

スマートフォンを利用した実空間上での活動支援を 目的とするコンテンツ提示手法に関する研究

武田, 十季

<https://doi.org/10.15017/1807045>

出版情報：九州大学, 2016, 博士（芸術工学）, 課程博士
バージョン：
権利関係：全文ファイル公表済



スマートフォンを利用した実空間上での
活動支援を目的とするコンテンツ提示手法
に関する研究

武田十季

Toki Takeda

2017年3月

目次

第1章 序論	1
第2章 関連研究	10
2.1 膨大なコンテンツの活用に関する研究	10
2.1.1 コンテンツのアノテーション技術	10
2.1.2 重要なコンテンツの抽出技術	10
2.1.3 コンテンツへの付加情報技術	11
2.2 コンテキストに基づくコンテンツ提示に関する研究	12
2.2.1 Web 上での操作・閲覧情報コンテキストとしたコンテンツ提示	12
2.2.2 位置情報と時刻情報をコンテキストとしたコンテンツ提示	13
2.2.3 生体情報や人の動作をコンテキストとしたコンテンツ提示	14
第3章 コミュニケーションを支援するプライベートコンテンツ提示手法	16
3.1 対面コミュニケーションにおけるコンテンツの利用	16
3.2 対話のきっかけを与えるプライベートコンテンツ提示手法	18
3.2.1 感情と記憶想起の関係	20
3.2.2 感情モデルを用いた感情の定量化	21
3.2.3 予備実験	23
3.2.4 提案インタフェース	27
3.2.5 実験：感情状態とコミュニケーションのきっかけ (仮説 I、II)	29

3.2.6	実験結果：(仮説 I、II)	32
3.2.7	実験：伝えたい記録の発見 (仮説 III)	34
3.2.8	実験結果 (仮説 III)	35
3.2.9	まとめ	37
3.3	対話中の話題に適したプライベートコンテンツ提示手法	38
3.3.1	Photo Ranking	40
3.3.2	評価実験	43
3.3.3	まとめ	47
3.4	相互の所有コンテンツを考慮したプライベートコンテンツ提示手法	48
3.4.1	写真グラフの構成	49
3.4.2	Random Walk with Restart モデルの応用	49
3.4.3	動作例	52
3.4.4	プロトタイプシステム画面	56
3.4.5	まとめ	56
第 4 章	拡張現実を利用した実空間コンテンツにおける情報獲得支援手法	60
4.1	実空間コンテンツを対象とした仮想融合のアプローチ	63
4.1.1	Pick and Lap 操作	63
4.1.2	処理の流れ	65
4.1.3	実空間コンテンツと融合表現モデル	67
4.1.4	融合表現モデル	70
4.2	ビジュアル・ジョイン	75
4.2.1	形式化	76
4.2.2	アイコン化モデルに基づくビジュアル・ジョイン	77
4.2.3	クリッピング・モデルに基づくビジュアル・ジョイン	81
4.2.4	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	83

4.3 評価	86
4.3.1 実験	87
4.3.2 結果と考察	88
4.3.3 まとめ	90
第 5 章 結論	92
参考文献	95
業績一覧	104

第1章 序論

2007年にAppleがiPhoneを発表してから、10年間で、スマートフォンは爆発的に普及した。日本の総務省の平成28年版情報通信白書 [1]によれば、日本におけるスマートフォンの所有率は2015年末の時点で72.0%に及ぶ。スマートフォンは、多くの人間が日常的に携帯して利用する汎用的なコンピュータである。

1946年に世界で最初の汎用的なコンピュータであるENIACが発表された当時、コンピュータは非常に巨大でありかつ高価であったため、一台のコンピュータを複数のユーザが共有して利用する形式で利用されるのが一般的だった [2]。そうしたことから、コンピュータを個人的な活動を支援するために利用されることはなかった。

1980年代から徐々に、コンピュータの小型化と低価格化が進み、1990年代から一般家庭にパーソナルコンピュータが普及した。パーソナルコンピュータは、ユーザー人が専有して利用するコンピュータであるため、個人の活動を支援する目的で利用されることも増えてきた。パーソナルコンピュータの種類には、机の上に備え付けて使用することを前提とするデスクトップ型と、持ち運びして利用できるノート型（ラップトップ型）がある。ノート型のパーソナルコンピュータは持ち運びができるが、一般のユーザが日常生活で常に持ち運んで利用することは行われぬ。つまり、パーソナルコンピュータを利用する時間と場所は限られているため、コンピュータが支援するユーザの活動の中心的内容は、文書作成、表計算、画像や動画の編集等に限定されていた。

一方、1990年代に普及した携帯電話は、2000年代には、音声通信のみならず、インターネットを利用したデータ通信が可能となり、様々な目的に利用できるようになった。しかし、携帯電話は、入力がテンキーと十字カーソルであるために操作性が低く、ディスプレイが小さく解像度が低いため視認性が悪く、プロセッサの処理能力が低いため複雑な処理を行えな

いという理由から、支援可能なユーザの活動は限定されたものであった。具体的には、携帯電話が支援可能な活動は、日常生活の隙間時間でのメッセージ送受信や、スケジュール確認等が主であった。

スマートフォンは、一世代前のパーソナルコンピュータと同等の処理能力を持つCPUを搭載し、従来の携帯電話では困難であった複雑な処理を実行できる。また、クラウドコンピューティングの普及に伴い、様々な処理がネットワークサービスとして提供されるようになり、大規模な計算資源を必要とする処理は、インターネット上のリモートのサーバで行い、処理結果のみを転送して利用する形態の処理が一般的になった。さらに、クラウド上のストレージを利用することにより、スマートフォンから莫大な記憶容量を利用可能である。また、スマートフォンは、高精細かつ高性能のタッチスクリーンインタフェースを搭載しており、入力インタフェースと出力インタフェースが一体化することで、デバイス一面をディスプレイとして利用できるようになった。ユーザが携帯可能なデバイスのサイズには制限があるため、タッチスクリーンによる入力インタフェースと出力インタフェースの融合により、モバイルデバイスの入出力のユーザビリティが飛躍的に向上した。

こうした背景の下、スマートフォンを利用することで、従来では実現できなかった様々な日常的な活動に対して、高度な支援が実現できるようになった。例えば、Google Maps [3] は、スマートフォンを利用して人間の移動という活動を様々な側面から高度に支援する。例えば、Google Maps は単に地図を提供するだけでなく、周辺の店舗の情報や、目的地への交通案内や予測到着時間を提示する機能を提供している。しかし、現在、スマートフォンが支援しているのは、ユーザの活動の一部分であり、今後、更に多くの種類の活動において、更に高度な活動の支援が期待される。

本研究では、スマートフォンを利用したユーザの高度な活動支援を対象とする。コンピュータが人間の活動を支援するための重要な概念として、コンテキストアウェアネス [4] がある。一般的に、コンピュータの動作原理は、ユーザがキーボードやマウスなどの入力デバイスを利用して明示的に情報を入力し、コンピュータは入力内容に応じて処理を行い、出力結果を

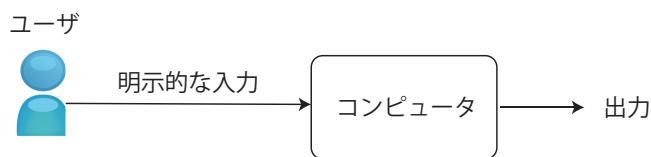


図 1: コンピュータへの明示的な入力と結果の出力

返すというものである (図 1)。しかし、日常生活でユーザの活動を支援するサービスを提供するためには、ユーザの明示的な入力を待つことなく、活動の中で適切なタイミングで、適切な処理を行うことが望ましい。

コンテキストウェアネスとは、コンピュータがユーザの状況を理解し、状況を考慮に入れた処理を行うアプローチである。コンテキストウェアネスを実現するためには、ユーザの明示的な操作だけではなく、ユーザの状態を反映するセンサ等のデータがシステムへ入力され、システムはそれらのデータを利用してユーザのコンテキストを推定し、コンテキストを考慮してよりユーザにとって満足度が高いと思われる処理を行う (図 2)。これまでにも、コンテキストウェアネスに関する研究は、1990 年代から取り組まれてきた [4]。例えば、Ryan ら [5] は、時刻、ユーザの位置情報や気温等をコンテキストとして、コンテキストが指定された条件になったら事前に関連付けられた情報をユーザにアラームとして提示する手法を提案している。また、Dey ら [6] は、場所や町の名前、天候の情報などをコンテキストとして、特定のコンテキストに応じた処理をユーザがあらかじめシステムに登録できる手法を示している。

いま、ユーザの日常的な活動の支援をコンテキストウェアネスの観点から捉えることを考える。一般的に、ユーザの日常的な活動は、その具体性に関して、異なるレベルから捉えることができる。活動の内容を具体性の高いレベルから捉えるほど、活動におけるユーザの目的や意図が明確に分かる。例えば、ユーザの「移動」という活動を考える。ユーザは何かの目的を持って「移動」という活動を行っているが、その目的は様々である。そこで、「移動」という活動をより具体性が高いレベルで捉えると、「歩行」または「飛行機 (による移動)」

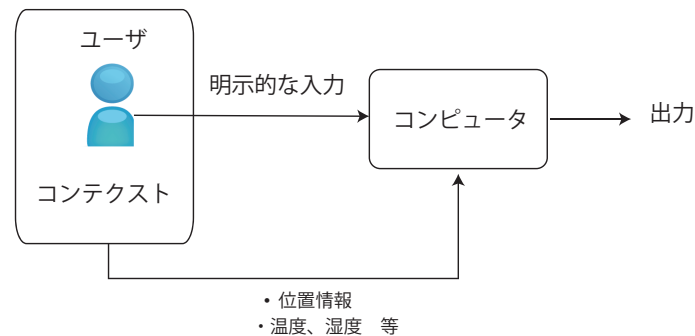


図 2: コンテキストアウェアネスにおける入力と結果の出力

というような活動として捉えることができる。「歩行」として捉えられた活動は、さらに具体性が高いレベルでは「散歩」や「通勤」などと捉えることができ、これらを捉えるレベルでは活動における目的がより明確になっている。活動を捉える具体性のレベルとその活動を捉える概念の関係を表す例を図 3 に示す。本研究では、高い具体性のレベルから捉えられた活動を「高次の活動」と呼ぶ。一方、低い具体性のレベルから捉えられた活動を「低次の活動」と呼ぶ。

従来のコンテキストアウェアネスに基づいて提案されてきた活動支援においては、「高次の活動」を考慮していないものが多い。例えば、リマインダはコンテキストアウェアネスの概念に基づいた代表的なサービスである。一般的なカレンダーアプリケーションにおけるリマインダ機能では、スケジュールに指定された時刻になったら関連付けられた情報をユーザに提示する。しかし、ユーザにとって、任意の時刻で必ずしもリマインダ情報が欲しいとは限らない。例えば、プレゼンテーションの最中にリマインダが表示されて、ユーザが他人に知られたくない情報が提示されてしまうことがある。これは、リマインダが、「プレゼンテーション」というユーザの「高次の活動」を考慮していないためであると考えられる。「プレゼンテーション」という「高度な活動」には、ユーザが集中して聴衆に意見を伝えようとしている、という意図が含まれている。システムがこのユーザの意図を把握できれば、発表に集中したいと

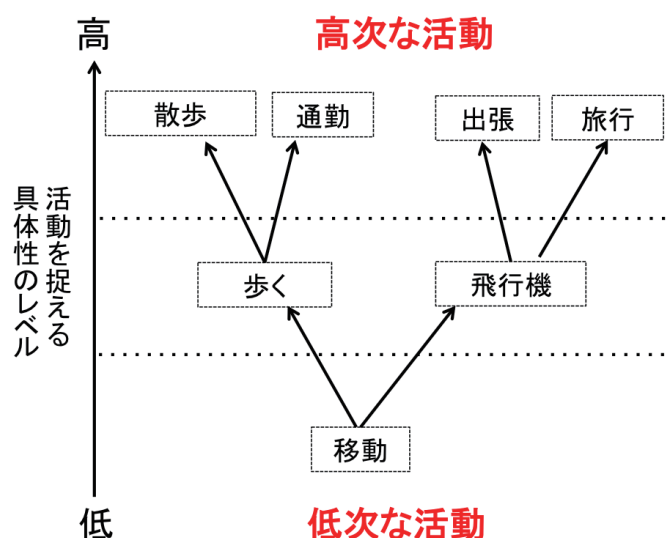


図 3: ユーザの活動の捉え方

いうユーザの要求を考慮したリマインドサービスを実現できるであろう。

高度な活動支援のためには、活動の目的や意図を明確に把握することが重要であるため、具体的性が高いレベルで捉えた「高次な活動」に基づいたサービスの設計が必要である。「高次な活動」を把握できるようになれば、人の要求を満たす情報提供が可能になる可能性が高い。スマートフォンは、ユーザが日常的に携帯して利用しており、ユーザは日常の中で様々な「高次な活動」を行っている。スマートフォンを利用して支援できる可能性のある「高次な活動」は多く存在する。また、スマートフォンには、カメラ、3軸加速度センサ、心拍計測センサ、センサが搭載されており、それらを利用して、ユーザの「高次な活動」を推定し、高度な支援が期待できる。したがって、スマートフォンは、ユーザの「高次な活動」を支援することに適したデバイスであると考えられる。本研究では、スマートフォンを利用してユーザの「高次な活動」を支援することを考える。

ここまで、スマートフォンを利用した活動支援の可能性について述べてきた。スマートフォンを利用した、実空間上の活動支援のためには、様々な方法が考えられる。例えば、ユーザ

のヘルスケアを対象としたウォーキングを支援することを考える。そのための、一つのアプローチとして、ユーザの心拍数や活動量に応じて、効果的な歩行のペースを提示するような機能を実現することである。もう一つのアプローチは、コンテンツの提示である。そのようなアプローチの例としては、ユーザの歩行スピードにあった楽曲を自動的に選曲して再生する機能を考えることができる。本研究では、ユーザの活動支援のためにコンテンツを利用するアプローチを採用する。

日常空間には、様々な数多くのコンテンツが存在する。ユーザは、日常的な活動を行う際に、様々なコンテンツを利用している。例えば、地図は代表的なコンテンツの一つであり、ユーザは地図を見ることで目的地の場所を確認できる。一般的に、コンテンツは、パブリックなコンテンツとプライベートなコンテンツに大別できる。パブリックなコンテンツとは、一般に公開されており不特定他者の利用を前提とするコンテンツである。パブリックなコンテンツの例として、ポスター、看板、情報誌、Web ページなどがあげられる。一方、プライベートなコンテンツとは、ユーザが個人として所有し、本人やその家族や知人などの、特定の他者の利用を前提とするコンテンツである。プライベートなコンテンツの代表例として、日記、スナップ写真等がある。

スマートフォンはデジタルコンテンツを扱うのに適した様々な特徴を有している。スマートフォンは高精細カメラを有しているものがあり、日常の中で写真や動画等の高品質なプライベートコンテンツを撮影できる。また、スマートフォンは高精細なディスプレイを有しており、撮影されたデジタルコンテンツを表示可能である。撮影したプライベートコンテンツが大量になったとしても、クラウドサービスを使えば、スマートフォンの内部メモリだけでなく、クラウド上のストレージ上にコンテンツを蓄積できる。ユーザは、必要なときにストレージにアクセスできるため、スマートフォン自体の記憶容量に制限されず、膨大なコンテンツを扱うことができる。一方、スマートフォンは、Wi-Fi や Bluetooth 等様々な通信機能を有しているため、ユーザが置かれた状況において適切な通信方法で Web ページなどの膨大なパブリックコンテンツにアクセスし、必要な情報を取得できる。さらに、スマートフォン

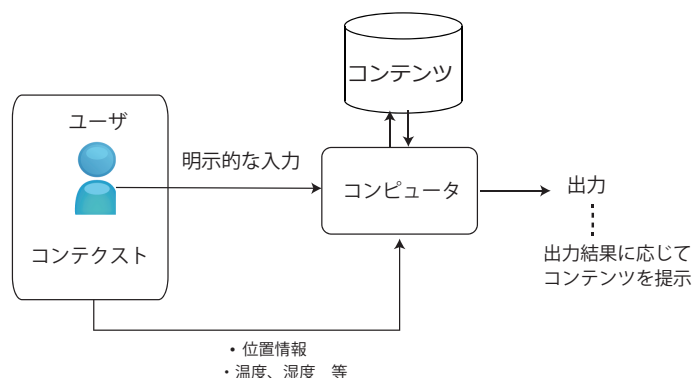


図 4: コンテキストに基づくコンテンツの提示

には、拡張現実 (AR) 技術 [7] を使えば、実空間上で物理的な媒体として存在しているコンテンツをより有効に利用できる可能性がある。以上のことから、スマートフォンは、コンテンツの利用に適したデバイスであると考えられる。

上記のように、スマートフォンは、ユーザが日常的に携帯する高性能のコンピュータであると同時に、コンテンツを扱うために有用な機能を提供するデバイスである。そこで、本研究では、スマートフォンを利用して、実空間上でユーザが行う「高次な活動」を対象に、活動の内容や目的に合致した形式でコンテンツを提示する手法を提案し、その有効性を検証することを目的とする。本研究で対象とするシステムの概要を図 4 に示す。

本研究で対象とする「高次な活動」には、多くの種類が存在する。本研究では、実空間での活動支援に対して、従来の研究で十分に組み込まれていないと考えられる、以下の 2 つの側面に取り組む。

- 対話コミュニケーションのためのプライベートコンテンツの活用の支援
- 実空間コンテンツを利用した情報取得のための拡張現実によるパーソナライズ

第一の側面は、プライベートなコンテンツを利用した活動の支援である。近年、スマートフォンのカメラを利用して、日常的に写真が撮影されるようになった。それに伴い、ユーザが

利用できるプライベートなコンテンツの数が膨大になっている。ユーザが利用できるコンテンツ数が増大するにつれ、ユーザがコンテンツの役割を十分に活かせない場合がある。膨大なコンテンツの中から、コンテンツを探す際に、目的のコンテンツを探し出すことは困難である。特に、人との対話中など、時間に制約がある場合は、コンテンツの探索はより困難になる。一般的に、ユーザがコンテンツを十分に活用するためには、コンテンツへのアクセシビリティや情報の獲得しやすさは非常に重要である。そこで、本研究では、プライベートなコンテンツを利用する代表的な場面として、人との対面コミュニケーション活動を想定する。ユーザは、対面コミュニケーションにおいて、記録したプライベートなコンテンツを利用して体験を共有することがある。ここで、対面コミュニケーションという活動をより具体性の高いレベルで捉えると、新たな話題を提供する目的の活動と、現在の話題に関連して対話を繋ぐ目的の活動が存在する。本研究では円滑なコミュニケーションを支援するために、これらの「高次な活動」に着目してユーザの目的を考慮し、ユーザがコンテンツへ効率的にアクセスできるコンテンツ提示手法を提案し、有効性の評価を行う。

第二の側面は、実空間コンテンツを利用した情報取得のための拡張現実によるパーソナライズである。実空間コンテンツとは、実世界上で物理的な媒体に印刷されたコンテンツである。代表的な実空間コンテンツには広告やポスター、情報誌がある。これらは、多くの人が利用するパブリックコンテンツである。実空間コンテンツは、様々な人が様々な活動のために利用されることが想定されることが多いため、ユーザにとって最適な表現になっているわけではない。実空間コンテンツはデジタルコンテンツのように可変でないため、ユーザが実空間コンテンツから知識や情報を獲得する際に、コンテンツが持つ情報量が多いため、ユーザは得たい情報をすぐに見つけられないことがある。また、情報獲得のためには、複数のコンテンツを相互に参照しなければならない場合もある。そうしたことから、実空間コンテンツをユーザの活動の目的に適した形で提示することが望ましい。ここで、実空間コンテンツを利用した情報獲得という活動をより具体性の高いレベルで捉えると、どの実空間コンテンツを、どんな順序で閲覧したか、という活動として見ることができる。それによって、ユー

ザの目的と要求する情報を推測することが可能となる。そこで本研究では、閲覧内容と閲覧順序によって、拡張現実を利用して実空間コンテンツをパーソナライズし、提示する手法を提案し、被験者実験により有効性を評価する。

本論文の構成は以下の通りである。2章では本研究に関連する研究について述べ、本研究の位置づけを明らかにする。3章では、対面コミュニケーションにおいて、写真を利用した体験の共有という活動のための、パーソナルコンテンツ提示手法について述べる。4章では、ユーザが実空間コンテンツに記載された情報を利用して意思決定を行う活動に着目し、スマートフォンを利用した拡張現実により、複数の実空間コンテンツをユーザの目的に応じて仮想的に融合する手法について述べる。5章では本論文の結論を述べる。

第2章 関連研究

2.1 膨大なコンテンツの活用に関する研究

これまでも膨大なコンテンツから、ユーザが重要なコンテンツを容易に見つけることができ、素早く必要な情報を取得できるように、コンテンツに注釈をつけるアノテーション技術や、重要なコンテンツを抽出する技術など、様々な観点から多くの研究が行われている。以下に、それらについて述べ、本研究の位置づけを明らかにする。

2.1.1 コンテンツのアノテーション技術

膨大なコンテンツに対して、全体を容易に把握できるように、ランキングや強調表現を利用して可視化を試みる研究は多く存在する [9] [10]。しかし、コンテンツを効率的に管理するためには、コンテンツから意味を抽出しタグを付与しておく必要がある。このようなアノテーションに関する研究は、従来から積極的に取り組まれてきた。Bao ら [11] は、スマートフォンやタブレットのカメラとマイクを利用して、映画を鑑賞しているユーザの表情と声を検出し、笑っているか、集中しているか等の状態を推定し、推定結果を映画コンテンツへタグ付けしている。Trohidis ら [12] は、楽曲コンテンツに対して、楽曲が持つリズム情報やメル周波数ケプストラム係数などを特徴量として、「幸福な」「落ち着いた」「悲しい」などの楽曲の感情表現を推定し、推定結果をタグとして付与している。このように、アノテーション技術が進み、様々なコンテンツにタグ情報を付与できるようになれば、コンテキストに最適なコンテンツとして推薦できるコンテンツの種類が広がる。

2.1.2 重要なコンテンツの抽出技術

コンテンツにアノテーションを付与するのみでなく、コンテンツの種類ごとに、重要な箇所を抽出する研究も活発に行われてきている。三浦ら [13] の研究では、料理の映像記録から、

