. 実空間コンテンツを対象とした仮想融合のアプローチ

3.1 Pick and Lap 操作

携帯端末を利用した実空間コンテンツの重ね合わせを実現するための操作であるPick and Lap操作について説明する。

Pick操作とは、カメラ付き携帯電話で実空間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．

たとえば，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．コンテンツは，様々な単位でとらえられる．旅行情報誌には 1 ページに複数の店舗や観光スポット等に関する情報が記載されているが，1 ページをコンテンツの単位ととらえることも，個々の店舗や観光スポットの紹介記事を単位ととらえることも，さらには旅行情報誌自体を 1 つの単位ととらえることもできる．Pick, Lap 操作が行われる対象は，コンテンツ単位であることから，统一的にモデルを構築するために対象コンテンツの単位の粒度を設定する．本研究では，個々の観光スポット，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 操作の対象となる単位とする．ここで，Pick 操作や Lap 操作の対象となるコンテンツの部分コンテンツ・セグメントと呼ぶ．このコンテンツ・セグメントに対するメタデータはシステム管理者側によって登録され，メタデータが登録されたコンテンツ・セグメントは，Pick 操作もしくは Lap 操作可能なコンテンツ対象となる．ユーザが携帯電話をかざした際に，コンテンツ・セグメント上に印を重ねて表示させることで，どのコンテンツ・セグメントが操作可能であるかを識別できるようにする．

本操作を実現するために，対象とするコンテンツ・セグメントを認識する必要がある．本研究では，コンテンツ・セグメントの認識対象として，コンテンツ・セグメントに含まれる 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．コンテンツ・セグメントと矩形画像の例を図 2 に示す．さらに，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，地図やカレンダー等は，コンテンツ全体を画像ととらえることができ，コンテンツ・セグメント自身が代表画像ととらえることができる．

Pick and Lap 操作が行われると，対象となったコンテンツ・セグメントの意味的な関係性に基づいて，融合処理を行う．このときの融合処理は，代表画像の移動，変形，合成によって行う．具体例を図 3 に示す．たとえば，旅行情報誌の店舗一覧に記載されている 1 つの店舗を表すコンテンツ・セグメントを Pick して地図のコンテンツ・セグメントに Lap すると，携帯端末の画面上では，地図の上に店舗の代表画像が縮小され重畳して表示される表現を考慮することができる．つまり，Pick コンテンツと Lap コンテンツが意味的に結び付き，それによって得られる相互の関係を表すために，代表画像による視覚表現をユーザに提示する．



図 2 コンテンツ・セグメントの例

Fig. 2 An example of contents segment.



図 3 Pick and Lap 操作に基づく仮想融合の実現例

Fig. 3 Example of virtual integration by Pick and Lap operation.

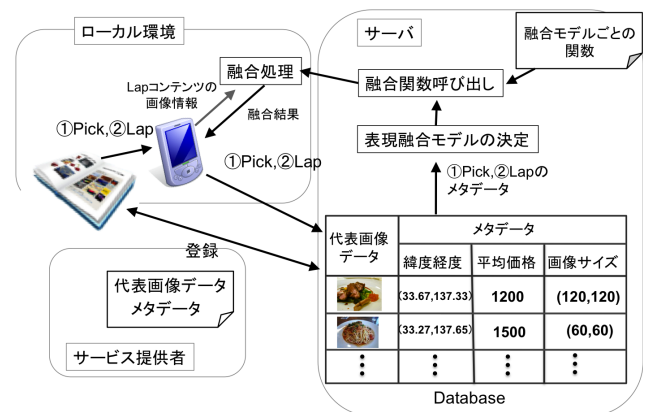


図 4 処理の流れ

Fig. 4 An overview of process flow.

3.2 処理の流れ

Pick and Lap 操作に基づく実空間コンテンツの仮想的な融合を行う処理の流れを図 4 に示す．システムは，サーバ・クライアント形式で実現される．クライアントは，エンドユーザが使用する携帯端末である．サーバには，システム管理者によって，融合可能な実空間コンテンツが有するそれぞれのコンテンツ・セグメントに対応する代表画像と，それに対応するメタデータが登録されているものとする．メタデータには，それぞれのコンテンツ・セグメントが表現する意味的な情報が含まれる．ここで，意味的な情報とは，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，座席数等の情報である．

本システムは，たとえば，出版社等が出版物のサービスの一環として運用することを想定している．そのような場

合、サービス提供者がシステム管理者としてサーバにメタデータを登録し、多数のエンドユーザがクライアント（携帯端末）でそれらを共有して利用することになる。

Pick 操作または Lap 操作を行う際には、携帯端末のカメラ機能を利用して、実空間コンテンツの代表画像を取得する。携帯端末によって取得した代表画像はサーバに送信され、サーバは対象とするコンテンツ・セグメントを検索し同定する。そして、同定されたコンテンツ・セグメントのメタデータを取得する。さらに、サーバで Pick コンテンツと Lap コンテンツの種類に基づいて表現融合モデルが決定され、融合モデルに適した融合関数が利用され、融合に必要なパラメータが計算される。このパラメータは、サーバからクライアントに送信され、クライアントの携帯端末上でコンテンツの仮想融合処理を行い、Lap コンテンツに対する拡張現実として利用者に結果を提示する。

撮影した代表画像からコンテンツ・セグメントを同定するのは重要な問題である。これに関しては、これまでも活発に研究が行われており、たとえば、中居ら [13] は携帯電話で撮影した画像に基づいて文書を高速かつ高精度に検索するための手法を提案している。本研究の目的は、コンテンツ・セグメントが正しく同定されたことを前提に、それらを利用してユーザに効果的な情報提供を行うための一般的な枠組みを開発することを目的とする。そのため、5 章で述べるプロトタイプ・システムでは、矩形領域に対する単純な画像類似検索手法を利用して実装を行った。また、矩形領域の抽出も画像に含まれる直線成分を利用した単純な手法を採用した [14]。

3.3 実空間コンテンツと融合表現モデル

コンテンツ・セグメントの代表画像は以下の 2 種類に分類できる。

- マップ型画像
- アイコン型画像

マップ型画像とは、意味的な広がりを持つ画像であり、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。ここで、画像の部分要素とは、画像上の、点、線、領域のことである。画像空間を、意味を持つ要素に分解する単位は、コンテンツのレイアウト構造に依存しており、点や線、領域の場合が考えられる。たとえば、マップ型画像の代表例である地図画像について考えてみる。地図画像上の点は、実空間上の地点（緯度経度）という概念を対応付けられる。また、地図画像上に描かれた線は、実空間上の道路等に対応付けられる。

マップ型画像は画像上の要素に対して他の概念を対応付ける。たとえば、地図上の要素として画像上の物理的な位置（これを物理座標と呼ぶ）を考えると、それぞれの物理座標には、現実空間における地点（緯度経度）を対応付けることができる。地図画像における物理座標と概念の対応付

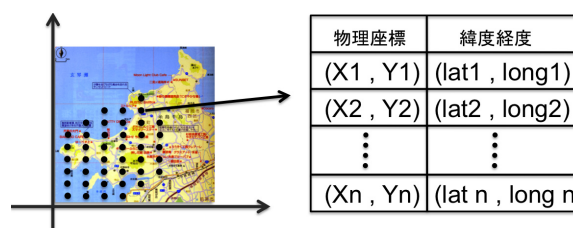


図 5 地図の物理座標と概念の対応付けの例

Fig. 5 Example of physical coordinates and assigned concepts of a map.