調理動作及び料理や素材の状態を示す部分を重要なシーンとして、料理中のハイライト抽出を行っている。森山ら [14] は、ドラマ映像の心理的内容に基づいた映像要約の生成を提案している。これは、映像の物理的様相、例えば、カット編集による効果、BGM、効果音、などのトラック構造に現れる心理的演出のパターンと、映像が伝える心理的内容の対応関係を用いて、心理的に重要な箇所のみを切り出して要約映像を生成するものである。中村ら [15] は、楽曲動画の特徴に加え、動画に投稿されるコメントを用いて視聴者の盛り上がりを考慮し、サムネイル動画の自動生成を行っている。ユーザがコンテンツを直感的に閲覧できるように、類似した順番でコンテンツを提示するスライドショーの研究も行われている。Liuら [16] は、写真の画像情報を用いて、現在閲覧している写真と類似した写真を次に提示することで、ユーザが複数写真を連続的に閲覧しやすくする試みを行っている。

しかし、これらのアノテーション技術や抽出技術のいずれの研究においても、万人に有用なアノテーション付与やコンテンツの抽出・提示手法を目指しており、そのためにコンテンツの物理的構造や物理的特徴に注目している。他方、本研究では、ユーザが任意の活動コンテキストの元で、コンテンツを利用する場合は、各ユーザの活動コンテキストに応じて提示されるコンテンツは変化することが望ましいと考え、コンテキストに基づくコンテンツ提示の研究に取り組む。

2.1.3 コンテンツへの付加情報技術

コンテンツに対して、主観的な情報であるユーザの感情情報や、動作・行動情報も付与することで、過去に記録したコンテンツから、ユーザに過去を効果的に振り返られるよう支援する研究も行われている。”Affect Aura” [17] は、生体信号を利用した感情推定の技術 [18] を応用し、感情の推定結果を、日常生活においてユーザへ提示することによって、利点を得ようとする新たな試みを行っている。この研究で、は、4日間にわたりデスクワークにおける1時間ごとの感情推定を行い、5日後に、4日間の活動内容を記述したコンテンツに加え、推定された感情状態を色と形で表現したコンテンツとして、ユーザへ視覚的に提示することで、

感情状態の可視化はユーザの振り返りの活動において役に立つことを実験的に示している。しかし、“Affect Aura”では、推定された感情状態の可視化のみに限られており、膨大なコンテンツから重要なコンテンツを抽出する、という視点からは取り組まれていない。Ramadeviら [19] は、写真を介したコミュニケーション時の映像や音声などを用いて、対話時の様子をメタデータとして写真に付与し、新たな写真の活用方法を提案している。例えば、任意の写真に対して、誰がどれくらい閲覧したか、どの部分を指しながらどのような単語を発したかなどのデータを付与している。しかし、この研究では、コミュニケーション時のユーザの動作・行動情報を取得することを目的としており、取得した情報を利用してコミュニケーションという活動に適した写真を提示する手法には取り組んでいない。

2.2 コンテキストに基づくコンテンツ提示に関する研究

「高次の活動」を支援するコンテンツ提示のためには、現在のユーザの活動内容に適したコンテンツを推定する必要がある。以下、対象とするコンテキストごとに、従来サービスや関連研究について述べる。

2.2.1 Web 上での操作・閲覧情報コンテキストとしたコンテンツ提示

Web 上でユーザが閲覧した情報や、試聴した映像や楽曲に基づいて、ユーザに新たなコンテンツを推薦する研究やサービスは数多く存在する。これらは、閲覧情報やコンテンツの特徴と類似したコンテンツを推薦する内容ベースフィルタリングや、過去のユーザの評価や購入履歴に基づいて、ユーザ間の類似度を計算し、類似したユーザの評価に基づいてコンテンツを推薦する協調フィルタリング [20]、という代表的な手法が用いられている。協調フィルタリングを利用した有名なサービスに Amazon の商品推薦や、Youtube の楽曲推薦がある。Amazon では、商品を購入したユーザに対して、同じ商品を購入したユーザが他に購入した商品を推薦する [21]。Youtube においても、試聴した楽曲に合わせて、ユーザの好みに適した他の楽曲を推薦するサービスが提供されている [22]。これらは、ユーザが、現在どんな商

品を買ったか、もしくは、どんな楽曲を聴いたか、という活動をコンテキストとしてコンテンツを推薦している。さらに、Giordano ら [23] は、Web 閲覧中の、ユーザの視線を抽出し、ユーザの注目するコンテンツを検出することで、高精度にユーザの興味や嗜好を推定する手法を提案している。従来はマウスのクリックやキーボードからの文字入力のみでしか、現在のユーザの活動を捉えられなかったが、ユーザの視線抽出により、同じ画面上でも、画面の中のどの部分を閲覧しているか、という詳細な活動を捉えることを可能にしている。

2.2.2 位置情報と時刻情報をコンテキストとしたコンテンツ提示

Swingdo [24] は、スマートフォンの GPS 機能とスケジュール機能とを対応づけることで、ユーザが任意の場所に到達したときに、ユーザがあらかじめ登録していたタスクを提示するサービスがある。このサービスは、ユーザが任意の場所に存在するというコンテキストに基づいて、ユーザが実行しなければならないタスクを知らせることで、ユーザがタスクを忘れてしまうことを防いでいる。

位置情報に基づいて、ユーザが興味を持つと予想されるコンテンツを推薦することを目的とした、POI(Point of Interest) 推薦の研究は盛んに行われている。POI とは、ユーザが興味を持つ場所のことである。Quan ら [25] は、過去の POI の訪問履歴に加えて、現在の時間と場所をコンテキストとした POI 推薦を実現している。Gao ら [26] は、位置情報や時間に加えて、POI に対する他者の感情情報も取り入れて、POI 推薦を行っている。

場所や商品に対するユーザの理解を支援するために、拡張現実 (AR) 技術 [7] を利用して、場所や商品に対応するコンテンツをユーザへ提示するサービスも多数登場している。例えば、観光客に対して観光地の案内や、その土地の特徴を効果的に伝えることで、観光を支援するサービスがある [27] [28]。これらのサービスでは、観光客がスマートフォンのカメラ機能で実空間を覗くと、風景にデジタルコンテンツが重畳され、観光客が滞在している場所の詳細な情報を、直感的に知ることができる。Google glass では、拡張現実 AR メガネを装着すると、メガネを通して現実風景の中に様々なコンテンツが重畳して表示される。天気予報を見

たいとする音声に反応して空に天気予報のコンテンツが表示され、目的地までのナビゲーションと連動して、現在地点から目的地までの道順を現実風景の中に表示する。Google glass は、スマートフォンのように手で操作する必要がなく、常にユーザのコンテキストに応じたコンテンツを提供できることが可能となるため、様々な場面での利用が期待されている [29] [30]。Google glass では、コンテンツを提供するのみでなく、ユーザが手を使うことなく指示を入力することができるように、目の瞬きや頭の動きを検知する研究も取り組まれている [31]。

2.2.3 生体情報や人の動作をコンテキストとしたコンテンツ提示

近年は、センサの小型化により、ユーザの生体情報や動作情報を容易に計測することが可能となってきた。それに伴い、生体情報や動作情報をコンテキストとしたコンテンツの提示サービスや研究も取り組まれてきている。

Nike plus [32] のサービスでは、靴の加速度センサ情報を bluetooth でスマートフォンに送信し、走った距離やペース、消費カロリーを算出して、算出結果に応じたアドバイスを、音声やアニメーション表示によって、ユーザへフィードバックする。

Runtasite for docomo のトレーニングアプリでは、着るだけで生体情報を測れる電極”hitoe” [33] を利用して、運動中の心拍と加速度を計測する。hitoe 電極で計測された信号は hitoe トランスミッタを介して bluetooth によってスマートフォンのアプリへ送信される。ユーザは運動時に専用のウェアを着るだけで、心拍数に応じたユーザの身体状態「有酸素性」「脂肪燃焼」等をフィードバックとして得ることができ、身体状態に合わせた効率的なトレーニングを行える。

Liu ら [34] は、自律神経活動を反映する心拍の変動に着目し、楽曲試聴中のユーザの心拍を計測して、ユーザの副交感神経活動を優位にするような楽曲を推薦し、リラックス状態へ誘導する試みを行っている。

Pan ら [35] は、指先の抵抗値の変化として捉えられる GSR(Galvanic Skin Reflex : 皮膚電気反射) を計測し、電子書籍へのブックマークを試みている。GSR は、外的な刺激に対し

て大きく反応する性質があるため、この研究では、読書を中断したかどうかの検知のために GSR を利用している。中断後、再びユーザが読書を開始する際に、中断した時点のページを提示することで、ユーザが円滑に読書を再開することを支援する。これらの研究では、読書の再開や、リラックスするための楽曲試聴、という活動を限定して、その活動を支援するコンテンツ提示に試みており、我々のモチベーションと同じである。本研究では、プライベートコンテンツを利用した対面コミュニケーションと、実空間上のコンテンツを利用した情報獲得という、今までに取り組みられていない活動に着目し、それらの活動特有の課題を明らかにし、最適なコンテンツ提示手法を提案する。

第3章 コミュニケーションを支援するプライベートコンテンツ提示手法

本章では、人が日常生活において、コンテンツを利用して他者へ情報共有する活動を想定したコンテンツ提示手法について述べる。本研究では、特に、写真やテキストなど、ユーザー個人が記録したコンテンツを利用して、対面コミュニケーションを行う活動に着目する。

3.1 対面コミュニケーションにおけるコンテンツの利用

SNS(Social Networking Service) やブログなどの記録メディアや、撮影機能を有するデバイスの普及により、個々人の日常生活の活動や状況をライフログとして取得・蓄積する環境が整いつつある。特に、人が日常生活の中で体験する「できごと」、つまり、見たことや聞いたことなどをデジタル記録として残すことは容易になり、体験時に残される記録コンテンツは膨大化している。写真の場合、2014年の調査 [38] では、週に1回以上写真を撮るユーザは全体の22.8%である。月に1回以上写真を撮るユーザは55.5%である。我々が、20代の大学生38名に写真に関するアンケートを行ったところ、全体の76%の学生が、スマートフォンから自分の写真200枚以上を閲覧可能であった。

多くの人々がスマートフォンを持ち歩くようになり、写真や文書記録などのコンテンツ管理は、Webサービス上で行うことができるようになった。Webサービスであれば、記憶容量に制限があるスマートフォンからでも、膨大なコンテンツを利用できる。これにより、ユーザは、場所や時間帯を問わずいつでも記録コンテンツへアクセスできる。例えば、写真や文書記録はInstagram や Facebook 等のSNS にアップロードされ、友人と体験を共有するために利用される。対面コミュニケーションにおいても、これらのコンテンツは利用されることがある。

しかし、ユーザがコミュニケーションの最中に、膨大なコンテンツの中から必要なコンテンツを探し出すことは困難である。大学生 38 名を対象にしたアンケートでは、57.8%の学生が、コミュニケーション中に所有する写真から、目的の写真を探し出すことに苦労した経験があった。また、目的の写真を見つけられずに探すことを諦めた体験のある被験者は 36.8%であった。さらに、Nielsen [39] らの研究では、ユーザが興味を失わずに待てるコンピュータの反応時間は 10 秒とされている。このように、コンテンツを用いた対面コミュニケーションにおいて、ユーザが要求するコンテンツを短い時間で見つけ出せ、効率的にコンテンツを利用可能とすることは重要である。

ユーザ自身が所有するコンテンツを効率的に管理できるように、専用ソフトウェアの開発に関する研究やサービスがある。専用ソフトウェアでは、撮影日、撮影時、タグ等を利用して写真を並び替えられる。また、振り返りたい対象が明確な場合は、日時や場所をキーワードとして写真を探索できる [40] [41]。しかし、振り返りたい対象が明確でない場合は、そもそも検索のためのキーワードを想起することができない。また、対面コミュニケーションでは、コミュニケーション時のコンテキストに合ったコンテンツが必要とされる。コミュニケーション時の話題は経過時間に伴い動的に変化する。対面コミュニケーション時のコンテンツ利用の課題は、対話を始めるきっかけとなるコンテンツを提供すること、そして、変化し続ける話題に応じて適切なコンテンツを提供することである。

そこで、本研究では、対面コミュニケーションにおける以下の 2 つの活動を想定し、それぞれに対して適切なコンテンツ提示手法を提案する。

1. 新たな対話のきっかけを与えるコンテンツ提示手法
2. 現在の話題に関連する話題提供のためのコンテンツ提示手法

一つ目は、新たな話題を提供しようとしている活動に対して、話題のきっかけを与えるようなコンテンツを提示することで、対面コミュニケーションを支援することを目的とする。これによって、友人との偶然な再会時や久しぶりに話す祖父母との電話時など、限られた時間

内や特別な状況下において、伝えたいことを思い出せないまま、共有できずに終わることを防ぐ。二つ目は、対話が始始してから、現在の話題に対して、次の話題のきっかけになり得るようなコンテンツを提示することで、対面コミュニケーションを支援することを目的とする。本章では、上記2つの活動に対して、それぞれコンテンツ提示手法を提案し、被験者実験を通して、その効果について検証を行う。

3.2 対話のきっかけを与えるプライベートコンテンツ提示手法

親しい友人と街中で偶然に出会った時、あるいは、遠方にいる両親や祖父母と久しぶりに電話で話をした際に、自身あるいは相手の過去の体験から会話が盛り上がることもある。一方で、後から振り返ると伝えなかった体験を、会話の中では思い出すことができず、会話を終えた後に思い出して、伝えられなかったことを後悔する場合もある。人は、互いに自分に関する情報を伝え合う自己開示によって、互いに理解しあい親密な人間関係を構築している [42]。このことを踏まえると、会話を盛り上げる話題があるにも関わらず、思い出せないまま会話を終えてしまうことは、コミュニケーションの観点から残念なことである。友人との偶然な再会時や久しぶりに話す祖父母との電話時など、限られた時間内や特別な状況下での対面コミュニケーションという活動コンテキストにおいては、伝えたい体験を素早く思い出したい、というユーザの要求が含まれると考えられる。

伝えたい体験をユーザに素早く思い出させることについて考える。日常会話において、人は、話し手として、客観的なできごとに加えて、その時に自分がどう感じたかなどの感情や意図といった主観を伝えようとする。こうした傾向は日記や Blog や SNS 上で「できごと」を共有する際にもみられる。聞き手としてもそれを期待する場合が多い。つまり、人は「できごと」と「感情」を同時に記録し伝えることで、他人への共有における情報の価値を高めている。特に、親しい関係にある者同士のコミュニケーションほど、感情のような主観的な内容はより重要な情報となる。

例えば、人は「友人が久しぶりに遊びに来てくれて嬉しかった」というように、過去のエ

ピソードを語ることがある。感情状態の観点から考えると、これは感情状態が高まったときの体験といえる。また、「誰も誕生日を祝ってくれなくて寂しく感じていたが、後にサプライズパーティが開かれてとても驚き、そして嬉しかった」というような体験を語ることもある。これは、一見、前者と似ているが、後者は感情状態が哀から喜へと大きく変化した際のエピソードである。以上のことから私たちが伝えたい体験は、少なくとも2つ、すなわち感情状態が高まる時、また大きく変化する時の体験であると推察される。我々は、ユーザが伝えたい体験の記録を素早く見つけ出すために、感情状態とその変化に着目する。

本研究では、伝えたい体験を素早く思い出させてコミュニケーションのきっかけを与えるインタフェース、“Feelboard”を提案する(図5)。Feelboardでは、まず、ユーザが端末上で定期的に、体験をコンテンツとして記録する。そして、その体験の周辺で最も高まった感情状態と、高まる直前の感情状態の2状態を、感情価-覚醒度の2次元座標値として入力する。ユーザが記録から体験を想起しようとする際には、感情の高まり度合いとその変化の大きさに基づいてランク付けされた記録が端末画面上に提示される。上位ランクの記録ほど、想起しやすい体験であり、かつ、自己開示を動機づけさせる(伝えたい)体験であると期待され、また、優先的に提示されるため見つけやすい。体験の一部の記録提示でも、体験の記憶を再構築するきっかけとして十分であることから[40]、結果として、伝えたい体験の記憶が素早く再構築される。

本研究の主な貢献は以下の通りである。

1. 感情状態が大きく高まる状態へ変化した際の記録の提示は、伝えたい体験の想起をより容易にすることを明らかにする。
2. 感情状態の高まる状態への変化の大きさに基づいてランク付けて並べた記録の提示は、時間順に並べた提示よりも、伝えたい体験の記録を見つけ出すことにおいて優れていることを明らかにする。
3. 感情状態の変化に基づいてランク付けた記録の提示によって、コミュニケーションの

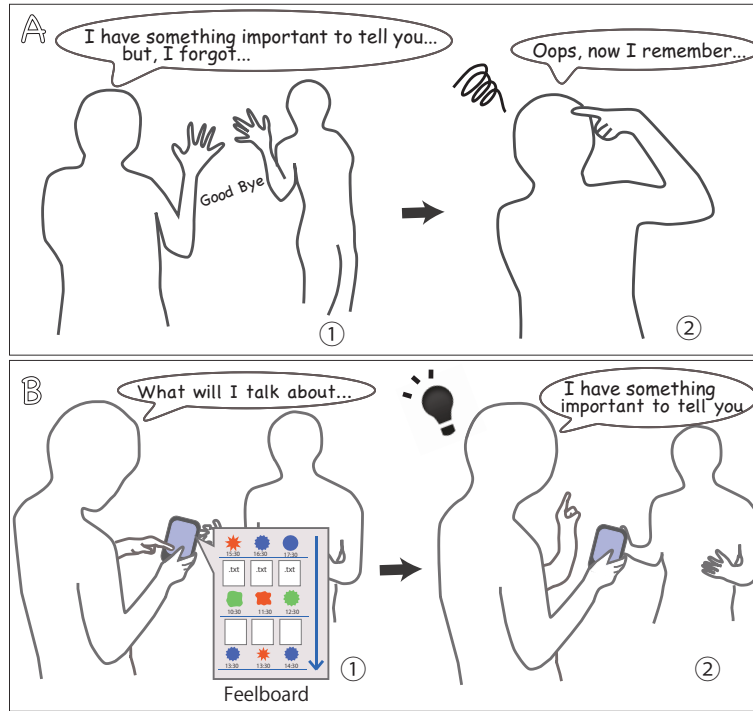


図 5: Feelboard の概要

きっかけを与えることを支援するインタフェース、Feelboard を提案する。

コミュニケーションを開始する際に、両ユーザが専用アプリを起動すると、ユーザの端末間で通信が開始し、システムは、コンテンツを利用したコミュニケーションが開始されるというコンテキストを検出できる。それに対応するユーザの要求の推測は、あらかじめシステムデザイナーによって設計される必要がある。本研究では、冒頭に述べた通り、ユーザの要求には、人に伝えたい体験を素早く思い出したい、という思いが含まれていると推測し、その要求に対して最適なコンテンツを提示するシステムを提案する。

3.2.1 感情と記憶想起の関係

心理学の分野では、感情と記憶に強い関連があることを示す研究が数多く存在する。例えば、覚醒度が高いと記憶に残りやすい覚醒度説 [43] や、ポジティブな感情を喚起したできご

とが記憶されやすいというポジティブ優位説 [44]、ネガティブな感情を喚起したできごとが記憶されやすいというネガティブ優位説 [45] 等が存在する。本研究ではこの人間の特性に着目し、感情状態を利用して想起しやすいできごとの記録を優先的に提示する。これによって、自己開示を動機づけさせて、コミュニケーションのきっかけを与えることを目指す。

工学分野では、生体信号を用いて感情を推定する研究が盛んに行われている [18] [46] [47]。McDuff ら [17] は、音声や表情、生体信号や操作履歴を継続的に取得するマルチモーダルセンサ と、感情を推定する分類器、そして、ユーザが振り返るためのインタフェースから構成されるシステムを構築した。彼らは、ユーザが過去のできごとの記憶を構築する際に、システムに蓄積された感情状態の情報を活用することが振り返りに役に立つことを示した。しかし、このシステムでは、時間順に記録が並べて提示されているのみで、膨大な記録の中から過去の体験を素早く探しだすことに焦点を当てていない。我々は、ユーザが、記録時の体験を素早く想起して対話できる、かつ、伝えたい体験であることを満たすような記録を抽出する方法を提案する。また、上記の記録を抽出するために、どのような感情状態が有効であるかの分析を行う。

3.2.2 感情モデルを用いた感情の定量化

日常生活における体験時の感情状態を測定可能な情報として抽出し、さらにユーザが自分自身の感情状態を評価可能とするために、感情を定義する必要がある。感情状態を表す指標として様々なモデルが提案されているが、代表的なモデルの一つとして、高覚醒-低覚醒 (Y 軸) を表す覚醒度と、ポジティブ-ネガティブ (X 軸) を表すバランス度の 2 軸から構成される感情モデルがある (図 6)。心理学の分野の従来研究では、この感情モデルの 2 軸は、記憶と強い関係があることが示されている [43] [44] [45]。また、この感情モデルは自己の感情状態のリアルタイムな評定に対して優れていることが示されており [36]、感情推定の従来研究でも多く用いられている [17] [37]。以上より、日常生活の中の感情状態を収集し、感情状態を用いて素早く体験を想起させるような記録提示を目的とする本研究においては、本感情モデル

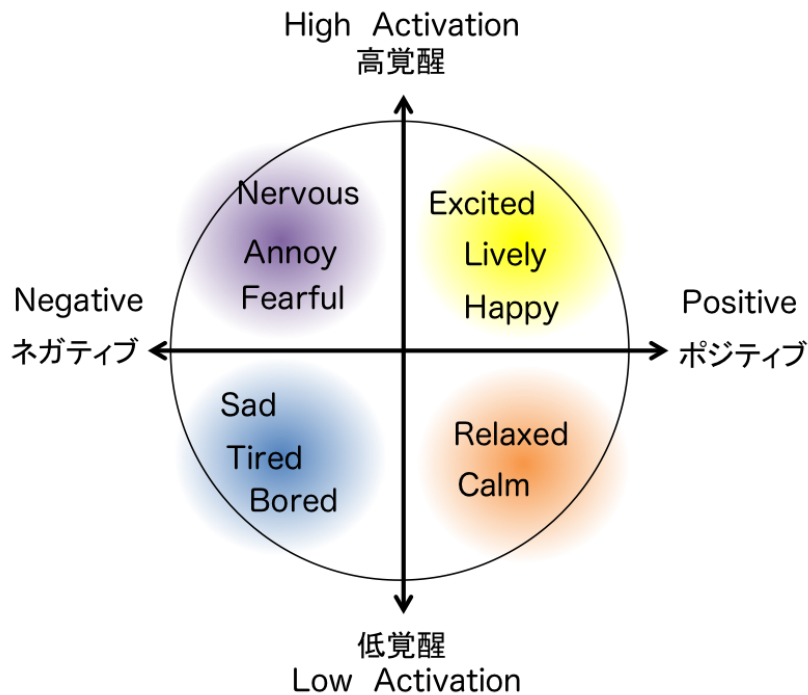


図 6: 感情モデル

を採用することは適しているとする。

本研究では、以下の仮説 I、II をたてた。ここで、感情状態が中立状態からより遠ざかることを感情状態が高まると定義する。

仮説 I 感情状態が高まり (i)、感情状態の変化が大きい (ii) ほど、その体験の記録が提示されるとその体験を容易に想起できる

仮説 II 伝えたい体験ほど、感情状態が高く (i)、感情状態の変化が大きい (ii)

これは、感情状態が高い時の体験ほど思い出しやすいことが従来研究にて示されており、伝えたい体験は、感情状態が高い時、また高い状態へ大きく変化する時の体験であることが期待されるからである。図 7 に i、ii を満たす感情状態を矢印で示す。

条件 i、ii を満たす感情状態を、感情モデルを用いて表すために、ユーザが時刻 t で感情モ

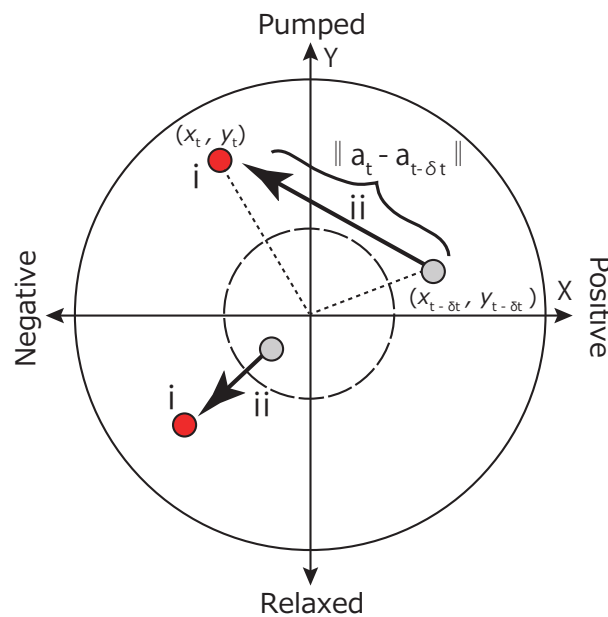


図 7: 伝えたい体験時の感情状態の仮説: i 感情の大きな変化が認められ、ii 変化先の感情状態が強い状態

デル上に入力する感情状態を、バランス軸を X 軸、覚醒度軸を Y 軸として、 -1 から $+1$ までの数値 $\mathbf{a}_t (= [x_t y_t]^T)$ で表すとする。これによって、ユーザ i の感情が高まる状態への変化値は、 $\|\mathbf{a}_{i,t} - \mathbf{a}_{i,t-\delta t}\|_2$ ($\|\mathbf{a}_{i,t}\|_2 > \|\mathbf{a}_{i,t-\delta t}\|_2$) と表すことができる。ここで、 $\|\cdot\|_2$ は L2 ノルムを表す。

3.2.3 予備実験

本予備実験の目的は、仮説 i と ii を共に満たす体験時のコンテンツを提示した場合は、そうでない体験時のコンテンツを提示した場合と比べて、より体験を想起しやすく伝えたい、という仮説が満たされるかを明らかにすることにある。

実験内容

仮説の検証に向けて、被験者の感情状態に変化を与えるために、被験者に対して映像による刺激を提示した。まず、被験者(Aとする)へ映像を提示し、その間、一定時間ごとに指示を与えて、その時の感情状態を評価させた。次に、映像視聴という体験を通して、シーンごとの「できごと・感情の想起しやすさ」と「伝えたさ」を評価するために、視聴から数日後、映像を見たことがない別の被験者(Bとする)に対して、ある条件で抽出された映像のサムネイルについて説明させる実験を行った。以下に、実験内容の詳細について述べる。

まず、被験者(A)11名(男性:8名、女性:3名)に対して約90分間(全6作品、1作品:5~20分)の映像を視聴させた。映像内容は、感情を効果的に喚起させるために、作品中の登場人物への感情移入を促すことが期待できる実写映画を選択した。被験者には、記憶に何かしらの影響を与えないようにするために、主観評価による視聴時の感情データの収集を目的としていることのみを伝えてある。映像視聴中には、約2~3分間隔でベルを鳴らし、ベルが鳴ったタイミング時の被験者自身の感情状態を評価させる。感情評価は前述した感情モデルを用いて評価させるが、評価と入力方法はRobbyら[36]の研究を参考に、図8のように感情モデルを端末のタッチディスプレイ上に提示して、モデル上を直接指でタッチすることで評価と入力を行わせる。被験者にはあらかじめ感情モデルの各軸の説明を行い、モデルの中心部をニュートラルとして、評価時の感情がニュートラルな状態からどれだけ離れているかで評価を行うよう伝えてある。

映像視聴から10日~13日経過後、映像を視聴した被験者(A)に映像から抽出したサムネイルを提示し、そのサムネイルについて映像を視聴していない別の被験者(B)に対して説明させた。サムネイルは図9に示すように、視聴時に入力された感情評価データを元に、仮説で立てた感情状態のシーンを切り抜いた「ハイライトシーンサムネイル」と、仮説と対局のシーンとして、感情の変化が小さく、変化後の感情状態がニュートラルに近い時のシーンを切り抜いた「比較シーンサムネイル」の2種類を含む。ここで述べるシーンとは、感情評価の入力区間である約2~3分間を指しており、1シーンにつき3つのサムネイルを提示する。

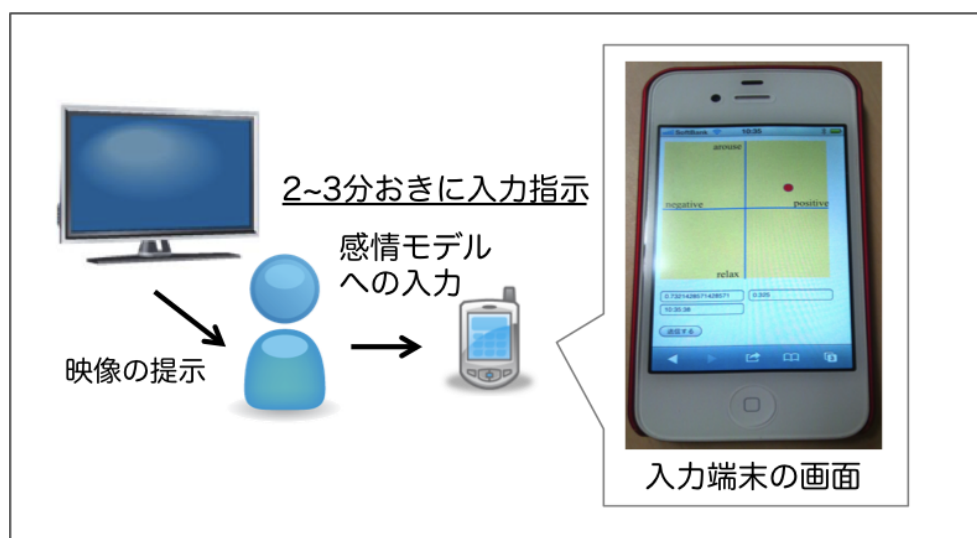


図 8: 感情モデルへ入力するために用いた端末画面

3つのサムネイルは、1シーンをさらに3分割した中から1サムネイルずつ選択したものである。このとき、作為的なサムネイル抽出にならないように、各分割区間内での中央地点周辺において登場人物が映っている部分をサムネイルとして用いた。

提示する「ハイライトシーンサムネイル」の抽出方法について述べる。いま、最も感情状態が高まった時の時刻を t_j 、その感情状態へ変化し始める直前の時刻を $t_j - \delta t_j$ としたとき、被験者 i が各時刻において入力する感情状態の値は \mathbf{a}_{i,t_j} 、 $\mathbf{a}_{i,t_j - \delta t_j}$ と表される。このときハイライトシーンサムネイルは以下の条件で抽出される。

$$\|\mathbf{a}_{i,t_j}\|_2 > \tau_1 \quad \text{かつ} \quad \|\mathbf{a}_{i,t_j} - \mathbf{a}_{i,t_j - \delta t_j}\|_2 > \tau_2 \quad (1)$$

ここで、 τ_1 は $\frac{1}{2N} \sum_{j=1}^N (\|\mathbf{a}_{i,t_j}\|_2 + \|\mathbf{a}_{i,t_j - \delta t_j}\|_2) \times C$ で算出した。 τ_2 は $\frac{1}{N} \sum_{j=1}^N \|\mathbf{a}_{i,t_j} - \mathbf{a}_{i,t_j - \delta t_j}\|_2 \times C$ で算した。Nはサムネイルの個数を表し、Cは0.8とした。式(1)を満たす記録の中から、 $\|\mathbf{a}_{i,t_j}\|_2 + \|\mathbf{a}_{i,t_j - \delta t_j}\|_2$ の値の降順で上位4つのサムネイルを「ハイライトシーンサムネイル」

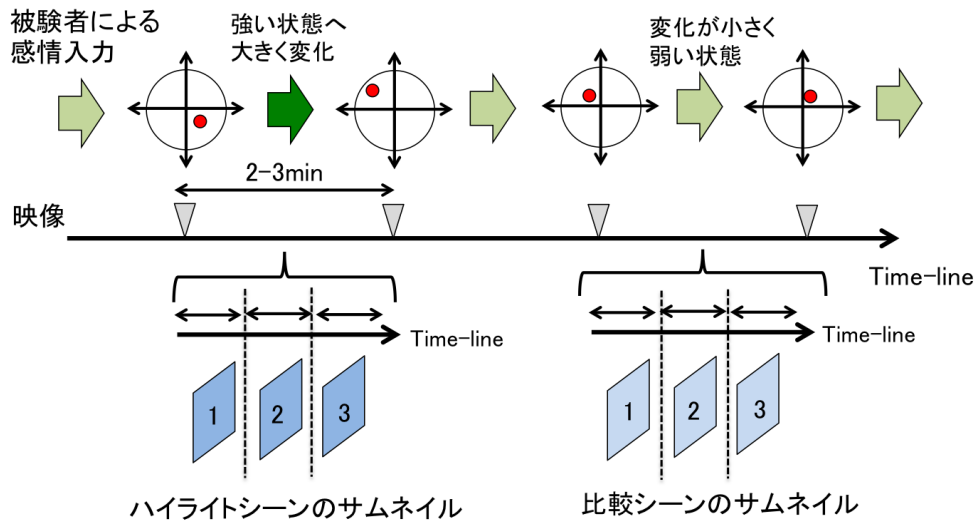


図 9: サムネイル抽出の概要

とする。比較シーンサムネイルは、

$$\|\mathbf{a}_{i,t_j}\|_2 < \tau'_1 \quad \text{かつ} \quad \|\mathbf{a}_{i,t_j} - \mathbf{a}_{i,t_j - \delta t_j}\|_2 < \tau'_2 \quad (2)$$

を満たす記録の中から、 $\|\mathbf{a}_{i,t_j}\|_2 + \|\mathbf{a}_{i,t_j - \delta t_j}\|_2$ の値の昇順で上位 4 つのサムネイルとして選択した。ここで、 τ'_1 と τ'_2 は C を 0.4 として、 τ_1 と τ_2 と同様の式で算出した。

被験者 (A) には、上記条件によって抽出された 4 シーン以内の「ハイライトシーンサムネイル」と、4 シーン以内の「比較シーンサムネイル」、計 8 シーン以内をランダムに提示し、各シーンに含まれる 3 つのサムネイルが何を表しているのか、聞き役の被験者 (B) に対して説明させた。各シーンの説明終了時に、説明した被験者 (A) に、図 10 に示す「1. サムネイル周辺のできごとを想起しやすかったか」「2. サムネイル周辺の感情を想起しやすかったか」「3. 相手に伝えたいと思ったか」という評価項目に対して 7 段階で評価を行わせた。



図 10: 主観評価項目

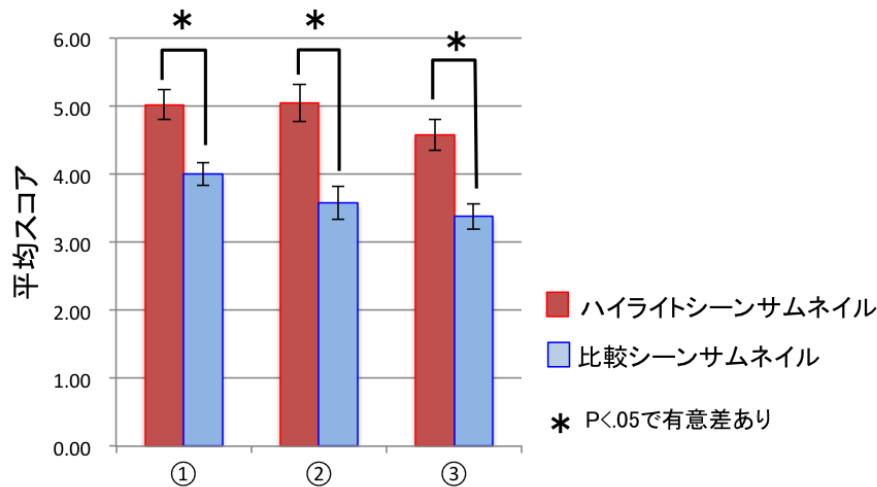
実験結果

実験結果を図 11 に示す。縦軸は「ハイライトシーンサムネイル」と「比較シーンサムネイル」に対する 7 段階評価の回答に、1～7 点でスコア付けを行ったものの全被験者 (11 名) の平均値を表している。3 つの評価項目に対して「ハイライトシーンサムネイル」と「比較シーンサムネイル」の結果に差があるか t 検定を行ったところ、いずれの評価項目においても、2 種類のサムネイルの間には $p < .05$ の有意差がみられた。

以上の結果より、I. 感情の大きな変化が認められ、II. 変化先の感情状態が強い、体験時のコンテンツを提示した場合はそうでない体験時のコンテンツを提示した場合と比べて、1. 「できごと」を想起しやすい 2. その時の「感情」を想起しやすい 3. 伝えたい、ということが示された。これは、仮説が妥当であることを示唆する。

3.2.4 提案インタフェース

予備実験より、感情状態 I、II を用いることが、できごと・感情を想起しやすく、伝えたいコンテンツを提示するために、有効であることが明らかになった。しかし、予備実験では、実験室環境下における映像コンテンツの体験に限定しており、日常生活においても、感情状態に着目したコンテンツ提示が、コミュニケーションに対して有効であるかは分からない。そこで本節では、日常生活で記録される体験のコンテンツを対象として検証を行う。日常生活におけ



	t値	p値
① 出来事の想起しやすさ	2.7994	0.01107
② 感情の想起しやすさ	4.9519	0.00007677
③ 伝えがい	3.1185	0.005413

図 11: 実験結果

る体験の記録方法と、コミュニケーション時の提示手法を備えたインタフェース”Feelboard”を設計し、Feelboard を用いて実験を行う。

ここで、体験の一部の記録提示でも、体験の記憶を再構築するきっかけとして十分であることから [40]、Feelboard は、全ての体験の記録を提示するのではなく、感情状態 i と ii を用いて、想起しやすく伝えたい体験の記録を優先的に提示することで効率的にコミュニケーションのきっかけを与える。そこで、仮説 I と II が満たされる場合、さらに以下の仮説をもとに Feelboard のインタフェースの設計を行う。

仮説 III 感情状態 i と ii を満たす時に残された記録を上位に提示することは、対話時に伝えたい記録を素早く見つけ出すことを促す

本節での実験は、仮説 I と II を検証するための実験と、仮説 III を検証するための実験から構成される。

a) 日常生活における体験の記録

ユーザはできごとと感情状態を記録し、その記録は Feelboard に日時と共に保存される。ユーザは、できごとの記録を残す時に、そのできごとを通して最も高まった時 (時刻 t) の感情状態と、その感情状態へと変化し始める直前 (時刻 $t - \delta t$) の感情状態の 2 つを Feelboard に記録する。 δt は任意の値である。感情状態の記録は、予備実験と同様、端末画面上に表示された感情モデルの 2 次元座標系上をポイントすることで入力させる。ユーザに最も高まった時の感情状態と変化開始直前の 2 点を入力させることで、記録の時間間隔が多少長い場合であっても、その時間区間における感情状態の最大値とその変化量を記録することが可能となる。

b) 提示手法

感情状態 i と ii に基づいて記録を提示するために、収集された記録は、感情の変化値 $\|a_{i,t} - a_{i,t-\delta t}\|_2$ の降順で順位づけられ、上位から提示される。Feelboard が提示する記録は、図 12 に示すように、ユーザが記録したできごとに関するテキストと、入力された感情状態を表したアイコン、そして記録日時から構成される。ユーザは画面をスクロールすることによって上位の記録から順番に閲覧することができる。感情状態の表現は、直感的に理解しやすいことを示した [17] らの研究を参考に、感情状態のバランス軸をバブルの色によって表現し、覚醒度軸をバブルの形で表現した。具体的には、ポジティブは暖色系であるオレンジ色で、ネガティブは寒色系である青色、高覚醒はバブルが破裂した形で、低覚醒はバブルが歪んだ形で表現する。バブルを表現するためのパラメータは、感情状態の最大値 $a_{i,t}$ を用いる。

3.2.5 実験：感情状態とコミュニケーションのきっかけ (仮説 I、II)

本実験では、日常生活での体験を対象とした被験者実験を行い仮説 I、II の検証に取り組む。

体験の記録方法

20 名の被験者 (男性:16 名、女性:4 名) に、スマートフォンもしくは PC を用いて、日常生活

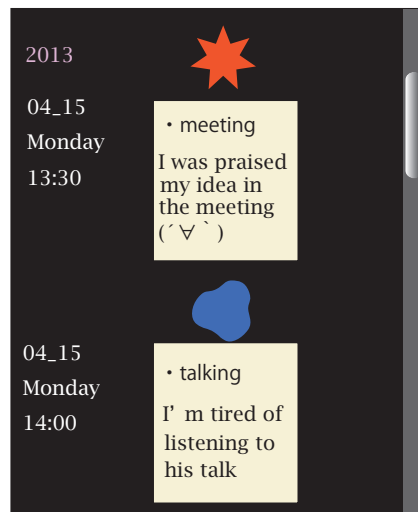


図 12: Feelboard は感情状態の変化の大きさに基づいて記録を並べて提示する。

の中でのできごとと感情状態の記録を行わせた。記録は約 30 分に 1 回の間隔で入力させ、入力時間帯は 1 日につき 7 時間で 1 週間継続させた。集まった記録数は被験者 1 人につき 100~140 記録となった。1 回に入力させる記録の詳細を以下に述べる。

- 「感情状態」の入力：約 30 分の体験を通して、最も感情が高まった状態の感情状態と、その感情状態へ変化し始める直前の感情状態の 2 点を感情モデル上に入力させた。
- 「できごと」の選択：約 30 分でのできごとを”会議”、”食事”、”デスクワーク”、”雑談”、”休憩”、”移動”、”趣味・遊び”、”その他”の中から選択させた。
- 「できごと」の詳細な入力：振り返り時にできごとを区別できるように、詳細な状況を記入させた。

また、記憶に影響を与えないように、被験者は記録の履歴を閲覧できないようにした。実験に用いた入力画面の例を図 13 に示す。

体験の想起

Input two states, 1)strongest state and
2) the state just before changing to the strongest state
over the last 30 minutes

Pumped

Negative

Positive

Relaxed

Event Option

Two-dimensional Emotion Space

What did you do?

Write, in detail, of your situation

I was talking with my mom about cooking.