けの例を図 5 に示す。本論文では、簡単化のために、マップ型画像内の要素として、物理座標のみを対象とする。コンテンツ・セグメントの代表画像がマップ型画像である場合、物理座標と概念との対応付けは、そのコンテンツ・セグメントのメタデータとして与えられているものとする。その場合、様々な対応付けが考えられるが、基本的には、マップ型画像の任意の点に対して意味座標上の点を対応付けることと考える。本研究では、単純化のため、線や領域に対する対応付けについては考慮しない。それらについての対応は、今後の課題である。

本手法では、マップ型画像における意味座標と物理座標の対応付けは、メタデータ内でリレーションとして与えられることを前提とする。リレーションの与え方としては、(1) 関数型、(2) 列挙型の 2 種類を考えることができる。関数型でリレーションを与える場合には、意味座標と物理座標の関係を相互に変換可能な計算式として表現する。デフォルメが行われていない地図では、意味座標は緯度経度となるが、この場合、緯度経度と物理座標を変換する関数が数式として定義できるため、関数型でリレーションを与えることができるものとする。この場合、意味座標と物理座標は連続値であるため、リレーションは無限集合と考える。一方、意味座標と物理座標の対応関係を数式による関数として表現できないマップ画像も存在する。たとえば、デフォルメが行われている略地図や、カレンダー等は関数型でリレーションを定義することは困難である。本研究では、そのようなマップ型画像に対しても適用可能とするために、列挙型でのリレーションの定義も可能とする。列挙型でリレーションを与える場合には、対象となるマップ型画像上にサンプリング点を定め、それらのサンプリング点が表す意味座標の値を明示的に登録する。この場合、リレーションは有限集合となる。

列挙型の定義では、サンプリング点の位置、数、粒度等をどのように決定するかが重要である。それらは、対象とするマップ型画像の特徴や目的に応じて変化する。たとえば、デフォルメが行われた略地図においては、ランドマークが記されている点や交差点等の特徴点が、サンプリングの対象と考えられる。本手法では、マップ型画像の対応関係を、関数型の変換だけではなく、列挙型の定義が可能なり

レーションとして表現することにより、多様な種類のマップ型画像に対応可能な一般的な論理モデルを提供することを目的としている。列挙型によってレーションの定義を行う際の、サンプリングの対象点や数、粒度の最適値の決定方法については、今後、様々な種類のマップ型画像を対象として具体的に検討していく予定である。

本論文で物理座標とは実空間上での画像としての局所座標であり、ミリメートルで与えるものとする。そのため本システムにおいては、画像の物理座標上での大きさとして、画像の左上点 $(0, height)$ 、左下点 $(width, 0)$ を、あらかじめシステムに登録しておくことを前提とする。本手法では、意味座標に基づいて、携帯端末のディスプレイ上で画像の融合処理を行う。したがって、与えられた緯度経度等の意味座標に対応付けられる物理座標に対して、それが携帯端末のカメラで取得した画像において、どの位置に対応付けられるかを計算する必要がある。我々は、そのために、2次元アフィン変換を利用する。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り、式(1)に示す行列演算として表現できる。

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

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。我々は、画像処理により、物理座標系における原点 $(0, 0)$ 、左上点 $(0, height)$ 、左下点 $(width, 0)$ のそれぞれに対して、カメラで取得した画像上でのカメラ座標系での座標 (x_0, y_0) 、 (x_1, y_1) 、 (x_2, y_2) を取得する。ここで、 $width$ は対象とする代表画像の幅を表し、 $height$ は高さを表す。このとき、代表画像における物理座標系上の点をカメラ座標系上の点に変換するアフィン変換行列 T は方程式(2)を解くことにより取得可能である。

$$\begin{bmatrix} x_0 & x_1 & x_2 \\ y_0 & y_1 & y_2 \\ 1 & 1 & 1 \end{bmatrix} = T \begin{bmatrix} 0 & 0 & width \\ 0 & height & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad (2)$$

アイコン型画像とは、意味的な広がりを持たず、それ全体として他の概念に関連付けられる画像である。たとえば、旅行情報誌中の店舗の外観を表す画像は、その店舗を表すアイコン型画像であると考えられる。アイコン型画像に対応付けられる概念は単一であるとは限らない。たとえば、店舗 A で食べることができるカルボナーラの画像に関しては、パスタの1種としての「カルボナーラ」に対応付けられると同時に、「店舗 A」に対応付けることも可能である。本研究では、アイコン型画像は、画像に対応付ける対象に関するメタデータが与えられているものとする。たとえば、店舗を表す代表画像の場合、店舗の緯度経度情報や、平均価格、開店時間等のメタデータが与えられているものとする。

3.4 融合表現モデル

本研究では、複数のコンテンツ・セグメントの代表画像を、ユーザの目的に応じて仮想的に融合して提示することを目標とする。代表画像の融合の形式には様々な形式が考えられる。本研究では、代表画像の融合の形式を融合表現モデルと呼ぶ。

前節ではコンテンツ・セグメントの代表画像を2種類に分類したが、複数のコンテンツ・セグメントを重ね合わせる組合せと順番には、多数のパターンが存在する。本節では重ね合わせるコンテンツ・セグメントの組合せと順番のパターンに基づいて、3種類の融合表現モデルを導入する。

3.4.1 アイコン化モデル

アイコン化モデルは、Pick コンテンツの代表画像アイコン型で、Lap コンテンツの代表画像がマップ型である場合に利用される融合表現モデルである。この操作は、アイコン型画像をマップ型画像上に重ねる操作ととらえられるので、アイコン型画像が表す概念のマップ型画像上での位置をユーザに提示する。具体的には、Pick コンテンツの代表画像であるアイコン型画像を適切な大きさに縮小し、Lap コンテンツの代表画像であるマップ型画像上で対応する点に移動して表示する。

図6(a)にアイコン化モデルの概念図と融合の具体例を示す。この例では、旅行情報誌で店舗 A のコンテンツ・セグメントを Pick し、地図 B のコンテンツ・セグメントへ Lap した場合を示している。ここで、携帯端末の画面上では店舗 A の代表画像を、地図 B 上で店舗 A が存在する位置に表示する。つまり、ユーザは興味を持った店舗を、閲覧したい地図へ重ね合わせることで、興味を持った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あっても略地図等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、様々な地図でも店舗の位置を確認可能となる。さらに複数の店舗を Pick して地図へ Lap すると、複数の店舗の場所を1度に確認可能となり、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繰り返し参照する必要がない。

上記のような融合を行う際に問題となるのは、アイコン型画像のメタデータとして与えられている緯度経度、マップ型画像の物理座標とが、異なる座標系の座標であることである。本研究では、このように、異なる座標系における属性値を対応付け、融合操作を実現可能な一般的な演算として、4章においてビジュアル・ジョイン演算を導入する。ビジュアル・ジョイン演算では、画像に対するメタデータをリレーショナル・データモデルにおけるレーションとして与え、結合演算により異なる座標系における属性値を対応付け、画像の融合を行うためのアフィン変換のパラメータを自動的に決定可能である。具体的には、システム管理者が、マップ画像の物理座標と意味座標の対応関係、

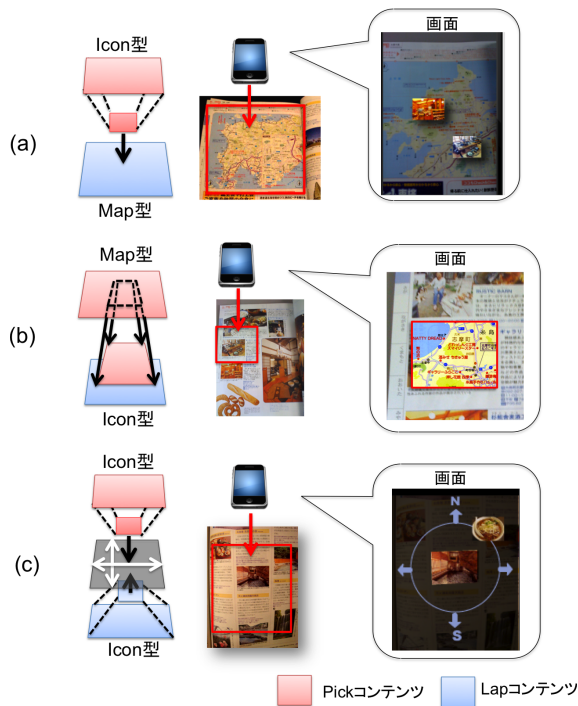


図 6 融合表現モデルの概念図と具体例：(a) アイコン化モデル，
(b) クリッピング・モデル，(c) アイコン・フィールド・モデル
Fig. 6 Overviews and examples of integration presentation models: (a) iconizing model, (b) clipping model, and (c) icon field model.

およびアイコン型画像の意味座標をメタデータとして与えておくことにより，エンドユーザは自動的にアイコン型画像とマップ型画像を融合可能となる。

3.4.2 クリッピング・モデル

クリッピング・モデルは，Pick コンテンツの代表画像がマップ型で，Lap コンテンツの代表画像がアイコン型である場合に適用される融合表現モデルである。この操作は，マップ型画像をアイコン型画像に重ねる操作ととらえられるので，マップ型画像の中でアイコン型画像が表す概念に関連する一定の領域のみを切り出（クリッピング）して，アイコン型画像の上に重畳して表示する。つまり，ユーザにとっては，興味のある Lap コンテンツの代表画像のみが，Pick コンテンツの代表画像の関連する部分に置き換わったように見える。

図 6(b) にクリッピング・モデルの概念図と融合の具体例を示す。この例では，旅行情報誌で，地図 B を Pick して店舗 A へ Lap した場合を示している。このときの融合表現では，ユーザが Pick した地図 B 上で店舗 A の場所がどこに存在するかを確認できるように，地図 B を店舗 A の場所を中心にセンタリングし，店舗 A のコンテンツ・セグメントの代表画像の領域に合わせてクリッピングして表示する。

一般に，旅行情報誌には複数の種類の地図が掲載されている。ユーザは単一の地図のみで場所を確認するとは限ら

ない。ユーザは目的によって異なる種類の地図を利用する場合がある。クリッピング・モデルでは，ユーザが，様々な地図の中からユーザの目的に適した地図を選んで Pick し，興味のある対象へ Lap することで，Pick した地図上においてそれぞれの対象がどのあたりに存在するかを確認可能となる。たとえば，駅付近の飲食店を探したいという要求を持ったユーザは，路線情報を強調して示された地図を Pick して，店舗に Lap していくと，店舗ごとに地図のクリッピング領域が変わることにより，各店舗の駅からの位置を確認可能となる。

3.4.3 アイコン・フィールド・モデル

アイコン・フィールド・モデルは，Pick コンテンツと Lap コンテンツの代表画像がどちらもアイコン型画像である場合に適用される融合表現モデルである。この操作は，複数のアイコン型画像を同一の空間に配置する操作ととらえられるので，ユーザは何らかの基準の下でそれらが表す概念を比較したいという意図を有すると考える。しかし，この操作でユーザが選択したコンテンツ・セグメントの代表画像にはマップ型画像が含まれていないため，アイコン型画像を配置するための基準が存在しない。そこで，ユーザが選択した複数のコンテンツ・セグメントに共通の属性を配置可能なマップ型画像を想定し，その上に複数のアイコン型画像を配置してユーザに提示する。ここで利用されるマップ画像を「仮想マップ画像」と呼ぶ。この際，仮想マップ画像の中心を，Lap コンテンツの代表画像の中心と一致させることにより，Lap コンテンツを中心として，Pick コンテンツの意味的な関係を視覚的に表示する。なお，ここで利用される仮想マップ画像は，あらかじめシステムが保持しているものとする。また，共通の属性を決定するためには，リレーショナルデータモデルにおける自然結合と同様に，属性名が同じ場合に共通の属性であると考えられる。属性名が異なる場合であっても共通の属性と考えられる場合があるが，そうした場合への対応は今後の課題である。

図 6(c) にアイコン・フィールド・モデルの概念図と融合の具体例を示す。この例は，店舗 A のコンテンツ・セグメントを，Pick して観光スポット C のコンテンツ・セグメントへ Lap した場合を表している。この例では，相互の位置関係を示すための仮想マップ画像を利用している。このマップを用いることで，観光スポット C を中心に，対象の方角を示す仮想的な空間が出現し，観光スポット C から見た店舗 A の方角が視覚的に分かりやすく表示される。具体的には，観光スポット C を中心にした円周上の位置で，観光スポットから見た方角を示す。この表現によって，たとえば，ユーザが特定の観光スポットに行き，さらにその観光スポット C 周辺の飲食店で食事したいと考えた場合，ユーザは興味を持った複数の店舗を Pick し，観光スポット C へ Lap すると，それぞれの店舗が観光スポットから，どの方角に位置するかを確認可能となる。旅行先で