Free Text Entry

Send

図 13: Feelboard の実験用体験記録画面

収集された記録の中から、感情状態 i、ii を満たす時の記録 (以下、Feelboard 記録と呼ぶ) と、感情の変化が小さく、変化後の感情状態がニュートラルに近い時の記録 (以下、比較記録と呼ぶ) の 2 種類を被験者に提示した。本実験の目的は、Feelboard 記録の提示が、伝えたい体験を想起しやすくさせること (仮説 I、II) を示すことなので、被験者には、提示される各記録に対して、その時の体験についてのメールを書かせ、主観評価を行わせた。メールを書くというコミュニケーションのタスクを与えた理由は、親しい友人とプライベートな体験について会話するところを第 3 者が観測するのは自己開示を妨げると考えたからである。また、同じ理由から、研究者はメールの内容を見ずに、メールの内容を加工した結果のみを取得することを被験者に伝え、その通り実践した。もう 1 つの理由は、体験によって話しやすい相

手が変わると考えられるが、メールの場合だと伝える相手を柔軟に選択することができるからである。

提示する記録の抽出方法は、予備実験と同様の方法である。式(1)を満たす記録の中から、 $\|\mathbf{a}_{i,t_j}\|_2 + \|\mathbf{a}_{i,t_j-\delta t_j}\|_2$ の値の降順で上位2つの記録を Feelboard 記録とし、式(2)を満たす記録の中からを満たす記録の中から、 $\|\mathbf{a}_{i,t_j}\|_2 + \|\mathbf{a}_{i,t_j-\delta t_j}\|_2$ の値の昇順で上位2つの記録として選択した。

抽出された2つの Feelboard 記録と2つの比較記録、計4つの記録を被験者にランダムな順序で提示した。被験者には、1つの記録につき話しやすい相手を思い浮かべ、その相手に対して、記録から思い出される体験について伝えるメールを書かせた。被験者には、1つの記録についてのメールを書き終えるごとに「記録からできごとを想起しやすかったか」「記録から感情を想起しやすかったか」「思い出した体験を相手に伝えたいと思ったか」という評価項目に対して7段階評価を行わせた。書かれたメールからは、形態素解析ツールを用いて解析して、文字数と名詞、動詞、形容詞のみを取得した。また、記録を見てからメールを書き始めるまでの時間も測定した。

3.2.6 実験結果：(仮説 I、II)

主観評価、及び、定量評価は、実験の中で指示に反して話し始めてしまった被験者1名を除いた19名で行った。

主観評価

図14に、主観評価の結果を示す。全ての評価項目においてt検定を行ったところ Feelboard 記録と比較記録との間に有意な差が認められた ($p < .001$)。これより、日常生活においても、仮説 I、II は満たされることが示された。このことから、感情状態 i、ii を用いて記録を提示することは、ユーザに伝えたい記録を想起しやすくさせ、コミュニケーションを促すことに効果があると考えられる。

定量評価

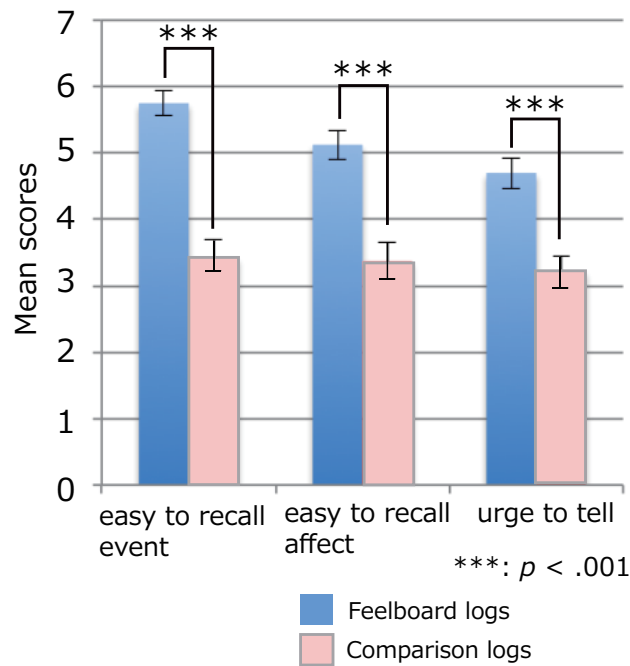


図 14: 被験者 19 名の 7 段階評価の平均値。

表 1 に Feelboard 記録と比較記録についてのメール内容の結果を示す。メールを書き始めるまでの時間に関しては、対応のある t 検定を行ったところ、Feelboard 記録の方が比較記録に比べて短い時間で書き始める傾向が見られた ($p = 0.056$)。このことから、感情状態 i、ii を満たす時の記録を提示することは、その時の体験を素早く思い出させ、円滑にコミュニケーションを開始させることに効果があることが期待される。一方で、それ以外の結果については特に傾向は見られなかった。この結果の原因の一つとして、メールを書く時間や、体験にまつわるエピソードの範囲に制限を与えなかったため、被験者が全てのメールの間で分量の偏りが生じないように書いた可能性が考えられる。

	Feelboard 記録	比較記録
書き始めるまでの時間 (秒)	12.60	19.86
書くスピード (文字数/秒)	1.19	1.22
文字数	190.34	181.28
名詞数	30.84	27.97
形容詞数	2.10	2.44
動詞数	16.36	15.86
感情語数	2.2	2.05

表 1: メール内容と書き始めるまで分析結果 (被験者 19 名の平均値)

3.2.7 実験：伝えたい記録の発見 (仮説 III)

本節では、仮説 III を明らかにするため、記録を感情状態でランク付けて並べる提案インタフェース、Feelboard を用いた場合に、制限時間内で選ばれる記録が真に伝えたい記録とどの程度一致するかで評価する。ここでは、真に伝えたい記録を、制限時間を与えずに選ばれた記録とすることとした。また、比較法として、時間順に並べられた記録のリスト (以下、従来の記録リストと呼ぶ) を用いる。

本実験では、仮説 I、II の実験で参加した被験者 20 名に対して、各被験者が記録した 6 日間のできごとと感情の記録の中から、ある記録を探し出させるという 6 つのタスクを行った。提示される記録のリストは、タスクごとに異なる方法で並び替えられている。1 つのタスクは、Feelboard を用いて、伝えたい記録を探させるタスクである。別の 1 つのタスクは、従来の記録リストを用いて、伝えたい記録を探させるタスクである。それ以外は偽のタスクとなっており、ランダムに並び替えられた記録の中から、会議に関する記録の数、楽しかった会話をした時の記録の数、何かを食べている時の記録の数、眠っている時の記録の数を数えさせた。偽のタスクを行わせた理由は、実験の意図を被験者に気づかせないこと、そして、被験者ごとに偽タスク時の 1 記録あたりの平均閲覧時間を測定し、伝えたい記録を探させるタスクの制限時間を設定するためである。偽タスク時には、任意の短い制限時間を与え、タスク終了後に閲覧できた記録数から 1 記録あたりの平均閲覧時間を算出している。各リストは図 12 に示すように 1 画面内に 2 記録ずつ提示され、被験者はスクロールしながら記録を閲覧す

Task no.	1	2	3	4	5	6
Group A	(Td, F)	(Td, F)	T1, F	(Td, L)	(Td, L)	T2, L
Group B	(Td, F)	(Td, F)	T2, F	(Td, L)	(Td, L)	T1, L
Group C	(Td, L)	(Td, L)	T1, L	(Td, F)	(Td, F)	T2, F
Group D	(Td, L)	(Td, L)	T2, L	(Td, F)	(Td, F)	T1, F

T1: Feelboard, T2: conventional list,
Td: dummy task F: First three days, L: Last three days

表 2: 実験設計の詳細

ることができる。表示内容、及び、操作方法については、タスク開始前にあらかじめ説明してある。

被験者 20 人は 4 つのグループ (A ~ D) に分けられる。各グループに対して、前半の 3 日間の記録を 3 種類の並び順のリストで提示し、後半の 3 日間の記録も同様に 3 種類の並び順のリストで提示し、リストごとに 1 つのタスクを行わせた。グループによって、前半と後半の記録の並び方とタスクの内容は異なる。これは、前半と後半で記憶の残り方の差が出る可能性を考慮して、Feelboard と従来の記録リストを平等な条件下で比較するためである。伝えたい記録を探させるタスクは、3 番目と 6 番目のタスクに設定した。被験者は記録の並び方については何も知らされていない。以上の実験内容をまとめたものを表 2 に示す。Feelboard と従来の記録リストを提示するとき、被験者には制限時間内に伝えたい記録を少なくとも 3 つ以上選び、印をつけるよう指示を与えた。制限時間は全ての記録の 60%のみを閲覧できるように、偽のタスク時に 1 記録あたりの閲覧時間 s を測定し、記録数 $*s*0.6$ と設定した。全てのタスクを終了後、今度は制限時間を設けずに伝えたい記録を 5 つ以上選ばせて、伝えたい順番で番号をつけさせた。

3.2.8 実験結果 (仮説 III)

制限時間を設けられていない条件下で、選択されて順位づけられた伝えたい記録を正解データとする。そして、Feelboard と従来の記録リストについて、制限時間内で正解データの上

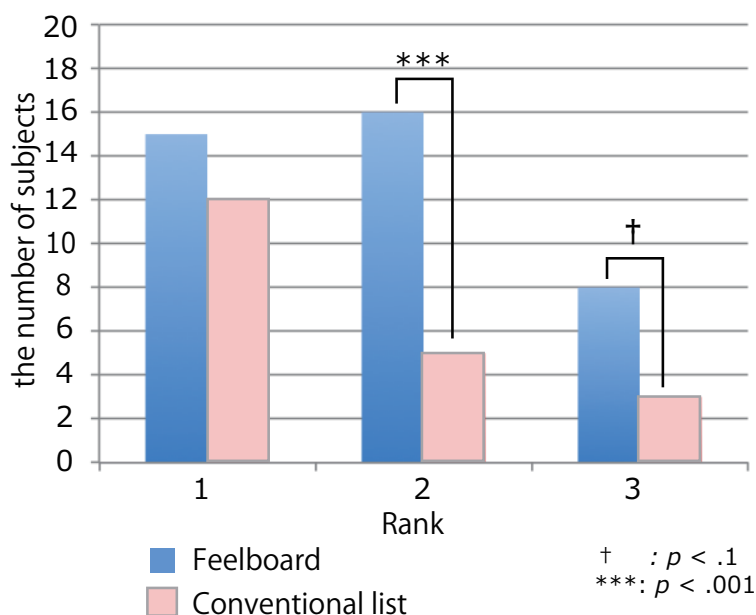


図 15: 正解データの各順位での記録を見つけることができた被験者数

位3つの記録を選ぶことができたかという点で比較した。図15には正解データの各順位の記録を選ぶことができた被験者数を示しており、縦軸は被験者数、横軸は正解データの順位を表す。

各棒グラフは、Feelboardを用いた時と従来の記録リストを用いた時の、正解データの各順位の記録を選ぶことができた被験者数を表している。いずれの順位においても、Feelboardを用いた時の方が正解データを選んだ被験者数が多くなっている。このことから、Feelboardは、従来の時間順に並べた記録の提示手法よりも短い時間でコミュニケーションを開始することを促進することが期待される。また、各順位におけるFeelboardと従来の記録リストの間で、カイ二乗検定を行ったところ、2位においては有為差が認められ($\chi^2 = 12.1, p = 0.00049$)、3位においては有意傾向が認められた($\chi^2 = 3.13, p = 0.076$)。しかし1位のみで、従来の記録リストにおける正解データを選ぶことができた被験者数が多くなっており、Feelboardとの間

に特に有為差は見られなかった ($\chi^2 = 1.02, p = 0.31$)。この結果の原因の1つとして、ある被験者はあらかじめ伝えたい記録を思い浮かべて、その記録を探し始めたため、残りの2つの記録を探し出す時間が足りなくなったことが考えられる。一方 Feelboard では、伝えたい記録が上位に含まれているため、多くの被験者が1位から3位までの正解データを探し出すことができたと考えられる。このことから、感情状態を用いて記録を並べて提示することは、ユーザが短い時間で様々な体験について話すことを可能にすると考えられる。これによって相手や状況に合わせて話題を選択できるようになり、会話を盛り上げることに役に立つと考えられる。

3.2.9 まとめ

Feelboard は、人との対面コミュニケーションにおいて新たな話題を提供するという、高次の活動コンテキストに限定した、コンテンツ提示手法である。我々はこの活動コンテキストに対して、体験に対する記憶と伝えたさが、体験時の感情状態と関連があることを利用し、感情状態に基づいたコンテンツ提示のデザインを行った。Feelboard は、ユーザが短い時間で伝えたい体験を思い出すことを支援でき、これは、時刻情報や位置情報などの、従来の低次のコンテキストのみでは、ユーザの要求が不明確なため実現できない。近年では、日常生活において、記録や写真などのコンテンツのみでなく、モバイルデバイスのセンサを利用して、人の感情情報も容易に残すことができるようになってきたため、Feelboard をサービスとして実現することは可能である。

3.3 対話中の話題に適したプライベートコンテンツ提示手法

本研究では、代表的なコンテンツである写真を対象として、対面コミュニケーションが開始してからの課題に取り組む。写真を利用した対面コミュニケーションでは、人はしばしば、現在の話題に関連した写真を探し、相手に見せることがある。これは、自身の自己開示のみが目的でなく、現在の話題に関連するので、相手が興味を持つ可能性があると考え、会話を盛り上げることが目的でもある。そこで、我々が想定する写真を利用した対面コミュニケーションとは、ユーザ同士が、提示された写真を話題に対話を始め、その対話によって次に提示したい写真を選択し、選択した写真を話題に、次の対話を始めることである。

写真を利用したコミュニケーションにおけるユーザの行動の具体例を図16に示す。いま、ユーザAとユーザBの間で行われる、写真を利用したコミュニケーションを考える。はじめに、ユーザAが写真を提示し、その写真を話題にして話したとする。このときユーザBは、ユーザAとの対話で想起した、自分の体験を伝えたいとする。これは、“photo-elicitation” [41] と呼ばれ、人が写真から記憶や考えを引き出される現象を意味する。ユーザBは自分の写真の中から、伝えたい体験と関連のある写真を探し出し、見つけた写真をユーザAに提示し、その写真を話題に次の話を始める。つまり、2者間における写真を利用したコミュニケーションは、ユーザ同士が提示された写真に基づいて対話を行い、対話の話題から想起される体験と関連のある写真を探し、次の写真を提示する、という流れで記述できる。

これらのユーザの行動を踏まえて、写真を利用したコミュニケーションにおけるユーザの要求に、現在の写真から想起される話題と関連のある写真を、素早く見つけたいという点があげられる。ユーザは、伝えたい体験を含む写真を、素早く見つけ出せることによって、対話を中断することなく次の話題への移行が可能になり、円滑なコミュニケーションを実現できる。アクセス可能な写真の枚数が増加するにつれ、目的の写真が見つからない場合や、自分が持っている写真の存在を忘れてしまう可能性も高くなるので、対話中に、話題と関連のある写真を、容易に見つけ出せることは重要である。そこで、本システムの要件を、現在の話題というコンテキストに対して、ユーザが伝えたい体験を含む写真を優先的に提示するこ

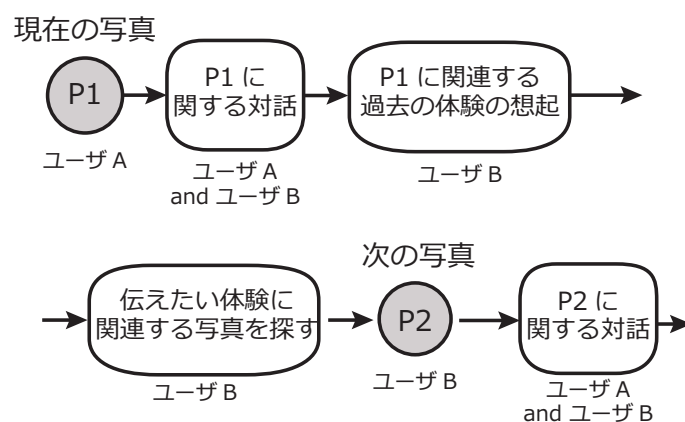


図 16: 写真を利用したコミュニケーション

ととする。

従来研究に、対話のコンテキストに合わせたリアルタイムな情報提示によるコミュニケーション支援の研究は存在する。Sommererら [48]の研究では、対話者の発話内容から音声処理技術によって単語を抽出し、その単語をクエリとしてインターネット上からアイコン画像を取得し、リアルタイムに画面上にアイコン画像を提示してコミュニケーションの支援を行っている。MeetBall [49]は、コミュニケーション参加者の発話内容から単語を抽出し、その単語に適合する画像をテーブルトップに映写して発話内容を視覚化することで会話の活性化を支援するシステムである。例えば、一部の参加者しか知らない事物が言及された場合でも、それがMeetBallによって映像化されることで全員が視認する事ができ、会話が中座してしまったり一部の参加者だけで会話が盛り上がったりしてしまうことを防いでいる。これらの研究では、コミュニケーション支援の情報として、インターネット上に公開された画像を利用しており、対話者自身の写真を利用していない。本研究では、対話コミュニケーションにおいて、対話者自身の写真への効率的なアクセスを支援するという点で大きく異なり、さらに我々は、ユーザが伝えたい体験を含む写真を優先的に提示できるかについての検証も行う。

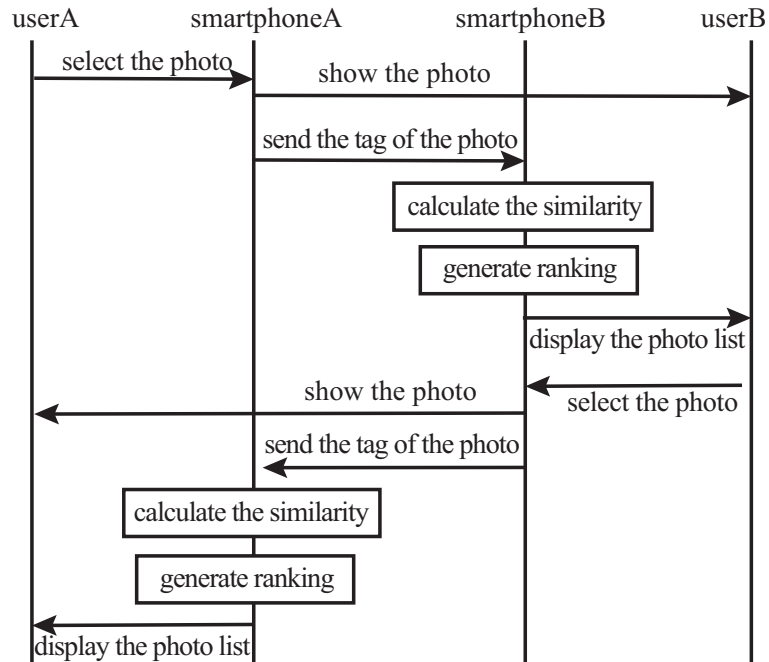


図 17: Photo Ranking のシステム概要

3.3.1 Photo Ranking

我々は、対面コミュニケーションのための写真アクセスシステム”Photo Ranking”を提案する。提案システムでは、現在のコンテキストに対して提示する写真の適切さを、現在提示されている写真との類似度として定義する。これは、前節で述べたように、現在提示された写真を話題にしたコミュニケーションにおいて、次に伝えたい写真とは、現在の写真と関連した体験を含む写真、と考えられるためである。ユーザは、各自のスマートフォンに自身の写真を所有し、各写真には、画像処理技術を利用した自動タグづけサービス等によって、あらかじめタグが付与されていることを前提とする。

システム概要を図 17 に示す。いま、二人のユーザが、各自のスマートフォン内の写真を用いて対話を行っており、ユーザ A が自身の写真の 1 枚を選択し、ユーザ B に提示したとする。写真が選択されると、各デバイス間で通信が開始され、ユーザ A が選択した写真のタグ情報

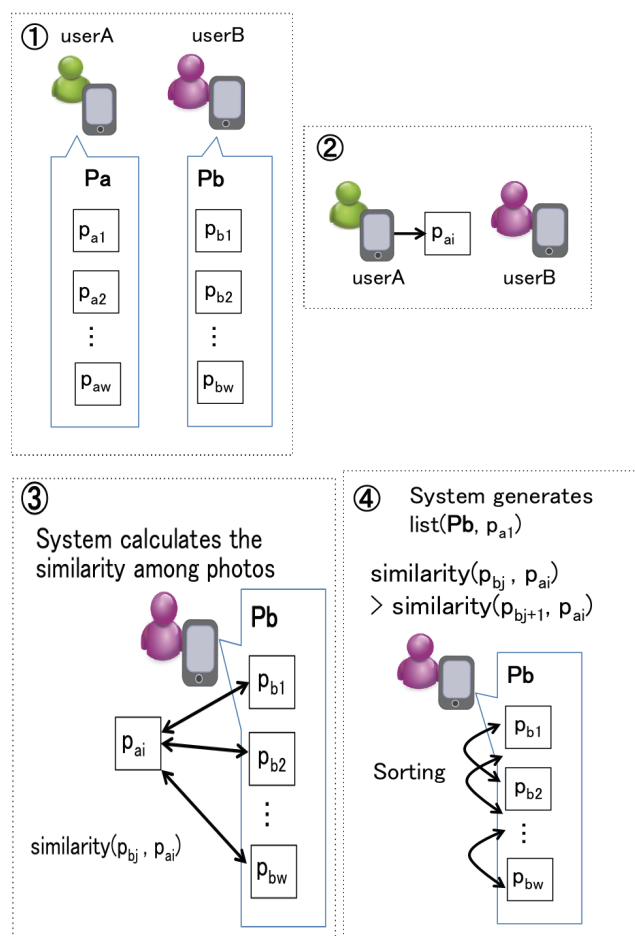


図 18: システムの形式化

がユーザ B の記録デバイス B に送信される。デバイス B では、ユーザ A が選択した写真と、デバイス B 内に保存された写真との類似度を、写真のタグ情報を利用して計算する。デバイス B 内の写真集合は、算出された類似度に基づいて、並び替えられる。ユーザ B は、並び替えられた順番で写真を閲覧し、次に提示したい写真を探し選択する。ユーザは、現在のコンテキストに適した写真が優先的に提示されることによって、効率的な写真アクセスが可能となる。

以下、本システムを形式化する。形式化概要を図 18 に示す。ユーザ A とユーザ B がそれ

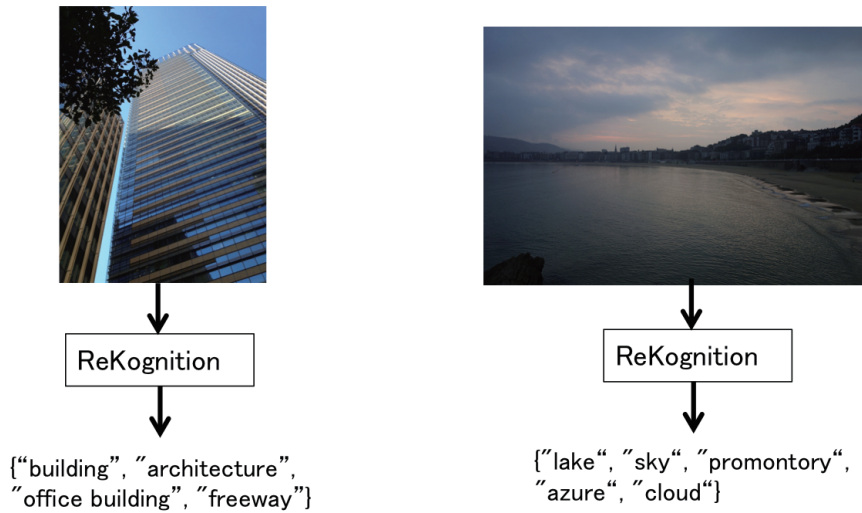


図 19: ReKognition API の出力例

それぞれ所有する写真集合を $P_a = \{p_{aj} | j = 1, \dots, w\}$ 、 $P_b = \{p_{bj} | j = 1, \dots, w\}$ とする。写真集合 P_a の中から、ユーザ A が選択した写真を p_{ai} 、写真集合 P_b の中から、ユーザ B が選択した写真を p_{bi} とする。

いずれかのユーザが選んだ写真を利用して、コミュニケーションが行われることとする。いま、ユーザ A が選択した写真 p_{ai} を利用して対話していた場合、ユーザ B の写真と写真 p_{ai} との類似度 $s(p_b, p_{ai})$ が算出される。求められた類似度に基づいて、並び替えられたランキングリストを、 $List(P_b, p_{ai}) = \langle p_{b1}, p_{b2}, \dots, p_{bw} \rangle$ と表記する。リストの並び順は、 $s(p_{bj}, p_{ai}) > s(p_{bj+1}, p_{ai})$ を満たす。

写真間の類似度 $s(\cdot, \cdot)$ は、写真から抽出されるタグ情報を利用して算出する。画像処理による写真のタグ付け技術に関する研究は盛んに行われており [50] [51]、近年では、自動タグ付けサービスも提供されている [52]。本研究では、自動タグ付け技術サービスの API “ReKognition” [52] を利用する。ReKognition API は、画像に含まれるオブジェクトやシーンを認識し、認識結果を単語として出力する。図 19 に入力画像に対する出力例を示す。本研究では、出力される単語をタグとして利用する。

タグ間の距離は、Normalized google Distance (NGD) [53] を利用する。NGD は、単語間の類似度の近似値を Google の検索結果の件数から求める手法である。いま、写真 p_{a1} 、 p_{b1} におけるタグ集合を $T(p_{a1}) = \{t_k | k = 1, \dots, m\}$, $T(p_{b1}) = \{t_l | l = 1, \dots, m\}$ としたとき、NGD によるタグ間の距離は以下の式にて求められる。

$$\begin{aligned} NGD(t_k, t_l) \\ = \frac{\max\{\log(h(t_k)), \log(h(t_l))\} - \log(h(t_k, t_l))}{\log N - \min\{\log(h(t_k)), \log(h(t_l))\}} \end{aligned} \quad (3)$$

ここで、 $h(t_k)$ と $h(t_l)$ は、タグ t_k または t_l を Google で検索したときのヒット数であり、

$h(t_k, t_l)$ は t_k かつ t_l を Google で検索したときのヒット数である。 N は Google のインデックス数である。NGD を用いた写真間の類似度 $s(p_{a1}, p_{b1})$ は、以下の式にて求める。

$$s(p_{a1}, p_{b1}) = 1 - \min_{t_k \in T(p_{a1}), t_l \in T(p_{b1})} \{NGD(t_k, t_l)\} \quad (4)$$

以上より算出される写真間の類似度を用いて、写真のランキングを生成する。

3.3.2 評価実験

本実験では、現在提示されている写真に対して、ユーザが伝えたい話題を含む写真を、提案システムが優先的に提示できるか、という観点で評価する。まず、被験者に、提示される写真と自身の各写真とを比較させ、「伝えたさ」の5段階で評価を行わせる。そして、主観評価によって並び替えたランキング (以下、主観評価ランキングと呼ぶ) と提案システムのランキングとの関係性によって評価を行う。

実験方法

事前準備として、被験者7名 (20代、男性5名、女性2名) を対象に、各被験者から日常

で撮影した写真 20 枚を収集した。収集する写真の枚数が少ないため、撮影対象のジャンルに偏りが生じないように、以下の手順で準備させた。まず利用頻度の高いタグが公開されている Flickr のタグクラウドサービス [54] から、固有名詞を除いて、意味的に重複しないタグを 20 個抽出した。そして被験者には、その中から 10 個以上のタグを用いて、タグと意味の近い写真を 20 枚選ぶよう指示した。

いま、7 名の被験者を A~G としたとき、被験者 A に対して、被験者 B~G の写真の中からランダムに選択した 9 枚の写真を提示する。被験者 A は提示される 9 枚の各写真についてコミュニケーションが行われるという仮定のもとで、自身が所有する写真 20 枚の各写真に対して「伝えたさ」の観点で 5 段階評価 (1:話したくない、5:とても話したい) する。被験者 A の写真 20 枚を用いて、提示された写真との類似度によって得られる提案システムのランキングと、主観評価のスコアの降順による主観評価ランキングの 2 種類を生成する。残りの被験者に対しても同様の方法でランキングを生成し、評価の合計サンプル数は 63(9 ランキング × 7 名) サンプルとなった。

実験結果

提案システムが、任意の写真に基づくコミュニケーションにおいて、被験者が伝えたいと思う写真を優先的に提示できるかについて評価を行う。本評価では、被験者の「伝えたさ」に関する主観評価ランキングに対する提案システムのランキングの一致性で評価を行った。

各写真には、主観評価のスコア、提案システムによる類似度の値が含まれる。評価のために、主観評価のスコアを基準にして、被験者ごとに 20 枚の写真を並び替えた。伝えたさのスコアが同点の写真は、その中でランダムに並び替えてある。並び替えた写真の上位 10 枚と下位 10 枚における、提案システムによる類似度の 63 サンプルの平均値は、それぞれ 0.0225、0.0112 であり、63 サンプルの中で t 検定を行ったところ、 $p < .001$ で有意差があった (図 20 左)。並び替えた写真を 3 分割した上位 7 枚、中央 7 枚、下位 6 枚における類似度の平均値は、それぞれ、0.0241、0.0162、0.0092 であり、上位と下位の間と、上位と中央との間には、いずれも $p < .001$ で有意差があった (図 20 左)。

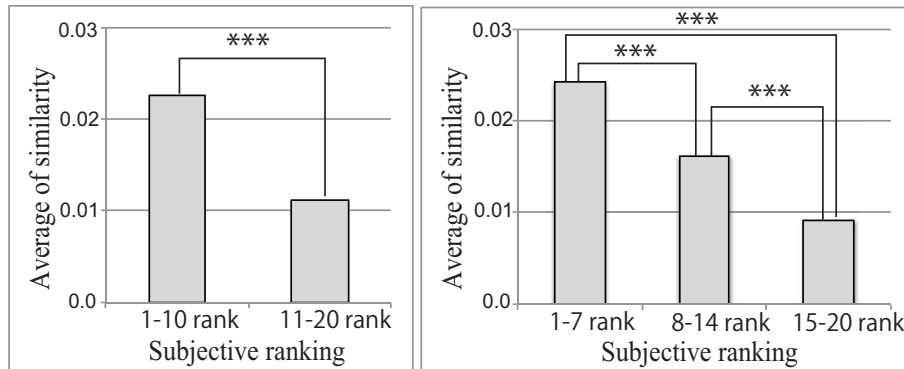


図 20: 主観評価ランキングの各順位範囲における、63 サンプルの写真類似度の平均値

さらに、DCG (Discounted Cumulative Gain) を用いて評価を行った。提案システムのランキングにおける順位 i の主観評価スコアを g_i とした上位 10 番目までの DCG は

$$DCG = g_1 + \sum_{i=1}^{10} \frac{g_i}{\log_2 i} \quad (5)$$

で表せる。比較対象にはランダムに並び替えた 3 種類の写真ランキングを用いた。ランダムランキングにおける上位 5 位までと上位 10 位までの DCG を求め、63 サンプルの DCG の平均値を用いて、提案システムとランダムランキングとの間で比較を行った。結果を図 21 に示すように、DCG@5 と DCG@10 のいずれにおいても、提案システムのランキングの方が、ランダムランキングに比べて有意に ($p < .001$) 高くなった。

以上の結果より、提案システムのランキングと主観評価ランキングには関係性があることが示された。このことから、提案システムは、現在提示された写真に対する話題において、ユーザが次に伝えたいと思う写真を優先的に提示可能なことが示唆され、コミュニケーションにおけるユーザの写真へのアクセス支援を期待できる。

しかし、ランキングの精度向上に向けて課題は残されている。1 つ目は、画像処理技術によるオブジェクトの認識精度である。例えば、図 22(a) の写真と、図 22(b) の写真において、システムのランキングと被験者の主観評価との間には大きな差があった。これは、被験者の主

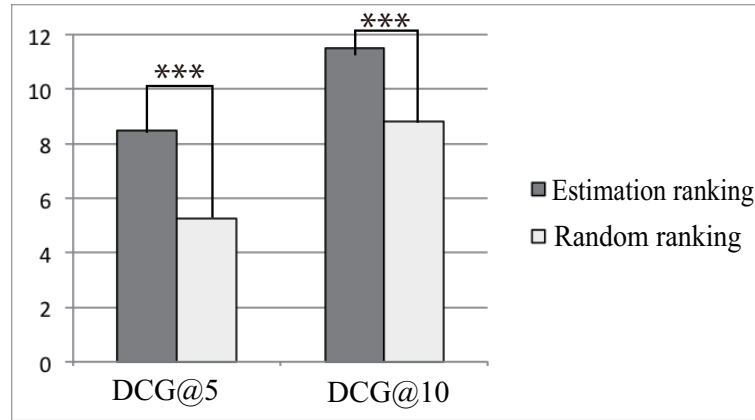


図 21: 主観評価ランキングと提案システムとの DCG 平均値と、主観評価ランキングとランダムランキングとの DCG 平均値

観では2枚の写真の間に「食べ物」としての繋がりを連想し、主観評価が高くなったと考えられるが、画像認識では図 22(b) を食べ物と認識できなかったためだと考えられる。2つ目は、必ずしも関連があるから伝えたいとは限らない点や、話したさの基準に個人差がある点があげられる。図 22(c) と図 22(d) は、ランキング結果の順位が高い一方で、主観評価が低かったときの例である。被験者は、2枚の写真は「食べ物の種類」が異なったため、関連性がなく伝えたいと思わなかった、もしくは、2枚の写真の間に「食べ物」としての関連性を見出したが伝えたいとは思わない、のいずれかの理由で低い評価をつけたと考えられる。

以上のように、コミュニケーションのコンテキストに適した精度の高い写真ランキングの実現には様々な課題はあるが、我々は、詳細な順位よりも、いかに上位に伝えたい写真を含めることができるかが重要だと考える。その理由として、ユーザに対して、短い時間でコンテキストに適した写真を見つけ出させるためには、上位から閲覧をし始めて少ない枚数で目的の写真に到達しなければならないからである。本ランキング手法の評価において、上位と下位との間に差があったことは、上位に伝えたい写真が多く含まれたことを示しており、コミュニケーションにおける現在の話題というコンテキストに基づいた、コンテンツ提示システムの実現可能性を示唆している。

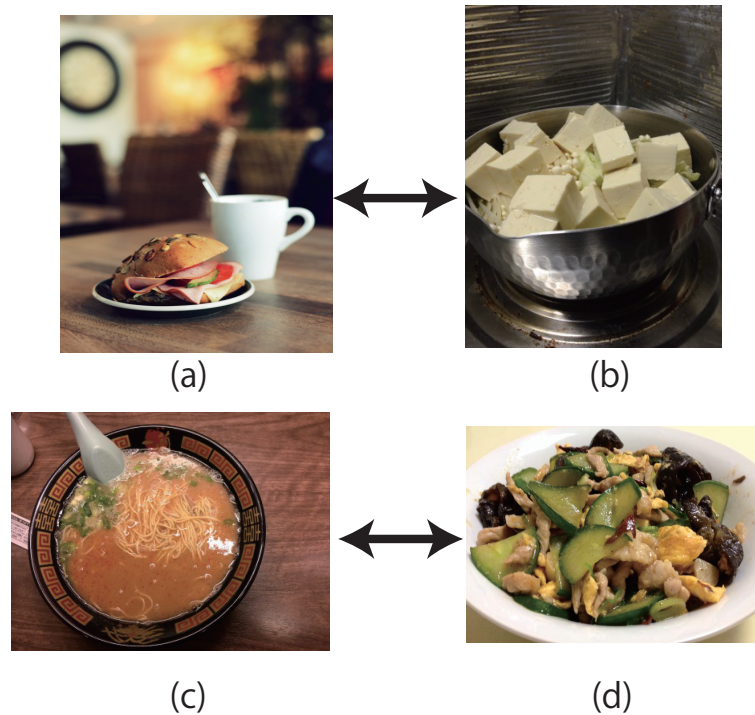


図 22: 主観評価ランキングと提案システムランキングの結果が一致しなかった写真の例。(a) サンドイッチ、コーヒー (b) 鍋、豆腐 (c) ラーメン (d) 食べ物、炒め物

3.3.3 まとめ

Photo Ranking は、コミュニケーションが開始し、ユーザが任意の写真に関する話題について話している、という高次の活動コンテキストに限定した、コンテンツ提示手法である。我々は、現在の話題に関連して、次の話題をユーザが円滑に提供できるよう支援するためのコンテンツ提示のデザインを行った。Photo Ranking では、現在のコンテキストに対して提示する写真の適切さを、現在提示されている写真との類似度として定義することで、ユーザが次に伝えたい話題を含む写真を優先的に提示することを実現している。このように、どんな写真に関する話題か、というコンテキストに限定することで、その瞬間生じるユーザの要求に適したコンテンツを提供できるようになる。

3.4 相互の所有コンテンツを考慮したプライベートコンテンツ提示手法

対話開始前に着目したコンテンツ提示手法 Feelboard の提案と、現在の話題に適したコンテンツを提示する手法 PhotoRanking の提案を行った。Feelboard の実験では、感情状態を利用することで、思い出しやすく伝えたい体験のコンテンツを優先的に提示できることを示した。PhotoRanking の実験では、現在話題としているコンテンツに類似したコンテンツは伝えたい傾向にあることを示した。

しかし、対話開始前においても、思い出しやすく伝えたい体験のコンテンツのみでなく、相手が所有するコンテンツも考慮した、コンテンツ提示が望ましい場合もある。なぜなら、話し手にとって伝えたい体験のコンテンツであっても、聞き手にとってはそれに関連する体験がなく、共感できない場合や、その体験に関連した次の話題がないため対話が続かない場合が考えられるからである。

そこで本節では、対話開始前において、思い出しやすく伝えたい体験のコンテンツに絞り込んだ後に、さらに以下の条件を満たすコンテンツ提示の実現方法について述べる。

1. 両対話者が所有するコンテンツの間で、関連性が高いコンテンツを上位に提示する
2. 聞き手が多く所有するコンテンツと類似したコンテンツを話し手の上位に提示する

PhotoRanking では、1枚の写真が提示されていることを前提としていたため、提示されている写真との類似度を算出して、類似度に基づいて並び替えれば良かった。しかし、対話開始前においては、比較対象が1枚の写真でないため、2枚の写真間の類似度を算出するのではなく、両対話者が所有する全写真間の関連性を考慮する必要がある。

そこで我々は、個々の要素が複雑に絡み合っているようなデータを表現するのに適している、グラフ表現を利用する。つまり、写真をノードとし、写真のタグ情報からノード間のエッジが張られるグラフを考える。また、聞き手が多く所有するコンテンツと関連の強いコンテンツを優先的に提示するために、同一ユーザ内の写真ノード遷移とユーザ間の写真ノード遷移の確率に対して重み付けを行う。

3.4.1 写真グラフの構成

本項では、ユーザ A、B が所有する写真から構成されるグラフ $G = (V, E)$ について示す。まず、ノード集合 V については、ユーザ A、B が所有する写真の集合をそれぞれ V_A 、 V_B とし、 $V = V_A \cup V_B$ と定義する。次に、2つの写真 v_i 、 v_j の間のエッジについて、 $e_{i,j} \in E = 1$ のとき、写真 v_i 、 v_j の間にエッジが張られていると定義する。

写真 $v_i \in V$ に付与された N_i 個のキーワードタグの集合を $T_i = \{t_k \mid k = 1, \dots, N_i\}$ としたとき、 $e_{i,j}$ の値は式 (6) とする。

$$e_{i,j} = \begin{cases} 1 & \text{if exists } (t_k \in T_i, t_l \in T_j) \\ & \text{such that } NGD(t_k, t_l) \geq \theta_{tag} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

ここで、 $NGD(t_k, t_l)$ は、 t_k と t_l の Normalized google Distance (NGD) である。NGD を用いることで全く同じタグでなくても、意味的に強い関連を持つタグが付与された写真間にエッジを張ることができる。なお、 $\theta_{tag} = 1$ としたとき、同じタグが付与された写真間のみエッジが張られる。図 23 に写真グラフ $G = (V, E)$ の例を示す。写真 $v_1 \sim v_5$ はユーザ A が所有する写真、写真 $v_6 \sim v_{10}$ はユーザ B が所有する写真を表す。結ばれた線は、相互のタグ情報の関連性に基づいて張られたエッジを表している。

3.4.2 Random Walk with Restart モデルの応用

グラフのノードの類似度を計算するために、Random Walk with Restart (RWR) [55] を適用する。RWR では問い合わせノード v_q を始点とするランダムウォークを行い、ノードに到達するたびに一定の確率 c で問い合わせノードに戻る。ここで、 \mathbf{p} を $n \times 1$ 行列とし、要素 p_i をノード v_i における存在確率とする。また、 \mathbf{q} を $n \times 1$ 行列とし、問い合わせノード v_q に対応する要素 q_q を 1、その他の要素を 0 とする。そして、 \mathbf{A} を列が正規化されたグラフの

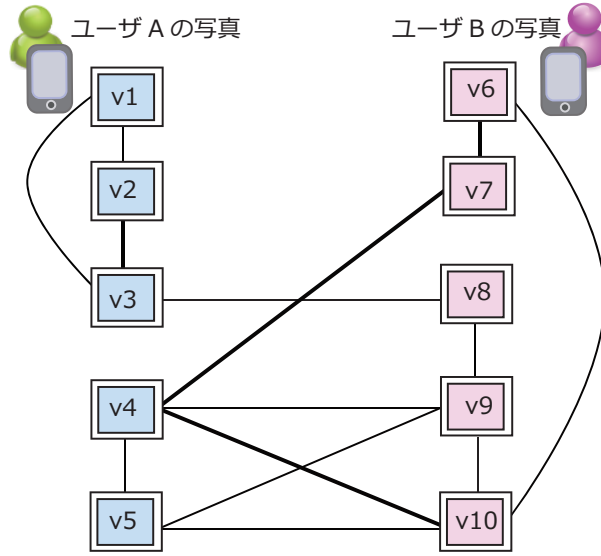


図 23: 写真グラフ

隣接行列とする。すなわち、要素 $A_{i,j}$ は式 (7) に示す様にノード v_j からノード v_i へランダムウォークする確率を表す。

$$A_{i,j} = e_{i,j} / \left(\sum_k e_{k,j} \right) \quad (7)$$

このとき、定常状態における各ノードにおける存在確率 \mathbf{p} は、式 (8) を再起的に収束するまで繰り返すことで計算できる。

$$\mathbf{p} = (1 - c) \cdot \mathbf{A}\mathbf{p} + c \cdot \mathbf{q} \quad (8)$$

対話開始前は、現在提示されている写真が存在しない、つまり、問い合わせノードが存在しない状態であるため、 $c = 0$ としてノードの存在確率 \mathbf{p} を計算する。ここで、問い合わせノード v_q を、対話中において現在提示されている写真とすれば、写真提示の度に、本数式にてノードの存在確率 \mathbf{p} を再計算することが可能である。求められるノードの存在確率を、2者の写真の集合 V の中から次に提示すべき写真のスコアと考え、推薦候補の写真のランキングに用

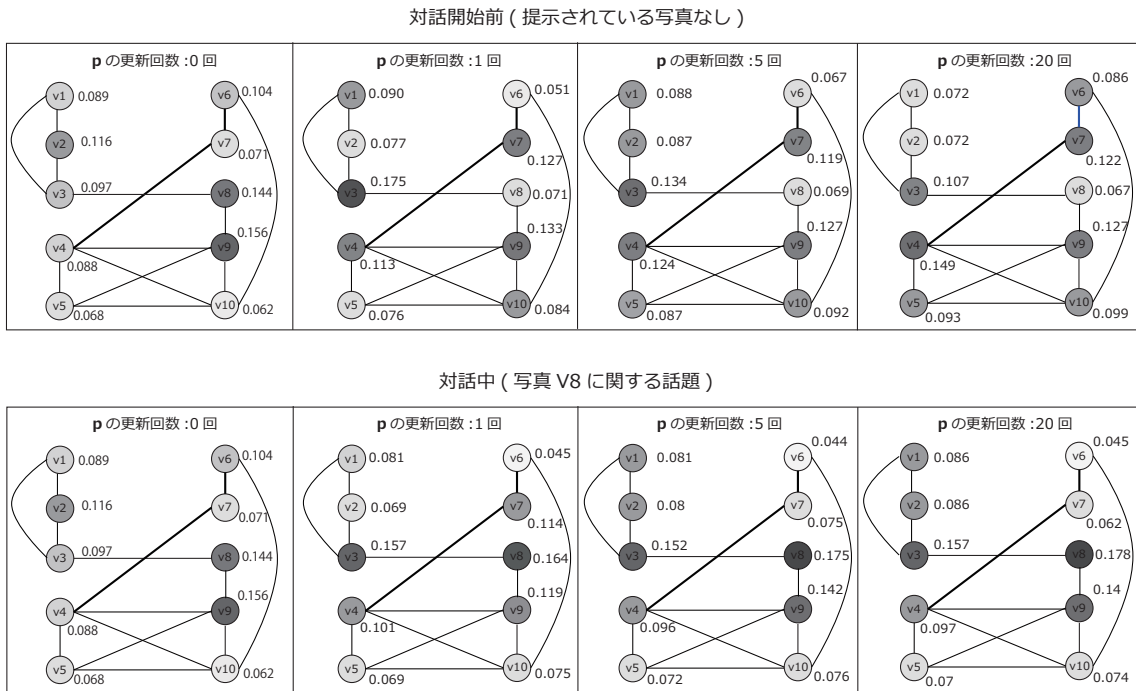


図 24: RWR の実行例

いることができる。

図 23 の写真グラフを用いて p の更新を行った際の、更新回数が 0 回、1 回、5 回、20 回の時のノード存在確率を図 24 に示す。図 24 上は対話開始前に計算されるノード存在確率の更新過程である。まず、対話開始前はランダムウォーク (リスタートなし) によりノードの存在確率 p が計算されるため、 p の更新回数の増加に伴い、他のノードと多くのエッジで繋がっているノード、すなわち両者の共通の話題となりやすい写真 (V_4 、 V_9 、 V_7 など) のスコアが高くなる。次に、図 24 下は、ユーザ B が写真 V_8 を選択して対話が始まった時に計算されるノード存在確率を示す。対話中は、問い合わせノード (V_8) と類似するノードの存在確率が高くなるため、 V_8 とエッジで繋がっている写真 (V_3 、 V_9) のスコアが高まり、ネットワーク上で遷移しにくい写真 V_6 などのスコアが下がる。

上式 (7)(8) では、全てのノード間遷移の確率に対して同じ重みを与えていた。しかしこの処理では、聞き手が任意のタグ情報を含む写真をほとんど所有していない場合でも、話し手