食事をする店を決める状況で、実際に複数の店舗の様子を見て実際に入る店舗を決めたい場合、この仮想マップを利用すると、どちらの方角に向かえば少ない労力で多くの店舗を確認可能である。もし、具体的な位置を確認したい場合は、店舗を地図へ重ね合わせて得られるアイコン化モデルによって、地図上で店舗の場所を確認可能である。このように、ユーザは目的に合わせた組合せで、Pick and Lap 操作を行い、さらに従来の携帯端末上のサービスでは必要であった入力操作や画面切替えといった手間なしに、効率的に情報を取得することが可能となる。

ここで対象を比較するユーザは、対象の詳細な情報を取得したいのではなく、比較対象のうちどれが一番ユーザの要求を満たしているかということを知りたいと考えられる。たとえば、ユーザが店舗どうしを比較するのは、一番安い店舗や落ち着いた店舗、大人数でも利用できる店舗を探している場合が考えられる。このような場合、店舗どうしを重ね合わせ、それぞれの概念を定量的に示した仮想 Map 上に店舗画像が表示されることで、ユーザはどの店舗に行くか意思決定を行うことが可能となる。もし店舗の具体的な位置を確認したい場合は、店舗を地図へ重ね合わせて得られるアイコン・モデルによって、地図上で店舗の場所を確認することができる。このように、ユーザは目的に合わせた組合せで、重ね合わせを行い、さらに従来の携帯端末上のサービスでは必要であった入力操作や画面切替えといった手間なしに、効率的に情報を取得することが可能となる。

対象とするアイコン型画像が表す内容によって様々な仮想マップ画像が考えられる。本論文では、仮想マップ画像はシステム管理者によって設計され、システムに登録されることを前提としている。また、同一のアイコン型画像の組合せに対しても、適用可能な仮想マップ画像は複数存在する場合がある。そうした場合は、ユーザに選択可能な仮想マップ画像一覧が提示され、ユーザはそれらから適用する仮想マップ画像を選択する。別の仮想マップ画像の例として、価格の仮想マップ画像が考えられる。これは、ユーザが、安い店舗を探している場合等に有効である。この場合、店舗どうしを重ね合わせ、平均価格を表す仮想 Map 上に店舗画像が表示されることで、ユーザは複数の候補を比較して店舗を選択できる。

4. ビジュアル・ジョイン

融合表現モデルを利用して実空間コンテンツを仮想的に融合するための処理の一般的な形式化としてビジュアル・

ジョイン演算を提案する。ビジュアル・ジョイン演算は、メタデータと代表画像を持つ実空間コンテンツを対象に、メタデータの意味的な関連に基づいて代表画像の合成処理によって、コンテンツの融合を行う演算と定義する。

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 [15] では、データをリレーションとして構造化し、リレーションを構成するタプルの集合演算に加え、結合演算、射影演算、選択演算等の演算を利用可能である。これらの演算は入力と出力はリレーションであることから、複数の演算を組み合わせることで、ユーザはデータベースとして蓄積されている複数のリレーションから目的に応じて様々なリレーションを導出できる。

前章で提案した融合表現モデルは、複数の画像からそれらの意味内容に基づいて 1 つの画像を構成する方式を規定している。本論文で提案するビジュアル・ジョイン演算は、リレーショナル・データモデルの結合演算が複数のリレーションから 1 つのリレーションを構成するように、複数のコンテンツ・セグメントの代表画像から、それらの意味内容に基づいて、1 つの画像を構成する演算である。ビジュアル・ジョイン演算を利用することで、ユーザは複数のコンテンツ・セグメントに含まれる情報を融合した視覚的な表現を取得可能となる。

ビジュアル・ジョイン演算は、代表画像を有するコンテンツ・セグメントを対象とし、代表画像に関するメタデータが与えられていることを前提とする。具体的には、ビジュアル・ジョイン演算では、メタデータに関してリレーショナル・データモデルにおける結合演算を適応して、代表画像間の意味的な関連付けを行い、代表画像の移動、拡大・縮小によって関係性を視覚的に表現することで、コンテンツを仮想的に融合する。

4.1 形式化

ビジュアル・ジョイン演算の形式的な定義を示す前に、コンテンツ・セグメントを形式化する。対象とするコンテンツ・セグメントは代表画像とメタデータを有するものとし、コンテンツ・セグメント c の代表画像を $c.i$ 、メタデータを $c.m$ と表記する。本研究では、コンテンツ・セグメントのメタデータは、リレーショナル・データモデルにおけるリレーションとして表現されるものとする。また、代表画像 i の型を $i.type$ と表記する。型の値は「icon」または「map」のいずれかである。

いま、 c_1 、 c_2 がコンテンツ・セグメントであるとき、ビジュアル・ジョイン演算 VJ を以下のように定義する。

$$VJ(c_1, c_2) = \begin{cases} VJ^I(c_1, c_2) & (c_1.i.type = \text{icon} \text{ かつ } c_2.i.type = \text{map}) \\ VJ^C(c_1, c_2) & (c_1.i.type = \text{map} \text{ かつ } c_2.i.type = \text{icon}) \\ VJ^F(c_1, c_2) & (c_1.i.type = \text{icon} \text{ かつ } c_2.i.type = \text{icon}) \\ c_2 & (c_1.i.type = \text{map} \text{ かつ } c_2.i.type = \text{map}) \end{cases} \quad (3)$$

ここで、 VJ^I は融合表現モデルとしてアイコン化モデルを利用したビジュアル・ジョイン演算である。同様に、 VJ^C は融合表現モデルとしてクリッピング・モデルを利用したビジュアル・ジョイン演算であり、 VJ^F は融合表現モデルとしてアイコン・フィールド・モデルを利用したビジュアル・ジョイン演算である。つまり、演算の対象となるコンテンツ・セグメントの代表画像の型の組合せによって、異なる融合表現モデルに基づいた演算が適用される。これらに関しては、以下で定義する。なお、引数となる2つのコンテンツ・セグメントの代表画像がともにマップ型であった場合は、融合結果としてコンテンツ・セグメント c_2 を返す。これは、特別な融合処理を行わず、重ね合わされた対象となったコンテンツ・セグメントそのものを結果として返すことを意味する。