がそのタグ情報を含む写真を多数所有していたら、相互のスマートフォン上にはそれらの写真が上位に提示される。話し手にとっては、自身の興味のある写真を選んだとしても、聞き手の過去の体験に対して関連性の低い写真であるため、対話が盛り上がらない、等と不満足に繋がる可能性が考えられる。そこで、聞き手が多く所有する写真と類似した写真を、話し手の表示画面の上位に提示するために、ユーザ内遷移とユーザ間遷移で重みを変えられるようにする。つまり、式(8)で説明した隣接行列の要素 A で定義されるノード v_j からノード v_i に遷移する確率について、完全なランダムウォークではなく、ユーザ内遷移とユーザ間遷移で重み α を付けることを考える。 $b_{i,j}$ を、

$$b_{i,j} = \begin{cases} 1 & \text{if } [v_i \in V_A \text{ and } v_j \in V_B] \text{ or } [v_i \in V_B \text{ and } v_j \in V_A] \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

として、 v_j から v_i がユーザ間遷移のとき 1、ユーザ内遷移のとき 0 となる値と定義すると、隣接行列の要素は下記のように計算できる。

$$A_{i,j} = \frac{(\alpha b_{i,j} + (1 - \alpha)(1 - b_{i,j}))e_{i,j}}{\sum_k (\alpha b_{k,j} + (1 - \alpha)(1 - b_{k,j}))e_{k,j}} \quad (10)$$

ここで、 α ($0 \leq \alpha \leq 1$) の値を大きくすると、ユーザ間遷移の重みが強くなり、対話相手とより関連の強い写真に関するスコアが高くなる。

3.4.3 動作例

タグ情報に基づいて写真間にエッジが張られていることを前提として、図 25 に示す写真グラフのサンプルを作成した。図 25 の左側はユーザ A の端末が保有する写真集合を表し、右側はユーザ B の端末が保有する写真集合を表している。各写真には、タグ情報として「猫」「魚料理」「パスタ」「旅館」「場所情報(北海道)」「プラモデル」の中から少なくとも 1 つ以上が与えられているとして、共通のタグを持つ写真間にエッジを張った。

以上の写真グラフのサンプルを用いて、式(8)によって各写真の存在確率を計算した。ここ

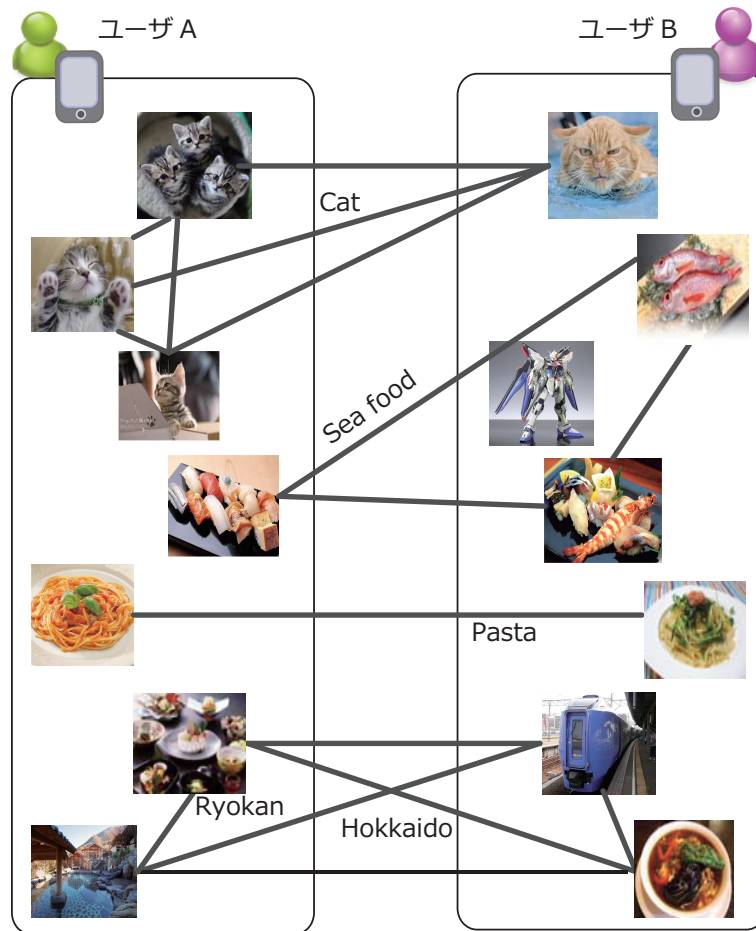


図 25: 写真グラフのサンプル

で、 $\epsilon_{i,j} \sim U(0, \epsilon_{\max})$ とし、エッジが張られていないノード間の遷移確率に対しても微小な値を与えることで、写真ネットワークのエッジが少ない場合でもランキング用スコアを安定して算出することができる。このとき式 (7) は以下の式 (11) のように表される。

$$A_{i,j} = \frac{(e_{i,j} + \epsilon_{i,j})}{(\sum_k e_{k,j} + \epsilon_{i,j})} \quad (11)$$

以上の計算から、存在確率の高い順に写真を並び変えた結果を図 26 に示す。図 26:左は、二

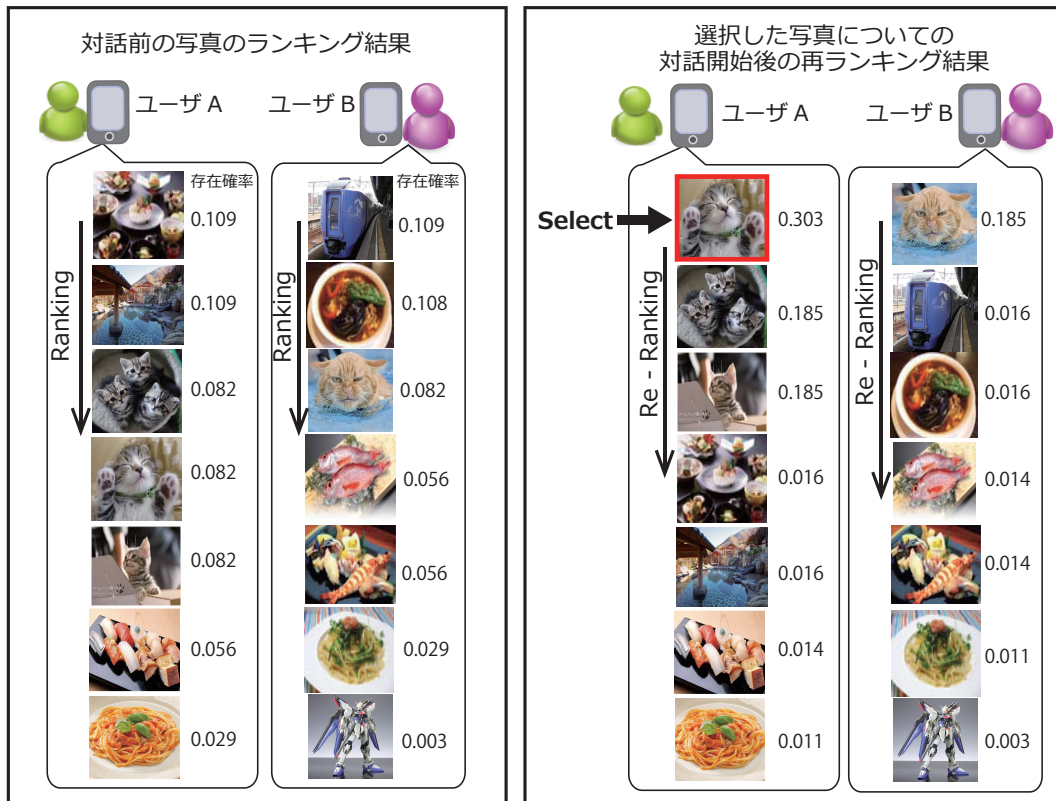


図 26: 対話前と対話開始後におけるスマートフォン A、B 画面上の写真ランキング結果

人のユーザが対話を開始する前の、各スマートフォンの画面上で写真が提示される順番に上から並べたものである。図から分かるように、エッジが多く張られた写真が上位にランク付けられている。ユーザ A の写真の上位には「北海道」での旅館の写真が提示されており、ユーザ B の写真の上位には「北海道」の鉄道の写真とスープカレーの写真が提示されている。これによって、ユーザのいずれかが上位の写真を元に相手に話題を提供すると、互いに北海道に訪れた経験があるという共通点を発見できる可能性がある。一方で「プラモデル」のようにユーザ A の全ての写真と関連のない写真は、ユーザ B の写真の下位に提示され、話題として提供しないことを勧めている。図 26:右は、ユーザ A が「猫」の写真を選びユーザ B と対話を開始した場合の、写真の再ランキング結果を表している。つまり、選択された「猫」の

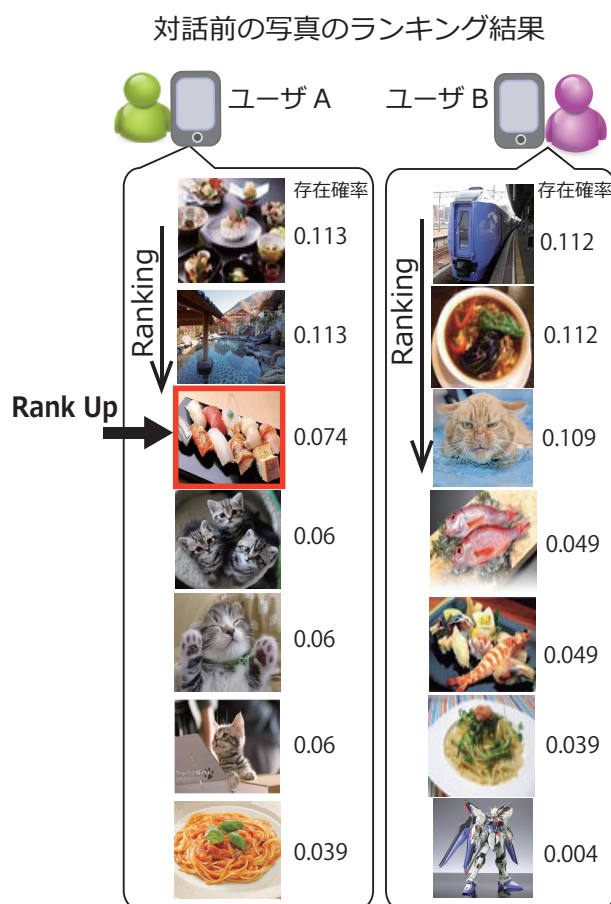


図 27: ユーザ内遷移の重みを 0.3、ユーザ間遷移の重みを 0.7 とした時の、対話前のスマートフォン A、B 画面上の写真ランキング結果

写真を始点ノードとしたランダムウォークによって計算した、各写真の存在確率の結果であり、ユーザ A、B の写真の上位には「猫」のタグが付与された写真が提示されている。このように、話題対象の写真と関連性の高い写真が、相互のスマートフォン画面上の上位に提示されることで、効率的な話題の提供と対話の促進が期待される。

今度は、ユーザ内遷移とユーザ間遷移で重みを考慮した式 (9)(10) に対して、 α を 0.7 に設定して実行した結果を図 27 に示す。図 26 と比較すると、ユーザ A のランキングにおいて、「魚料理」の写真の存在確率が高くなり上位にランク付けられ、一方で「猫」の写真の存在確

率が低くなり順位が下がっている。重み付けを行う前では、ユーザ A が「猫」の写真を多く持っているため、ユーザ A のスマートフォン上では「猫」が上位にランキングされていたが、ユーザ B は「猫」の写真を1枚のみしか持っておらず、ユーザ A 自身の興味を優先したランキングになっていた。しかし、ユーザ内遷移とユーザ間遷移に重み付けを行うことで、ユーザ A のランキングでは、ユーザ B が「猫」の写真よりも「魚料理」の写真を多く持っていることを考慮したランキング結果となった。以上より、本手法ではパラメータの変更を行うことで、自身の興味よりも対話相手の興味を優先したランキングに切り替えることが可能となる。

3.4.4 プロトタイプシステム画面

図 28～31 にプロトタイプシステムの画面を載せる。左側がユーザ A が所有する写真、右側がユーザ B が所有する写真であることを想定する。画面上で任意の写真を選択すると、その写真を問い合わせノードとして各写真の存在確率を計算しランキングを構築する。

図 28 は、対話開始前のときの、ランキング結果画面である。ユーザ B がお酒や鉄道、北海道のカレーうどんの写真を所有していることから、ユーザ A の画面では、カクテルや北海道の旅館の写真が上位に提示されている。一方でユーザ B の画面では、ユーザ A が猫の写真を多く所有していることから、猫の写真が上位に提示されている。図 29 は、ユーザ B が猫の写真を選択したときの結果画面である。ユーザ A の画面の上位にも猫の写真が表示されている。図 30 は、ユーザ B がスイーツの写真を選択したときの結果画面である。ユーザ A、B の画面の上位に、スイーツの写真が表示されている。

3.4.5 まとめ

本節では、対話開始前においても、Feelboard による思い出しやすく伝えたい体験のコンテンツの提示のみでなく、相手が所有するコンテンツも考慮して、両対話者が所有するコンテンツで関連性の高いコンテンツも優先的に提示する方法について示した。また、両対話者が

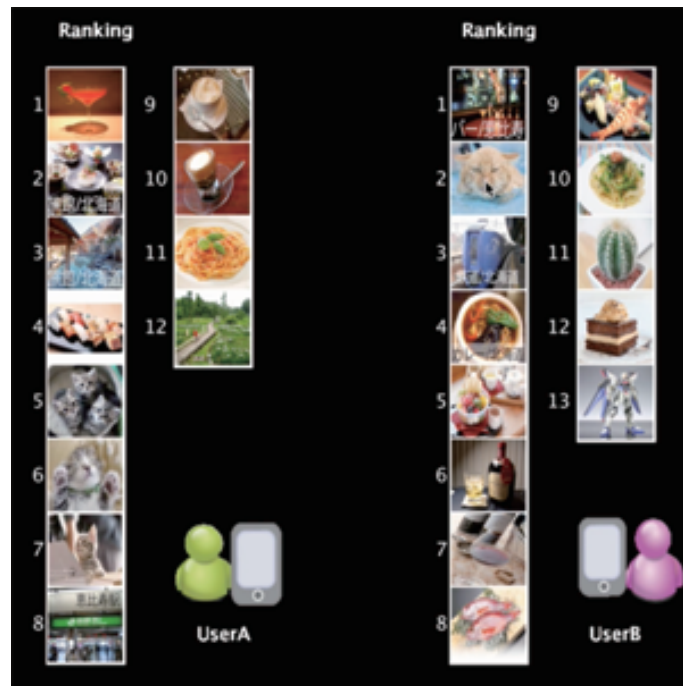


図 28: 対話開始前の表示画面

所有する全コンテンツ間の関連性を考慮するために、コンテンツをノードとするグラフ表現を導入し、RWRを適用することで、対話開始前と開始後においても、一つの計算式にてランキングスコアを算出することを可能とした。ユーザが、思い出しやすく伝えたいコンテンツを重視するか、聞き手の体験との関連性を重視するかを、パラメータとして調整できるように、サービス提供を行う際には設計することが重要である。

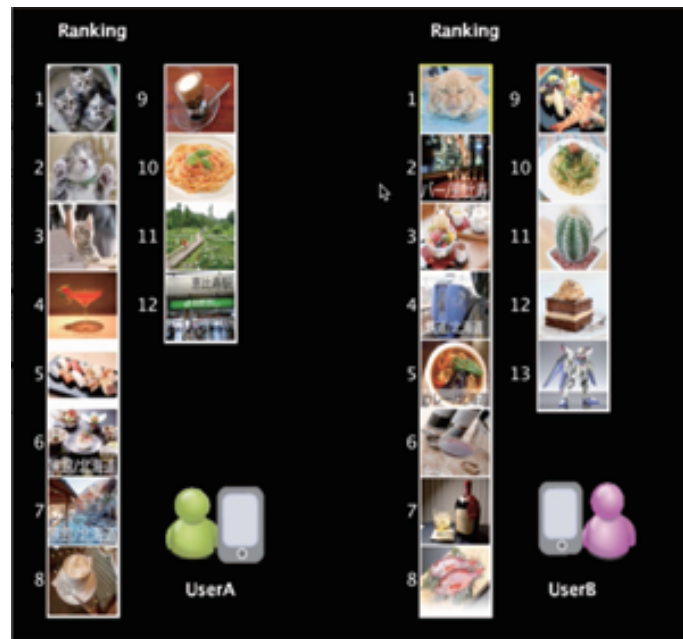


図 29: ユーザ B が猫の写真を選んだ時のランキング結果



図 30: ユーザ B がスイーツの写真を選んだ時のランキング結果



図 31: ユーザ B が旅行の写真を選んだ時のランキング結果

第4章 拡張現実を利用した実空間コンテンツに おける情報獲得支援手法

本章では、人が日常生活において、コンテンツを利用した情報獲得の活動を想定した、コンテンツ提示手法について述べる。人は、任意の目的達成のために様々なコンテンツを参照して情報を獲得している。その際、人は任意の目的に対して適切なコンテンツを選択して閲覧する。参照するコンテンツは、PC やスマートフォン上のデジタルコンテンツのみでなく、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) も含まれる。

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多い。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、情報を融合することで必要な情報を獲得している。その一方で、実空間コンテンツは、可変的でなく新たに情報を付加することができず、融合された情報は可視化されないため、ユーザにとって負担となることがある。

実空間コンテンツを利用した情報獲得の活動支援を行うために、本研究では、拡張現実 (AR) 技術 [7] を利用して、スマートフォン上で複数の実空間コンテンツを重ね合わせる操作”Pick and Lap 操作” (図 32) を提案し、それらのコンテンツを意味的な関連に基づいて仮想的に融合する、仮想融合モデルを提案する。我々は、コンテンツを参照する行為や、複数のコンテンツを相互に参照するときの順番に、ユーザの要求が含まれていることに着目する。例えば、



図 32: pick and lap 操作

店舗紹介を見た後に地図を見たユーザは、その店舗に行きたい、という要求が存在すると推測できる。Pick and Lap 操作では、実空間コンテンツを、スマートフォンのカメラで撮影する (Pick) ことで、何のコンテンツを閲覧しているかという活動コンテキストを認識する。そして、他の実空間コンテンツ上にスマートフォンのカメラでかざす (Lap) と、次にどのコンテンツを参照したか、という活動コンテキストを認識することができる。このとき、Pick したコンテンツと Lap したコンテンツの意味的な関係に基づき、ユーザがどんな情報を求めているかという要求を推定することができる。Lap 操作が完了すると、この2つのコンテンツを、意味的な関係に基づいて仮想的に融合することで、ユーザの要求に合致した形式で情報を提示する。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. スマートフォンのカメラを利用した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. スマートフォン上のアプリケーション・システムとしての実装方法

本研究では、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仮想融合モデルを提案する。すなわち、スマートフォン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。

従来研究においても、ユーザが興味を持つオブジェクトを単位として、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。スマートフォン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」[56]である。のぞき穴ディスプレイでは、スマートフォン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スマートフォン上のディスプレイに表示する手法である。HOTPAPER [57]では、スマートフォン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、スマートフォン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など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スマートフォンで撮影すると、付与されたコンテンツが表示される。ここで、ユーザがスマートフォン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 [58]らは、実空間上の地図にあらかじめ、ドットを格子状に配置し、各格子内の領域を全て画像として登録しておくことによって、スマートフォン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 [59]では、スマートフォン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 [60]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、スマートフォンで覗くと、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づけ、それらをスマートフォンのディスプレイで提示している。このため、あらかじめ対応づけられた、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み合わ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。

4.1 実空間コンテンツを対象とした仮想融合のアプローチ

4.1.1 Pick and Lap 操作

スマートフォン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、スマートフォンのカメラ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 されたコンテンツ」と呼ぶ。一方、Lap 操作は、Pick されたコンテンツを、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 されたコンテンツを重ね合わせる対象となるコンテンツを「Lap されたコンテンツ」と呼ぶ。

例えば、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、様々な単位で捉えられる。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捉えることも、個々の店舗や観光スポットの紹介記事を単位と捉えることも、さらには旅行情報誌自体を1つの単位として捉えることもできる。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、統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をコンテンツ・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。ユーザがスマートフォンをかざした際に、コンテンツ・セグメント上に印を重畳して表



図 33: コンテンツ・セグメントの例

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こと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33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、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34に示す。例えば、旅行情報誌の店舗一覧に記載されている一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、スマートフォンの画面上では、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えることができる。つまり、PickコンテンツとLapコンテンツが意味的に結びつき、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。

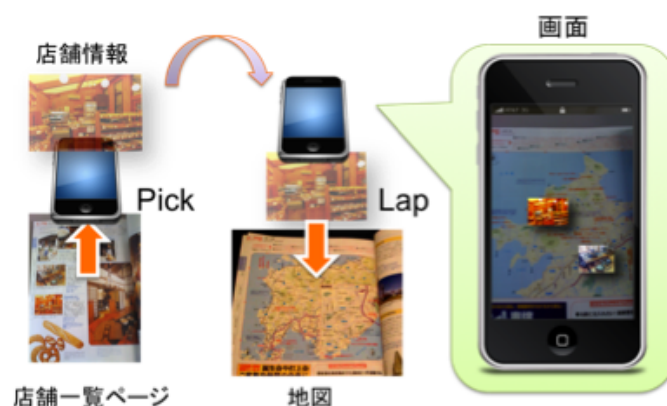


図 34: Pick and Lap 操作に基づく仮想融合の実現例

4.1.2 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 35 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スマートフォンである。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、座席数などの情報である。

本システムは、例えば、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（スマートフォン）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、スマートフォン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。スマートフォンによって取得した代表画像はサーバに送