4.2 アイコン化モデルに基づくビジュアル・ジョイン

ビジュアル・ジョインはメタデータに基づく意味的な関連付けと意味的な関連に基づく画像の移動・変形と合成によって行われる。融合表現モデルとしてアイコン化モデルを利用した場合のビジュアル・ジョインを定義する。

今、コンテンツ・セグメント c_1, c_2 に対して、アイコン化モデルに基づくビジュアル・ジョイン演算 VJ^I を以下のように定義する。

$$VJ^I(c_1, c_2) = c_r \quad (4)$$

ここで、 c_r は以下のように定義されるコンテンツ・セグメントである。

$$c_r.m = VJ_m^I(c_1, c_2) \quad (5)$$

$$c_r.i = VJ_i^I(c_1, c_2) \quad (6)$$

ここで、 VJ_m^I は結果となるコンテンツ・セグメントのメタデータを構成する演算であり、 VJ_i^I は結果となるコンテンツ・セグメントの代表画像を構成する演算である。これらの演算は以下で定義する。

マップ型画像を代表画像とするコンテンツ・セグメントのメタデータは、物理座標を主キーとするリレーションで表現される。マップ型画像は、物理座標に対して他の概念が対応付けられている画像であり、対応付けられた概念は

リレーションの属性として表現される。一方、アイコン型画像を代表画像とするコンテンツ・セグメントのメタデータも、リレーションで表されているとする。アイコン画像のリレーションのタプル数は1である。

マップ型画像のメタデータとアイコン型画像のメタデータの意味的な関連を表すメタデータを構成する演算 VJ_m^I は以下のように形式的に定義する。

$$VJ_m^I(c_1, c_2) = OJ(c_1.m, c_2.m) \quad (7)$$

ここで、 OJ はリレーショナル・データモデルにおける外部結合を表す。ここで、外部結合としたのは、アイコン化モデルが適用されたマップ型画像に対しても、マップ型画像としての特性である物理座標と概念の対応付けを保持するためである。

リレーショナル・データモデルの結合演算は形式的には、以下のように定義できる [15]。

$$R \bowtie_F S = \{t * u | t \in R, u \in S \wedge P_F(t, u)\} \quad (8)$$

ここで、リレーション R と S はリレーションであり、 $t * u$ はタプル t と u を連結したタプルを表し、 $P_F(t, u)$ は t と u が結合条件 F を満足するとき真となる述語である。結合条件 F とは、 R の属性 A_i と S の属性 B_i の比較演算子 θ による比較条件 $A_i \theta B_i$ として定義される。マップ型画像の物理座標に対応付けられる概念は、連続値の場合と離散値の場合の2種類に大別できる。たとえば、地図の物理座標に対応付けられる概念である緯度経度は連続値であり、カレンダーの物理座標に対応付けられる概念である年月日は離散値である。マップ型画像の物理座標に対応付けられる概念が連続値である場合、物理座標と、それに対応付けられる概念の関係を、関数 f として表現できる場合がある。このとき、マップ型画像のメタデータを表すリレーションは無限集合となり、アイコン型画像のメタデータを表すリレーションはタプル数が1の有限集合となる。この場合でも、マップ型画像の連続値の属性と、アイコン型画像の連続値の属性が比較条件を構成するが、関数 f の逆関数により、結果となるリレーションを求められる。一方、物理座標に対応付けられる概念が連続値であっても、略地図等のように、物理座標と対応付けられる概念との対応付けを、関数として表現できず、対応関係がタプルの有限集合として表現される場合がある。このとき、アイコン画像のリレーションにおける結合条件の属性値と、マップ画像のリレーションにおける属性値が厳密に一致しない場合がある。このような場合に対応するために、この外部結合における結合条件は、等号ではなく、比較する2要素が最近傍であるという条件を使用する。等号の条件を満足するデータは、最近傍の条件を満足するため、連続値である場合にも対応可能である。なお、マップ型画像のリレーションが有限集合であるような場合には、最近傍要素の探索に kd

木 [16] や R 木 [17] を利用することや、マップ型画像のリレーションを、概念と対応付けられる矩形領域として表現し、結合対象となるアイコン画像の属性値との矩形内包含を行うことにより、効率的に処理できると考えられる。

次に、画像の合成演算について述べる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・縮小等の幾何変換を行い、適切な配置で、適切な順番で重ね合わせることで実現する。代表画像の幾何変換処理は、2次元アフィン変換によって表現する。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コンテンツ・セグメントを c_1, c_2 とするとき、演算結果となるコンテンツ・セグメントの代表画像を構成する演算 VJ_i^I を以下のように定義する。

$$VJ_i^I(c_1, c_2) = \text{Over}(\text{Conv}(c_1.i, T), c_2.i) \quad (9)$$

ここで、 Over は 2 個の 2 次元画像を引数としてとり、1 番目の画像を 2 番目の画像上に重畳させる関数である。また、 $\text{Conv}(im, T)$ は画像 im 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関数である。

上記の演算で使用する 2 次元アフィン変換行列 T は、対象となるコンテンツの意味的な関係に基づいて決定する。なお、簡単化のために、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式 (10) で表現できる。

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} t_x & 0 & \alpha \\ 0 & t_y & \beta \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (10)$$

ここで、 t_x, t_y はアイコン型画像の X 方向、Y 方向の縮小率であり、 α は X 方向の移動距離、 β は Y 方向の移動距離を表す。図 7 にこの変換によって行われる画像合成の概念図を示す。これらのパラメータは、以下のように求められる。

$$\begin{aligned} \alpha &= IJ(c_1.m, c_2.m).物理座標.x - c_1.i.w/2 \\ \beta &= IJ(c_1.m, c_2.m).物理座標.y - c_1.i.h/2 \\ t_x &= icon_w/c_1.i.w \\ t_y &= icon_h/c_1.i.h \end{aligned} \quad (11)$$

ここで、 IJ は引数となるリレーションの内部結合を取得する演算を表している。なお、この内部結合における結合条件は、等号ではなく、比較する 2 要素が最近傍であることである。さらに $c_1.i.w, c_1.i.h$ は、それぞれコンテンツ・セグメント c_1 の代表画像の幅および高さを表す。さらに、 $icon_w, icon_h$ は、それぞれマップ画像上でのアイコンの幅と高さを示す定数である。たとえば、図 8 に示すようなメタデータを有する店舗情報を表すコンテンツ・セグメント c_1 、および地図を表すコンテンツ・セグメント c_2 を重ね

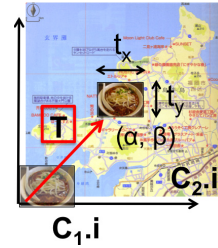


図 7 代表画像の幾何変換の例

Fig. 7 Example of transformations on a representative image.