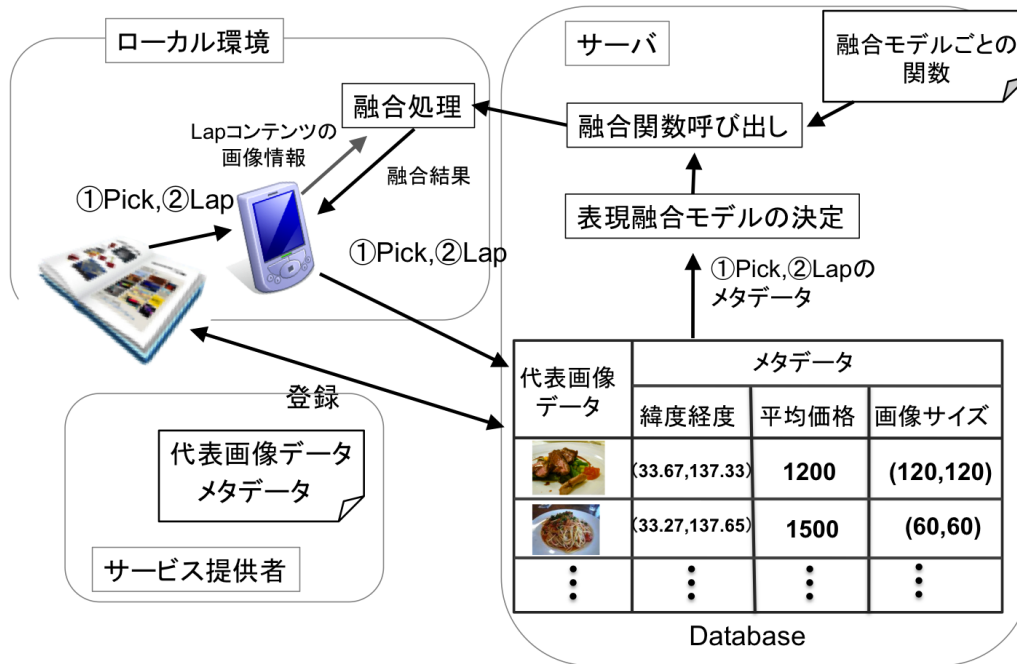


図 35: 処理の流れ

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、クライアントのスマートフォン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、例えば、中居ら [61] はスマートフォン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 [62]。

4.1.3 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、意味的な広がりをもつ画像であり、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。例えば、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づける。例えば、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）を対応づけることができる。地図画像における物理座標と概念の対応付けの例を図36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に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。

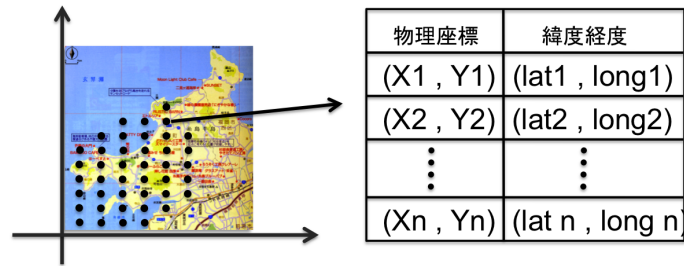


図 36: 地図の物理座標と概念の対応付け

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。例えば、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、粒度など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。例えば、デフォルメが行われた略地図に於いては、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リ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。列挙型によってリ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、粒度の最適な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。本手法では、意味座標に基づいて、スマートフォン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、それがスマートフォン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(12)に示す行列演算として表現できる。

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (12)$$

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、物理座標系に於ける原点 $(0, 0)$ 、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(13)を解くことにより取得可能である。

$$\begin{bmatrix} x_0 & x_1 & x_2 \\ y_0 & y_1 & y_2 \\ 1 & 1 & 1 \end{bmatrix} = T \begin{bmatrix} 0 & 0 & width \\ 0 & height & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad (13)$$

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。例えば、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。例えば、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」を対応付けることができると同時に、「店舗 A」を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。例えば、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。

4.1.4 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み合わせと順番には、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・セグメントの組み合わ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。

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して捉えられるので、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 37 (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、スマートフォンの画面上では店舗 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、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一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 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リ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、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

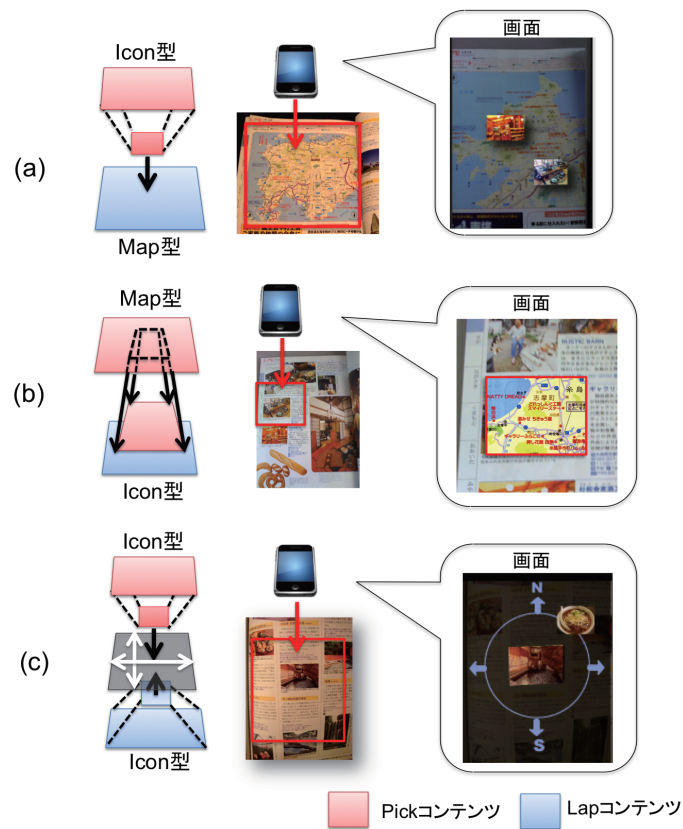


図 37: 融合表現モデルの概念図と具体例: (a) アイコン化モデル (b) クリッピング・モデル (c) アイコン・フィールド・モデル

融合可能となる。

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して捉えられるので、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 37(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。例えば、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。

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して捉えられるので、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える。

図 37(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 の方角が視覚的にわ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、例えば、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み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み合わせで、Pick and Lap 操作を行い、さらに従来のスマートフォン上のサービスでは必要であった入力操作や画面切り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。例えば、ユーザが店舗同士を比較するのは、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、店舗同士を重ね合わせ、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み合わせで、重ね合わせを行い、さらに従来のスマートフォン上のサービスでは必要であった入力操作や画面切り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。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できる。このように、ユーザは目的に合わせた組み合わせで、Pick and Lap 操作を行い、さらに従来は必要であった入力操作や画面切り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み合わせに対しても、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、店舗同士を重ね合わせ、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。

4.2 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して定義する。

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 [63] 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。

前節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一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一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、一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。

4.2.1 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・セグメントである時、ビジュアル・ジョイン演算 VJ を以下の

ように定義する。

$$VJ(c_1, c_2) = \begin{cases} VJ^I(c_1, c_2) & (c_1.i.type = \text{icon} \text{ かつ } c_2.i.type = \text{map}) \\ VJ^C(c_1, c_2) & (c_1.i.type = \text{map} \text{ かつ } c_2.i.type = \text{icon}) \\ VJ^F(c_1, c_2) & (c_1.i.type = \text{icon} \text{ かつ } c_2.i.type = \text{icon}) \\ c_2 & (c_1.i.type = \text{map} \text{ かつ } c_2.i.type = \text{map}) \end{cases} \quad (14)$$

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み合わせによって、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共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。

4.2.2 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。

$$VJ^I(c_1, c_2) = c_r \quad (15)$$

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。

$$c_r.m = VJ_m^I(c_1, c_2) \quad (16)$$

$$c_r.i = VJ_i^I(c_1, c_2) \quad (17)$$

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。

$$VJ_m^I(c_1, c_2) = OJ(c_1.m, c_2.m) \quad (18)$$

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 [63]。

$$R \bowtie_F S = \{t * u \mid t \in R, u \in S \wedge P_F(t, u)\} \quad (19)$$

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。例えば、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於ける結合条件の属性値と、マップ画像のリレーションに於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、この外部結合に於ける結合条件は、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 [64] や R 木 [65] を利用することにより、効率的に処理可能である。

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によって実現する。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^T を以下のように定義する。

$$VJ_i^I(c_1, c_2) = \text{Over}(\text{Conv}(c_1.i, T), c_2.i) \quad (20)$$

ここで、Over は 2 個の 2 次元画像を引数として取り、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、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。

上記の演算で使用される 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 (21) で表現できる。

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} t_x & 0 & \alpha \\ 0 & t_y & \beta \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (21)$$

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 38 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。

$$\begin{aligned} \alpha &= IJ(c_1.m, c_2.m).物理座標.x - c_1.i.w/2 \\ \beta &= IJ(c_1.m, c_2.m).物理座標.y - c_1.i.h/2 \\ t_x &= icon_w/c_1.i.w \\ t_y &= icon_h/c_1.i.h \end{aligned} \quad (22)$$

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、この内部結合に於ける結合条件は、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

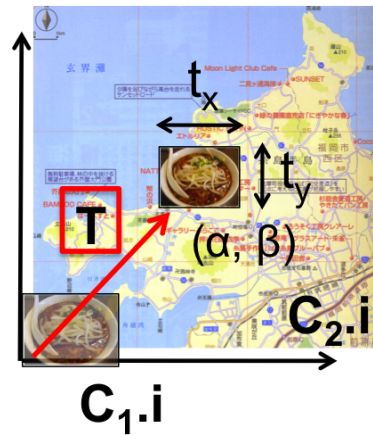


図 38: 代表画像の幾何変換の例

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 の代表画像の幅及び高さを表す。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。例えば、図 39 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が結合され、店舗は地図上の物理座標 $(X1, Y1)$ と対応づけられ、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。

4.2.3 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

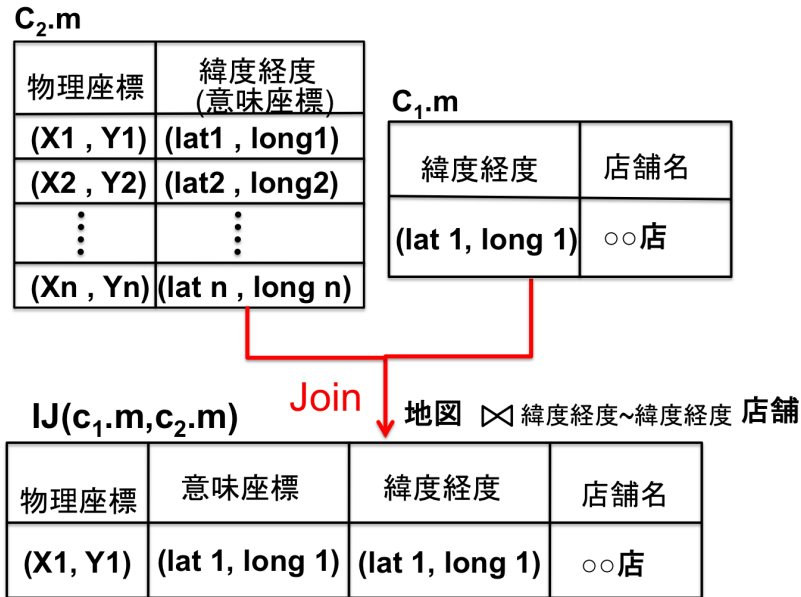


図 39: コンテンツ・セグメントのメタデータの内部結合の例

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。

$$VJ^C(c_1, c_2) = c_r \quad (23)$$

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。

$$c_r.m = VJ_m^C(c_1, c_2) \quad (24)$$

$$c_r.i = VJ_i^C(c_1, c_2) \quad (25)$$

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。

$$VJ_m^C(c_1, c_2) = OJ(c_1.m, c_2.m) \quad (26)$$

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、この外部結合に於ける結合条件は、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。

$$VJ_i^C(c_1, c_2) = \text{Clip}(\text{Over}(\text{Conv}(c_1.i, T), c_2.i), c_2.i) \quad (27)$$

ここで、 Clip は 2 個の 2 次元画像を引数として取り、1 番目の画像を 2 番目の画像の大きさを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。

$$\begin{aligned} \alpha &= -IJ(c_1.m, c_2.m).\text{物理座標}.x + c_1.i.w/2 \\ \beta &= -IJ(c_1.m, c_2.m).\text{物理座標}.y + c_1.i.h/2 \\ t_x &= 1 \\ t_y &= 1 \end{aligned} \quad (28)$$

4.2.4 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共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。

$$VJ^F(c_1, c_2) = VJ^I(c_1, VJ^I(c_2, VJ^C(VM(c_1, c_2, M), c_2))) \quad (29)$$

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通りである、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

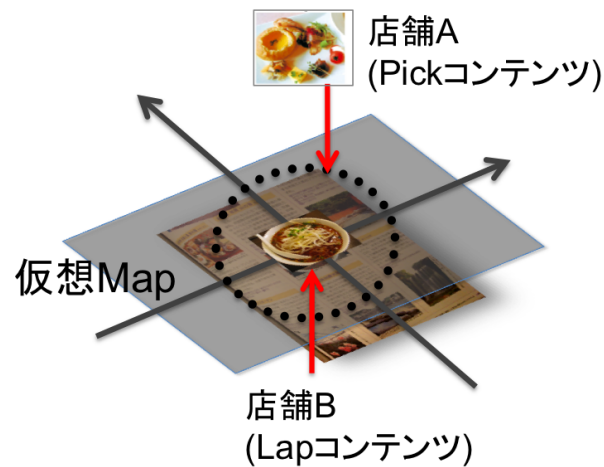


図 40: 方角を表す仮想マップの例

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 40 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す仮想マップ画像が考えられる。図 41 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わかりやすく提示可能である。

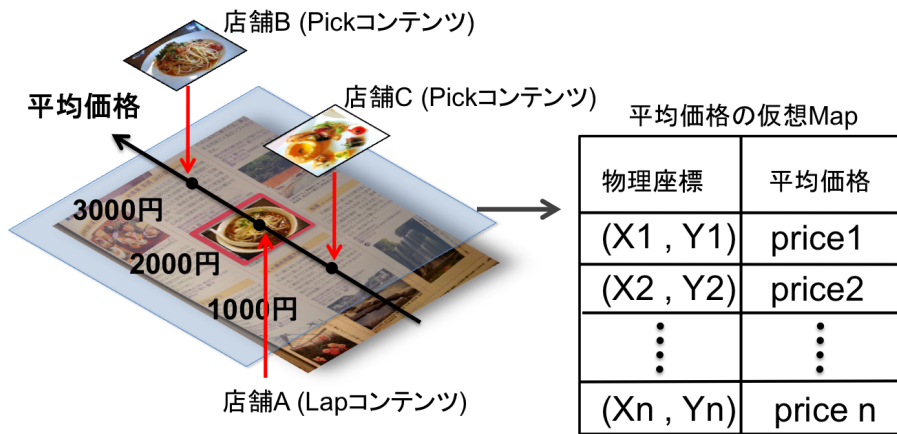


図 41: 平均価格を示す仮想 Map

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内、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仮想マップを自動的に選択することは今後の課題の一つである。

4.3 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、スマートフォン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・セグメントが一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図 42 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 42 の (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。

4.3.1 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、全てのパターンを実行した

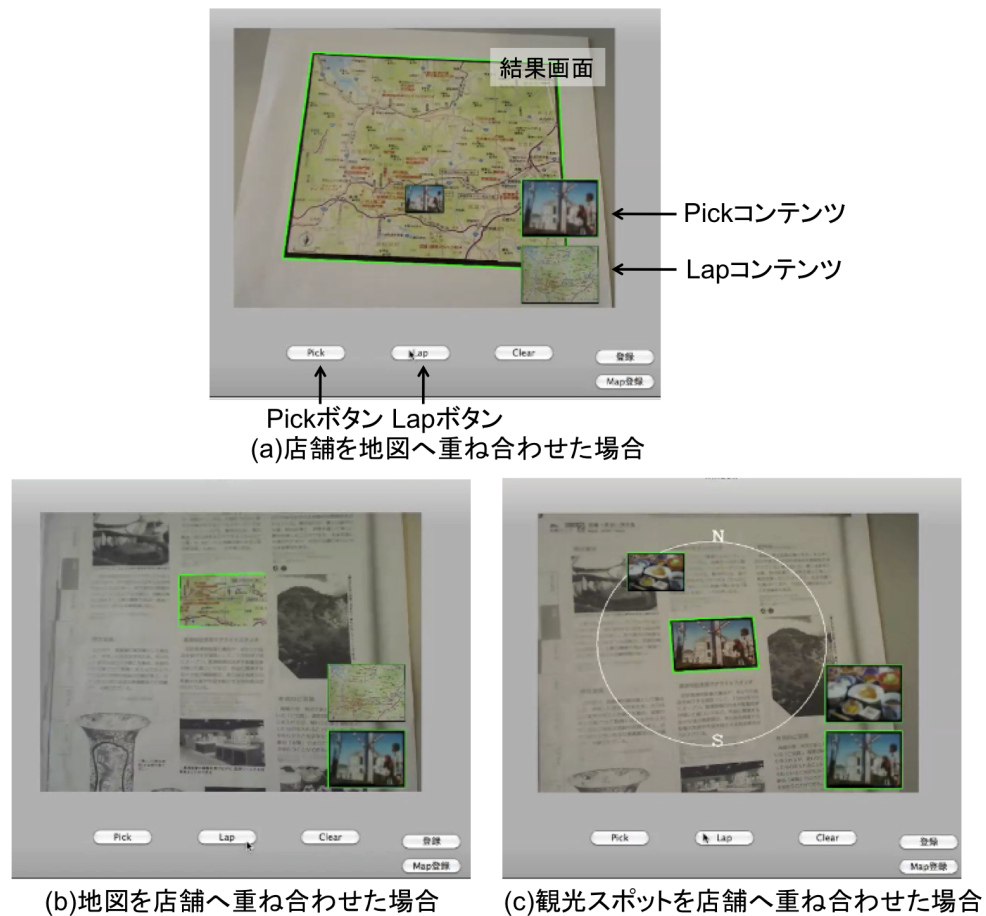


図 42: プロトタイプ・システムを利用したビジュアル・ジョインの実行例

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。

4.3.2 結果と考察

結果を図 43,44,45 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

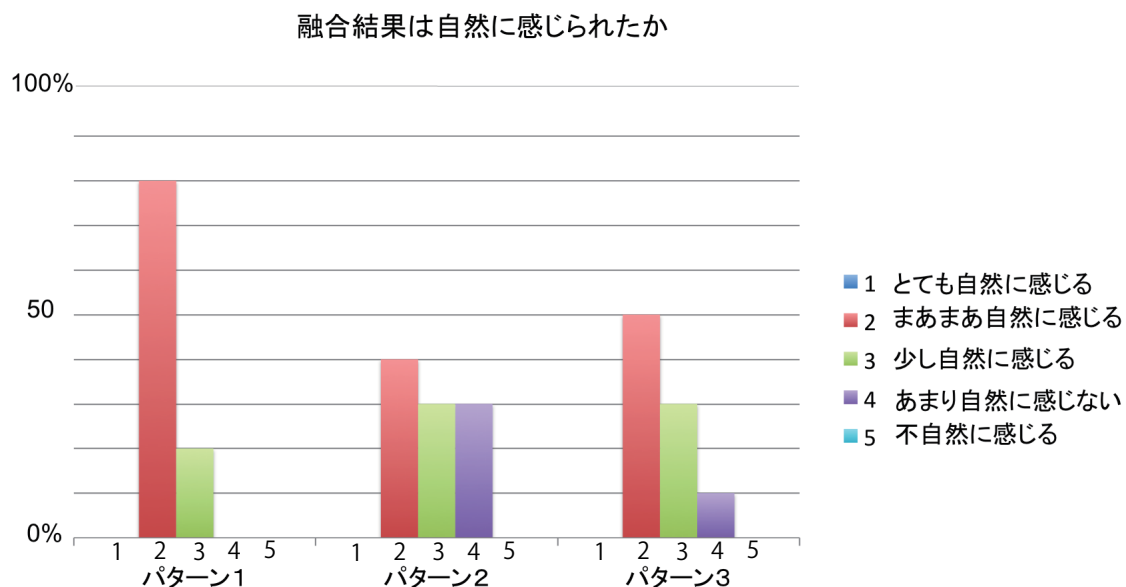


図 43: 各パターンで得られた結果に対する印象の結果

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。しかし、表示の問題以外では、「一回の操作で必要な情報が得られて嬉しい」「見ただけですぐに分かる」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、重ね合わせる操作に対しても自然に感じられる

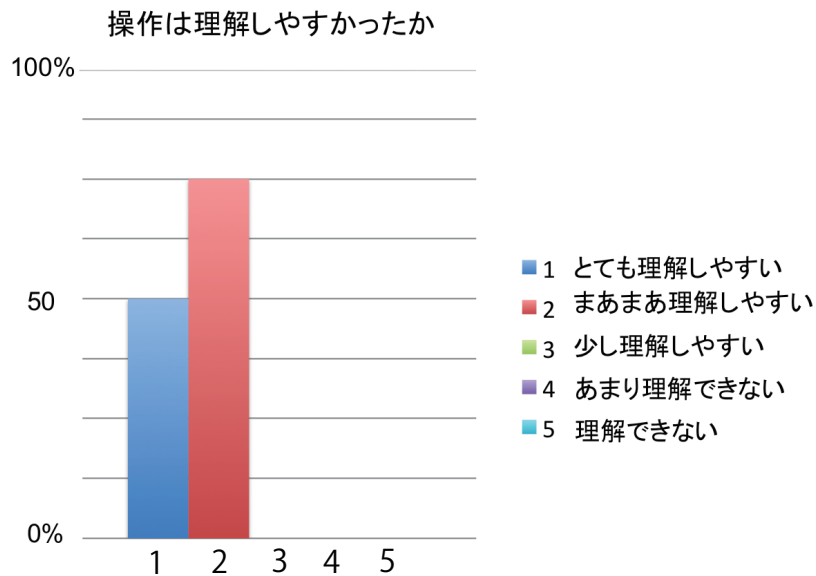


図 44: Pick and Lap 操作の理解しやすさに関する結果

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。

4.3.3 まとめ

本研究では、実空間上における情報取得の活動に限定したコンテンツ提示手法を提案した。実空間のコンテンツから情報を獲得する活動では、人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多い。しかし実空間コンテンツは、可変的でなく新たに情報を付加することができず、融合された情報は可視化されないため、ユーザにとって負担となることがある。そこで情報獲得という活動に対して、スマートフォン上で複数の実空間コンテンツを重ね合わせる操作”Pick and Lap 操作”を提案し、それらのコンテンツを意味的な関連に基づいて仮想的に融合する、仮想融合モデルを提案した。Pick and Lap 操作では、ユーザがどの