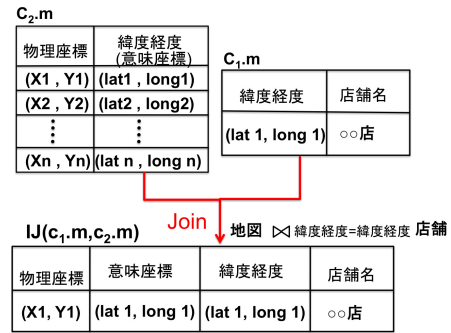


図 8 コンテンツ・セグメントのメタデータの内部結合の例

Fig. 8 Example of inner join of metadata of content segments.

合わせたとする。いま、店舗を地図へ重ね合わせると、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、店舗は地図上の物理座標 $(X1, Y1)$ と対応付けられ、アフィン変換のパラメータとして使用される。なお、図中の記号「 $\sim$ 」は最近傍要素であるときに真となる比較演算を表している。

4.3 クリッピング・モデルに基づくビジュアル・ジョイン

融合表現モデルとしてクリッピング・モデルを利用したビジュアル・ジョインについて述べる。

クリッピング・モデルによる融合処理は、アイコン化モデルによる融合処理と類似している。代表画像の変換処理としては、アイコン化モデルと同様に、Pick コンテンツの代表画像に対してアフィン変換を施し、Lap 画像の領域に重ね合わせる。アイコン化モデルとの違いは、Pick コンテンツの拡大縮小を行わず、Lap 画像の領域よりはみ出た部分は除去する点にある。

コンテンツ・セグメント c_1, c_2 に対して、クリッピング・モデルに基づくビジュアル・ジョイン演算 VJ^C を以下のように定義する。

$$VJ^C(c_1, c_2) = c_r \quad (12)$$

ここで、 c_r は以下のように定義されるコンテンツ・セグメントである。

$$c_r.m = VJ_m^C(c_1, c_2) \quad (13)$$

$$c_r.i = VJ_i^C(c_1, c_2) \quad (14)$$

ここで、 VJ_m^C は結果となるコンテンツ・セグメントのメタ

データを構成する演算であり、 VJ_i^C は結果となるコンテンツ・セグメントの代表画像を構成する演算である。これらの演算は以下で定義する。

メタデータを構成する演算 VJ_m^C は以下のように形式的に定義する。

$$VJ_m^C(c_1, c_2) = OJ(c_1.m, c_2.m) \quad (15)$$

ここで、 OJ はリレーショナル・データモデルにおける外部結合を表す。なお、この外部結合における結合条件は、等号ではなく、比較する2要素が最近傍であることである。

また、演算結果となるコンテンツ・セグメントの代表画像を構成する演算 VJ_i^C を以下のように定義する。

$$VJ_i^I(c_1, c_2) = \text{Clip}(\text{Over}(\text{Conv}(c_1.i, T), c_2.i), c_2.i) \quad (16)$$

ここで、 Clip は2個の2次元画像を引数としてとり、1番目の画像を2番目の画像の大きさでクリッピングする関数である。また、2次元アフィン変換行列 T のパラメータは以下のように決定される。

$$\begin{aligned} \alpha &= -IJ(c_1.m, c_2.m).\text{物理座標}.x + c_1.i.w/2 \\ \beta &= -IJ(c_1.m, c_2.m).\text{物理座標}.y + c_1.i.h/2 \\ t_x &= 1 \\ t_y &= 1 \end{aligned} \quad (17)$$

4.4 アイコン・フィールド・モデルに基づくビジュアル・ジョイン

融合表現モデルとしてアイコン・フィールド・モデルを利用した際のビジュアル・ジョインについて説明する。

アイコン・フィールド・モデルが適用されるのは、PickコンテンツとLapコンテンツの代表画像がともにアイコン型である場合である。この場合、アイコン型画像を配置するマップ型画像が指定されていないため、対象となる複数のコンテンツ・セグメントを対応付けられる意味的な広がりを持ったマップ型画像（仮想マップ画像）を利用して、アイコン型画像を配置することによって、コンテンツ・セグメントの意味的な関係を視覚的に提示する。仮想マップは、他のマップ型画像と同様に、物理座標に異なる概念が対応付けられている画像である。システムにはあらかじめいくつかの汎用的な仮想マップが登録されており、PickコンテンツとLapコンテンツのメタデータの共通属性によって適切な仮想マップが選択され、マッピングされる。

コンテンツ・セグメント c_1, c_2 に対して、アイコン・フィールド・モデルに基づくビジュアル・ジョイン演算 VJ^F を、アイコン化モデルに基づくビジュアル・ジョイン演算 VJ^I およびクリッピング・モデルに基づくビジュアル・ジョイン演算 VJ^C を用いて、以下のように定義する。

$$VJ^F(c_1, c_2)$$

$$= VJ^I(c_1, VJ^I(c_2, VJ^C(VM(c_1, c_2, M), c_2))) \quad (18)$$

ここで、 $VM(c_1, c_2, M)$ はマップ画像集合 M からコンテンツ・セグメント c_1, c_2 に対して、適切な仮想マップを返す関数である。

上記のアイコン・フィールド・モデルに基づくビジュアル・ジョインの定義の、直感的な説明は以下のとおりである。まず、関数 VM によって求められた仮想マップをPickコンテンツとして、 c_2 をLapコンテンツとしてクリッピング・モデルに基づくビジュアル・ジョイン演算を適用する。この操作により、 c_2 の代表画像の領域に仮想マップの画像が表示される。さらに、この結果をLapコンテンツとして、 c_2 をPickコンテンツとして、アイコン化モデルに基づくビジュアル・ジョイン演算を適用することにより、仮想マップの中心に c_2 の代表画像がアイコン化されて表示される。この結果をLapコンテンツとして、 c_1 をPickコンテンツとしてアイコン化モデルに基づくビジュアル・ジョイン演算を適用することにより、マップ画像上の適切な位置に c_1 の代表画像が表示される。これにより、仮想マップが提示する意味空間上で、 c_1 と c_2 の関連性を視覚的に確認できる。

アイコン・フィールド・モデルに基づくビジュアル・ジョインでは、適切な仮想マップを導出することが重要である。そのためには、利用者にとって理解しやすい仮想マップ画像のデザインと、上記の定義における関数 VM の実現方法が重要となる。

仮想マップとしては、多くのコンテンツ・セグメントが対応付け可能な広範囲な意味空間を表すマップ型画像が適していると思われる。また、相対的な意味関係を表示するのに適した汎用的な仮想マップを考えることもできる。図9は、方角を表すための汎用的な仮想マップを用いて店舗Aと店舗Bの位置関係を表示した例である。

仮想マップ画像の他の例としては、店舗の平均価格の関係を表示する仮想マップ画像が考えられる。図10では、店舗B、Cを店舗AへLapしたとき、店舗の平均価格を表す仮想マップ画像を利用した表現を示されている。この図では、店舗Aを中心とした平均価格を示す直線上に、店

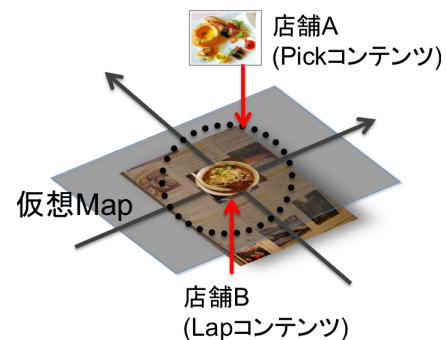


図9 方角を表す仮想マップの例

Fig. 9 An example of virtual map for representing direction.

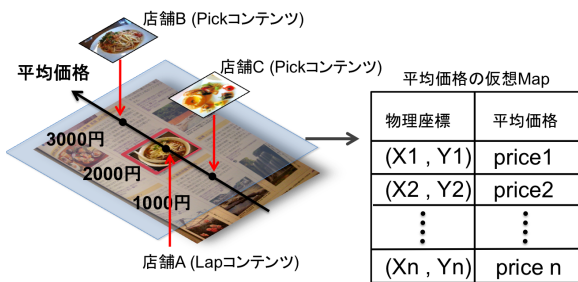


図 10 平均価格を示す仮想 Map

Fig. 10 An example of virtual map for average price.

舗 B, C のアイコン画像が配置され、店舗の価格帯を直感的に分かりやすく提示可能である。

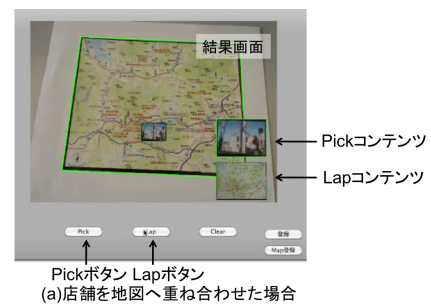
仮想マップ集合 M から、コンテンツ・セグメント c_1 , c_2 に適した仮想マップを返す関数 $VM(c_1, c_2, M)$ は、仮想マップ集合に定義されている仮想マップのうち、そのメタデータが c_1 , c_2 の両方のメタデータと共通の属性を持つものが選択される。候補が複数存在する際には、候補の中からユーザが手動で対象とする仮想マップを選択することを前提とする。コンテンツの重ね合わせる順番や組合せから文脈を考慮することで、ユーザの意図に合った仮想マップを自動的に選択することは今後の課題の 1 つである。

5. 評価

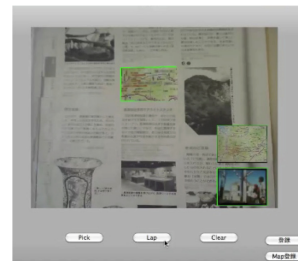
提案手法の有効性を評価するために、プロトタイプ・システムの実装を行い、プロトタイプ・システムを利用して評価実験を行った。

プロトタイプ・システムは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、携帯端末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、コンテンツの認識手法を開発することが目的ではないため、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。本研究では、簡単化のために、コンテンツ・セグメントが 1 つの代表画像を有する状況を対象とし、画像を認識することがコンテンツ・セグメントを認識することになる。コンテンツ・セグメントが複数の画像を含むような場合は、コンテンツ・セグメントが画像を含まないような状況への対応は今後の課題である。

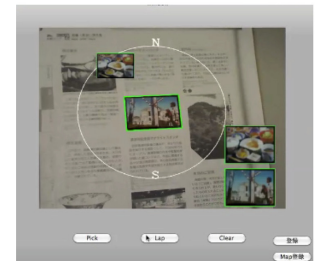
本プロトタイプ・システムで、3 種類の融合表現モデルを適用した際の実行例を図 11 に示す。ウインドウ下部にある Pick ボタンをマウスでクリックすることにより、コンテンツ・セグメントを Pick できる。また、Lap ボタンを押すと、Pick コンテンツを Lap できる。画面右には、その時点での、Pick コンテンツと Lap コンテンツの代表画像が表示される。カメラが矩形を認識している間は、矩形に対して黄緑色の枠が表示される。図 11 の (a), (b), (c) は



(a) 店舗を地図へ重ね合わせた場合



(b) 地図を店舗へ重ね合わせた場合



(c) 観光スポットを店舗へ重ね合わせた場合

図 11 プロトタイプ・システムを利用したビジュアル・ジョインの実行例

Fig. 11 Examples of visual join on the prototype system.

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。

5.1 実験

プロトタイプ・システムを用いて、本手法の有効性を評価するために、被験者実験を行った。本論文で提案するビジュアル・ジョイン演算は、我々が提案する 3 種類の融合モデルに基づいて、利用者に提示する表現を出力する。したがって、提案手法の有用性を示すためには、本システムの出力結果が利用者にとって自然な表現であるかが重要であると考えた。また、この演算を行うための重ね合わせる操作が理解しやすいことが重要であると考えた。

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、すべ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。

5.2 結果と考察

結果を図 12, 図 13, 図 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

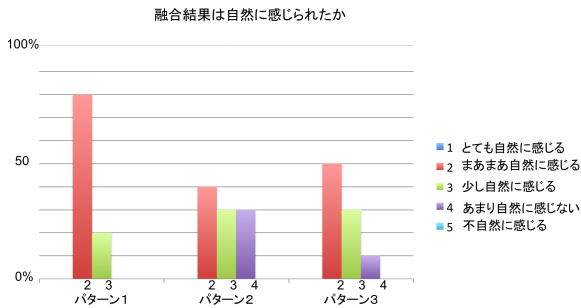


図 12 各パターンで得られた結果に対する印象の結果

Fig. 12 The results of the subjective evaluations for each pattern.

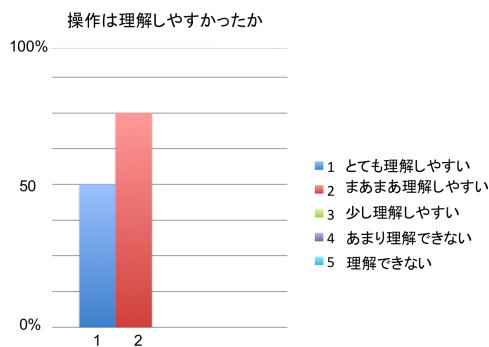


図 13 Pick and Lap 操作の理解しやすさに関する結果

Fig. 13 The result of the easiness for Pick and Lap operation.