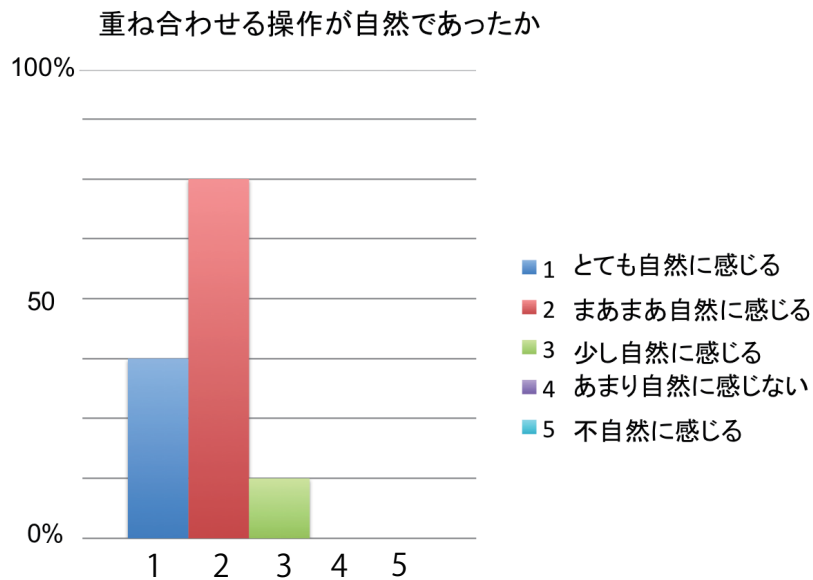


図 45: コンテンツの重ね合わせ操作への自然さに関する結果

コンテンツ同士を比較しているかという、高次な活動のコンテキストを捉えることができる。そのコンテキストから、ユーザが要求する情報を推定し、仮想融合モデルによって要求を満たすコンテンツを提示することで、ユーザが直感的に情報を獲得できるようにした。Pick and Lap 操作と仮想融合モデルは、情報獲得における動的に変化する活動コンテキストに対して、最適なコンテンツ提示を行うことで、ユーザの情報獲得の活動を支援できる可能性を示した。

第5章 結論

近年、通信機能を搭載したコンピュータとしてのスマートフォンの利用者が爆発的に増加している。こうした背景の下、スマートフォンを利用した日常生活における様々な活動の支援が期待されている。ユーザの日常的な活動の支援を実現するためには、コンピュータが自動的にユーザの状況を理解して、状況に応じた適切な処理を行うコンテクストアウェアの概念が重要である。一方、スマートフォンは、Web ページや写真等の様々なコンテンツを利用するために多くの機能を有している。本論文では、スマートフォンが有するこれらの特徴を融合して、ユーザに対して活動の支援を行うコンテンツ提示手法を提案し、有効性を評価した。

ユーザの活動を支援するコンテンツ提示システムでは、システムがユーザの活動状況をコンテクストとして理解し、コンテクストに基づいて適切にコンテンツを提示することが必要である。しかし、従来の多くのコンテクストアウェアに基づくシステムでは、活動を具体性の低いレベル、つまり「低次元活動」のレベルで捉えていたため、活動におけるユーザの目的や意図を十分に考慮することができず、適切なコンテンツを提示して活動を支援できない。本研究では、代表的な実空間上の活動の中から、コミュニケーションと情報取得という二つの活動を対象として、活動を具体性の高いレベル、つまり「高次元活動」のレベルで捉えることで、活動におけるユーザの目的や意図を考慮したコンテンツ提示手法を提案した。

本論文で提案した手法の概要を図 46 に示す。第 3 章では、コミュニケーションを支援するためのコンテンツ提示手法として、Feelboard、及び、Photo Ranking を提案した。Feelboard では、対面コミュニケーションにおいて新たな話題を提供しようとしている、というユーザの意図を含む「高次元活動」を対象にコンテンツ提示手法を提案した。新たな話題を提供する、という活動に着目することで、ユーザは短い時間で伝えたい体験を思い出したい、という要求を持っていると推測することができる。そこで Feelboard では、コンテンツの記録時

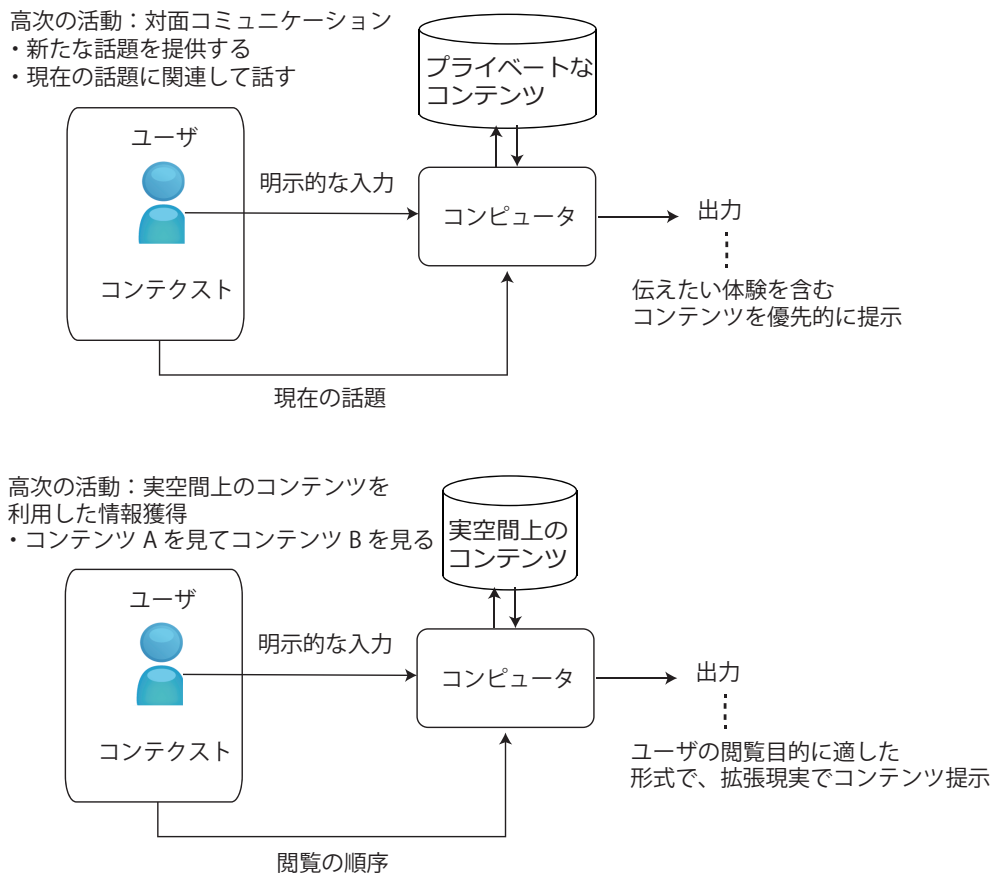


図 46: 本研究で提案した、高次の活動に基づくコンテンツ提示手法の概要

の感情情報を利用することで、思い出しやすく伝えたいコンテンツを優先的に提示することを実現している。Photo Ranking では、対面コミュニケーションにおいて現在の話題に関連して話そうとしている、というユーザの意図を含む「高次の活動」を対象にコンテンツ提示手法を提案した。現在の話題に関連して話す、という活動に着目することで、ユーザは短い時間で、現在の話題に関連した伝えたい体験を思い出したい、という要求を持っていると推測することができる。そこで、Photo Ranking では、現在提示されているコンテンツと類似し、ユーザが伝えたいと思うコンテンツを優先的に提示することを実現している。

第 4 章では、実空間コンテンツからの情報取得を支援するための拡張現実を利用した仮想

的なコンテンツ融合手法について述べ、Pick and Lap 操作、及び、融合表現モデルを提案した。Pick 操作は、ユーザが何のコンテンツを閲覧しているかという活動の内容の把握を可能とし、さらに Lap 操作は、何と何のコンテンツを比較しているかという、より「高次の活動」の内容を把握を可能としている。相互比較している対象コンテンツが分かれば、それぞれのメタデータから、ユーザはどんな情報を獲得したくて比較しているのか、というユーザの目的と意図を推測できる。例えば、店舗同士を比較していた場合は、店舗のコンテンツにはそれぞれ価格情報や評価情報がメタデータとして登録されており、ユーザは価格が安い店舗、もしくは、評価の良い店舗を探したい、という要求を持っていると推測できる。そこで、融合表現モデルでは、ユーザが比較しやすい形式で、スマートフォン上でコンテンツを仮想的に融合し表示する。これによって、ユーザは、直感的な重ね合わせ操作によって、素早く求めている情報を獲得することができる。

本研究で提案した手法によって、スマートフォンを、ユーザが指定したコンテンツを提示する単純な情報提示を行う装置としてではなく、ユーザが行っている実空間上の活動のコンテキストに基づいて、適切なコンテンツを適切な形態で提示することにより、ユーザの実空間上の活動を支援する装置として利用できる。本手法では、コンテキストを高次の活動の概念レベルから捉え、推定されたコンテキストに基づいてコンテンツを提示することで、活動内容に連動した高度な支援を可能とするものである。

今後、センサ技術の発達により、より詳細かつ正確にユーザの高次の活動の状態を推定可能となると期待される。また、ヘッドマウントディスプレイ等の普及により、より一層現実空間と仮想空間の融合が進展していくと考えられる。こうした中で、本論文で提示した、高次の活動の推定とそれに基づくコンテンツの推薦やパーソナライズは今後益々重要になっていくと考えられる。

謝辞

この研究は、九州大学大学院芸術工学府コンテンツ・クリエイティブデザインコース牛尼研究室において、牛尼剛聡准教授の御指導のもとに行われたものであります。

思えば今から9年前の平成20年、学部4年生として牛尼研究室に所属して以来、牛尼剛聡准教授には大変お世話になりました。牛尼剛聡准教授には、研究に向かう姿勢に始まり、研究テーマの設定や解くべき問題に対する考え方、そして論文としてまとめるまでに至り、適切な御助言と多大な御指導を賜りました。この研究活動を通して得られた経験がなかったら、私の物事へ対する見方や考え方は、今とは大きく異なっていたに違いありません。心から感謝の意を捧げます。また、査読をしていただきました、富松潔教授、鶴野玲治教授には、本学位論文に対して多くの有益な御指導、御助言を賜りましたこと心から深く感謝申し上げます。研究議論では、多くの新たな側面からご意見を頂き、研究を深めることができました。

社会人博士として、仕事と学業とを並行して進める中で、日本電信電話株式会社の上司や先輩、同僚、後輩の多くの方々から、研究活動への励まし、そして体調へのお気遣いを頂きました。今の研究は皆様の支えなしには、達成することができませんでした。皆様に深くお礼申し上げます。

最後に、これまで私をあたたく応援してくれた父、いつも美味しい料理を作って家で待っていてくれる母、親友のようにいつも相談に乗ってくれる兄の3人にも、心から感謝の意を捧げます。

参考文献

- [1] 総務省「通信利用動向調査」, <http://www.soumu.go.jp/johotsusintokei/statistics/statistics05.html>
- [2] 坂村健, 痛快!コンピュータ学, 集英社 (2002).
- [3] Google Maps, <https://www.google.co.jp/maps>
- [4] Schilit, B., Adams, N., Want, R.: Context-aware computing applications, WMCSA'94, pp.89-101 (1994).
- [5] Ryan, N. S., Pascoe, J., Morse, D. R.: Enhanced Reality Fieldwork: the Context-aware Archaeological Assistant, in Gaffney, V., van Leusen, M., Exxon, S.(eds.) Computer Applications in Archaeology 1997, (1998).
- [6] Dey, K. A., Abowd, D. G.: CybreMinder: A Context-Aware System for Supporting Reminders, HUC2K, pp.172-186 (2000).
- [7] Azuma, R.T.: A Survey of Augmented Reality, Presence, vol.6 no.4, pp.355-385 (1997).
- [8] Cardiograph, http://macropinch.com/cardiograph_android/
- [9] Viégas, B. F., boyd, D., Nguyen, D. H., Potter, J., J. Donath, J.: Digital Artifacts for Remembering and Storytelling: PostHistory and Social Network Fragments, HICSS'04, p.40109a (2004).

- [10] Hangal, S., Lam, S. M., Heer, J.: MUSE: Reviving Memories Using Email Archives, UIST'11, pp.75-84 (2011).
- [11] Bao, X., Fan, S., Varshavsky, A., Li, K. A., Choudhury, R. R.: Your Reactions Suggest You Liked the Movie: Automatic Content Rating via Reaction Sensing, UbiComp '13, pp.197-206 (2013).
- [12] Trohidis, K., Tsoumakas, G., Kalliris, G., Vlahavas, I.: Multi-label classification of music into emotions, EURASIP Journal on Audio Speech and Music Processing, pp.325-330 (2011).
- [13] 三浦宏一, 浜田玲子, 井出一郎, 坂井修一, 田中英彦.: 動きに基づく料理映像の自動要約, 情報学論コンピュータビジョンとイメージメディア研究会, 44, SIG 9, pp.21-29 (2003).
- [14] 森山 剛, 坂内 正夫.: ドラマ映像の心理的内容に基づいた要約映像の生成, 電子情報通信学会論文誌, Vol.J84-D-2, No.6, pp.1122-1131(2001).
- [15] 中村聡史, 山本岳洋, 後藤真孝, 濱崎雅弘.: 視聴者反応と音楽的特徴量を用いたサムネイル動画の自動生成, WebDB Forum 2012, A2-3 (2012).
- [16] Liu, D., Yan, S. and Zhang, H.: Next photo please: towards visually consistent sequential photo browsing, MM '11, Proceedings of the 19th ACM international conference on Multimedia, pp.973-976 (2011).
- [17] McDuff, D., Karlson, A., Kapoor, A., Roseway, A., Czerwinski, M.: AffectAura: An Intelligent System for Emotional Memory, SIGCHI'12, pp.849-858 (2012).

- [18] Layale, C., Hazem, H.: A Survey of Ground-truth in Emotion Data Annotation, IEEE Int'l Conference on Pervasive Computing and Communications Workshops, pp.697-702 (2012).
- [19] Ramadevi, V., Prasenjit, D., Ankit, S. and Phanindra, P.: The Picture Says It All! Multimodal Interactions and Interaction Metadata, ICMI '11, pp.14-18 (2011).
- [20] Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-Based Collaborative Filtering Recommendation Algorithms, Proceedings of the 10th International Conference on World Wide Web, pp.285-295 (2001).
- [21] Linden, G., Smith, B., York, J.: Amazon.com recommendations: item-to-item collaborative filtering, IEEE Internet Computing, vol.7, pp.76-80 (2003).
- [22] Davidson, J., Liebald, B., Liu, J., Nandy, P., Vleet, V. T., Gargi, U., Gupta, S., He, Y., Lambert, M., Livingston, B., Sampath, D.: The YouTube video recommendation system, RecSys'10 Proceedings of the fourth ACM conference on Recommender systems, pp.293-296 (2010).
- [23] Giordano, D., Kavasidis, I., Pino, C., Spampinato, C.: Content based recommender system by using eye gaze data, ETRA '12 Proceedings of the Symposium on Eye Tracking Research and Applications, pp.369-372 (2012).
- [24] Swingdo, <https://swingmail.co/swingdo/jp/>
- [25] Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, M. N.: Time-aware point-of-interest recommendation, SIGIR'13, pp.363-372 (2013).

- [26] Gao, H., Tang, J., Hu, X., Liu, H.: Content-Aware Point of Interest Recommendation on Location-Based Social Networks, AAAI'15, pp.1721-1727 (2015).
- [27] 八戸観光ナビ, <https://hachinohe-kanko.com>
- [28] 関ヶ原観光 Navi, <http://www.rekimin-sekigahara.jp/main/sekigaharanavi/>
- [29] Hernandez, J., Picard, W. R.: SenseGlass: using google glass to sense daily emotions, UIST'14, pp.77-78 (2014).
- [30] McNaney, R., Vines, J., Roggen, D., Balaam, M., Zhang, P., Poliakov, I., Olivier, P.: Exploring the acceptability of google glass as an everyday assistive device for people with parkinson's, CHI'14, pp.2551-2554 (2014).
- [31] Ishimaru, S., Kunze, K., Kise, K., Weppner, J., Dengel, A., Lukowicz, P., Bulling, A.: In the blink of an eye: combining head motion and eye blink frequency for activity recognition with Google Glass, Proceedings of the 5th Augmented Human International Conference Article No.15 (2014).
- [32] Nike plus, <https://www.nike.com/>
- [33] NTT 技術ジャーナル, 2014.11, pp.16-20 (2014).
- [34] Liu, H., Hu, J., Rauterberg, M.: Follow Your Heart: Heart rate controlled music recommendation for low stress air travel, Interaction Studies, vol.16, pp.303-339 (2015).
- [35] Pan, K.X.J. M., Chang, J., Himmetoglu, H. G., Moon, A., Hazelton, W.T., MacLean, E.K., Croft, A. E.: Now where was I? Physiologically-Triggered Bookmarking, CHI'11, pp.363-372 (2011).

- [36] Robby, C., Ellen, D.C., Susie, S., Edelle, M., Martin, S., Marc, S.: FEELTRACE: An Instrument for Recording Perceived Emotion in Real Time, Proceedings of the ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research, pp.19-24, Textflow, Belfast, 2000.
- [37] Morris, M., Guilak, F.: Mobile Heart Health: Project Highlight, PERVASIVE computing, vol.8, issue.2 (2009).
- [38] マイボイス株式会社, <http://www.myvoice.co.jp/biz/surveys/18615/>
- [39] Nielsen, J.:The need for speed, (1997); Available at <http://www.useit.com/alertbox/9703a.html>
- [40] Brush, A.J., Meyers, B.R.,Tan, D.S., Czerwinski, M.: Understanding memory triggers for task tracking, In Proc. CHI '07, pp.947-950 (2007).
- [41] Harper, D.:Talking about pictures: a case for photo elicitation, Visual Studies, vol.17, pp.13-26 (2002).
- [42] Collins, N.L., Miller, L.C.: Self-disclosure and liking.: A meta-analytic review, Psychological Bulletin, vol.116, pp.457-475 (1994).
- [43] Cahill, L., McGaugh, J. L.: A novel demonstration of enhanced memory associated with emotional arousal, Conscious Cogn, pp.410-421 (1995).
- [44] Jersild, A.: Memory for the pleasant as compared with the unpleasant. Journal of Experimental Psychology, vol.14, pp.284- 288 (1931).
- [45] Hanawalt, N. G., Gebhardt, L. j.: Child-hood memories of single and recurrent incidents, The Journal of Genetic psychology, vol.107, pp.85-59 (1965).

- [46] Katsis, D.C., Katertsidis, N., Ganiatsas, G., Fotiadis, I. D.: Toward emotion recognition in car-racing drivers: A biosignal processing approach, *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, vol.38, no.3, pp.502-512 (2008).
- [47] Kim, J., Andre, E.: Emotion recognition based on physiological changes in music listening, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.30, no.12, pp.2067-2083 (2008).
- [48] Sommerer, C., Mignonneau, L., Lopez-Gulliver, R.: Riding the net, *Siggraph'01*, p.133 (2001).
- [49] 中茂睦裕, 中山彰, 小林稔.: テーブルトップへの画像刺激による遠隔コミュニケーションの分析, *ヒューマンインタフェースシンポジウム 2006 論文集 2*, pp.695-700 (2006).
- [50] Lemstrom, K., Korte, J., Kuusi, P., Kyheroinen, P., Paivakumpu, P.: PICSearch: A Platform for Image Content-based Searching Algorithm, In *The Sixth Int. Conf. in Central Europe on Computer Graphics and Visualization*, Plzen, (1998).
- [51] Lieberman, H., Rozenweig, E., Singh, P.: Aria: An Agent for Annotating and Retrieving Images, *IEEE Computer*, vol.34, pp.57-62 (2001).
- [52] ReKognition API, <https://rekognition.com>
- [53] Cilibrasi, R., Vitanyi, P. M. B.: The google similarity distance, *IEEE Trans. Knowl. Data Eng.*, pp.370-383 (2007).
- [54] Flickr, <http://www.dicta.org.uk/ned/photos/tags/index.html>

- [55] Tong, H., Faloutsos, C.: Center-piece subgraphs: problem definition and fast solutions, KDD'06, pp.404-413 (2006).
- [56] Fitzmaurice, W.G.: Situated information spaces and spatially aware palmtop computers, Communications of the ACM, vol.36, no.7, pp.39-49 (1993).
- [57] Erol, B., Graham, J., Antunez, E., Hull, J. J.: HOTPAPER demonstration: multimedia interaction with paper using mobile phones, Proceeding of the 16th ACM international conference on Multimedia, pp.983-984 (2008).
- [58] Rohs, M., Schoning, J., Kruger, A., Hecht, B.: Towards real-time markerless tracking of magic lenses on paper maps, Proceedings of the 5th Intl.Conference on Pervasive Computing, pp.69-72 (2007).
- [59] Hecht, B., Rohs, M., Schoning, J., Kruger, A.: Wikeye - using magic lenses to explore geo referenced Wikipedia content, Proceedings of the 3rd International Workshop on PERMID (2007).
- [60] Schoning, J., Rohs, M., Kruger, A.: Mobile Interaction with the 'Real World ', MobileHCI'09, no.106 (2009).
- [61] 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).
- [62] 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).
- [63] 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, 昭晃堂発行 (1996).

- [64] Bentley, L. J.: Multidimensional binary search trees used for associative searching, Comm. ACM, vol.18, no. 9, pp.509–517 (1975).
- [65] Guttman, A.: R-Trees: A Dynamic Index Structure for Spatial Searching, Proceedings of SIGMOD'84, pp.47-57 (1984).

業績一覧

【査読付き論文】

- ・ 武田十季, 牛尼剛聡, 鶴野玲治.: 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, 情報処理学会論文誌 : データベース (TOD), Vol.5, No.1, pp.17-30 (2012).
- ・ Takeda, T., Ushiyama, T., Tsuruno, R., Yamamoto, K.: Actual map based interface for browsing content on mobile devices, Proceedings of SIGGRAPH 2009, research poster, Article No.52 (2009).
- ・ Takeda, T., Ushiyama, T.: Private Photo Recommendation System for In-person Conversation, Proceedings of 11th International Conference on Digital Information Management (ICDIM) 2016, pp.271-275 (2016). **【ベストペーパー賞受賞】**