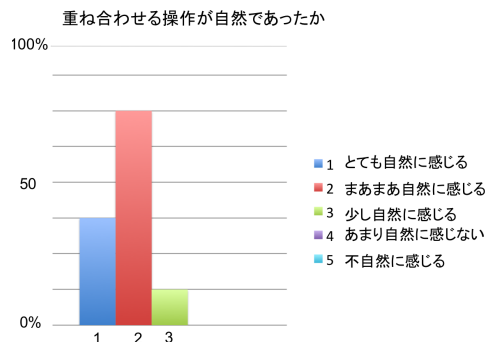


図 14 コンテンツの重ね合わせ操作への自然さに関する結果

Fig. 14 The result of the acceptability for overlapping operation of contents.

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、1度見たら分かるけれど、初めて見たとき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

くになってしまう」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つといえる。しかし、表示の問題以外では、「1回の操作で必要な情報が得られて嬉しい」「見たですぐに分かる」といった、融合表現から得られる情報に関しては肯定的な意見も多く得られた。

また全体を通した操作性の理解度に関する評価では、ほとんどの被験者が1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、重ね合わせる操作に対しても自然に感じられると答えた被験者が多かったことから、本手法は、表示に関して改善の余地はあるが、実空間上での複数のコンテンツ閲覧においては有効であることが示された。

6. まとめと今後の課題

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として Pick and Lap 手法を提案し、この操作によって実空間コンテンツを仮想的に融合する演算としビジュアル・ジョインを提案した。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、有効性の検証を行った。

今後の課題を以下に示す。

本研究のプロトタイプでは、Pick 操作から Lap 操作が行われるまでの時間は短時間であったが、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合せに幅が広がり、利用シーンの可能性も広がると考えている。

本論文では、Pick コンテンツと Lap コンテンツはともに実空間コンテンツであることを前提としてきたが、デジタル・コンテンツと実空間コンテンツの融合、デジタル・コンテンツどうしの融合にも応用できると考えられるため、デジタル・コンテンツどうしおよびデジタル・コンテンツと実空間コンテンツの融合に関しても考えていきたい。

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、様々な組合せが考えうる中で、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あげられる。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ている。

今回実装したプロトタイプでは、表示されたマップ型画像に対する対話的な操作を提供していない。たとえば、クリッピング・モデルを利用して表示された地図に対して、

中心にマークを表示し、対話的に地図の拡大・縮小、表示領域のスクロール等の対話的な操作を提供することで、ユーザの利便性が向上すると考えられる。今後、このようなコンテンツの融合した結果に対する対話的な操作を導入することを予定している。

提案手法を実システムに適用する際には、メタデータをどのように効率的に設定するかが重要となる。今後、コンテンツ・セグメントのデータから代表画像のメタデータを自動的に取得する、もしくはユーザのメタデータの作成を支援する機構を開発することを検討する予定である。

参考文献

- [1] GMAP, available from <http://gmap.jp/gmap/>.
- [2] ALPSLAB, available from <http://www.alpslab.jp/>.
- [3] ゼンリン いつも NAVI, 入手先 <http://www.zenrin-datacom.net/mobile/>.
- [4] Azuma, R.T.: A Survey of Augmented Reality, *Presence*, Vol.6, No.4, pp.355-385 (1997).
- [5] 武田十季, 鶴野玲治, 牛尼剛聡: Pick and Lap: カメラ付き携帯電話を用いた実空間コンテンツの重ね合わせによる効果的な情報閲覧, 第2回データ工学と情報マネジメントに関するフォーラム, A6-2 (2010).
- [6] セカイカメラ, 入手先 <http://sekaicamera.com/>.
- [7] 実空間透視ケータイ, 入手先 <http://kazasu.mobi/>.
- [8] Fitzmaurice, G.W.: Situated information spaces and spatially aware palmtop computers, *Comm. ACM*, Vol.36, No.7, pp.39-49 (1993).
- [9] Erol, B., Graham, J., Antunez, E. and Hull, J.J.: HOT-PAPER demonstration: multimedia interaction with paper using mobile phones, *Proc. 16th ACM international conference on Multimedia*, pp.983-984 (2008).
- [10] Rohs, M., Schoning, J., Kruger, A. and Hecht, B.: Towards real-time markerless tracking of magic lenses on paper maps, *Proc. 5th Intl. Conference on Pervasive Computing*, pp.69-72 (2007).
- [11] Hecht, B., Rohs, M., Schoning, J. and Kruger, A.: Wikiye - using magic lenses to explore georeferenced Wikipedia content, *Proc. 3rd International Workshop on PERMID* (2007).
- [12] Schoning, J., Rohs, M. and Kruger, A.: Mobile Interaction with the 'Real World', *MobileHCI 2009*, No.106 (2009).
- [13] 中居友弘, 黄瀬浩一, 岩村雅一: 特徴点の局所的配置に基づくデジタルカメラを用いた高速文書画像検索, 電子情報通信学会論文誌 D, Vol.J89-D, Vol.9, pp.2045-2054 (2006).
- [14] 山下大二, 富松 潔, 金 大雄, 牛尼剛聡: 虫メガネメタファーに基づく携帯電話上でのコンテンツ閲覧インタフェース, 日本データベース学会論文誌, Vol.8, No.1, pp.56-70 (2009).
- [15] 北川博之: データベースシステム 情報系教科書シリーズ 第14巻, 昭晃堂 (1996).
- [16] Bentley, J.L.: Multidimensional binary search trees used for associative searching, *Comm. ACM*, Vol.18, No.9, pp.509-517 (1975).
- [17] Guttman, A.: R-Trees: A Dynamic Index Structure for Spatial Searching, *Proc. SIGMOD'84*, pp.47-57 (1984).



武田 十季 (正会員)

2011年九州大学大学院芸術工学府芸術工学専攻修士課程修了。同年日本電信電話(株)入社。現在 NTT サイバーソリューション研究所に勤務。



鶴野 玲治 (正会員)

1985年大阪府立大学卒業。1987年同大学大学院修了(情報科学)。1993年博士(工学)。1987年近畿大学助手、1992年九州芸術工科大学講師・助教授を経て、現在、九州大学大学院芸術工学研究院准教授。コンピュータグラフィックス, NPR, コンテンツ制作支援技術等の研究に従事。芸術科学会, 画像電子学会, ACM, EUROGRAPHICS等の各会員。



牛尼 剛聡 (正会員)

1999年名古屋大学大学院工学研究科情報工学専攻博士課程後期課程満了。博士(工学)。1999年九州芸術工科大学助手。2007年九州大学大学院芸術工学研究院助教。現在、九州大学大学院芸術工学研究院准教授。電子情報通信学会(シニア会員), ACM, IEEE-CS, 日本データベース学会, ヒューマンインタフェース学会各会員。

(担当編集委員 宮森 恒)