

# 拡張現実感を用いた看板画像からの 店舗情報アクセス手法に関する研究

総合情報学研究科  
知識情報学専攻

インタラクションデザインの理論と実践

北村 茂生

# 目 次

<b>1</b>	<b>序論</b>	<b>1</b>
1.1	本研究の背景	1
1.2	街中における検索行為	2
1.3	本研究の目的	3
<b>2</b>	<b>関連研究</b>	<b>5</b>
2.1	AR を用いた情報提示に関する研究	5
2.2	情報の視認性に関する研究	6
2.3	看板認識に関する研究	6
2.4	物体検出に関する研究	7
2.5	本研究の位置付け	7
<b>3</b>	<b>デザイン指針</b>	<b>13</b>
3.1	これまでの取り組み：減算型表示	13
3.2	対象とする状況	15
3.3	提案システムの要件	16
3.4	提案手法：Search by Snap	17
<b>4</b>	<b>減算型表示を用いたプロトタイプの実装</b>	<b>18</b>
4.1	システムの概要	18
4.2	システムの動作	18
<b>5</b>	<b>リアルタイム看板認識 API の実装</b>	<b>21</b>
5.1	システムの概要	21
5.2	対象とするデータ	21
5.3	看板領域の抽出	21
5.4	抽出した看板の分類	24
5.5	API の実装	26
5.6	フィールド実験	26
<b>6</b>	<b>Search by Snap のプロトタイプ</b>	<b>29</b>
6.1	システムの概要	29
6.2	システムのインターフェース	29
6.3	システムの動作	30
<b>7</b>	<b>減算型表示を用いた評価実験</b>	<b>33</b>
7.1	実験の目的	33
7.2	実験の概要	33
7.3	実験の手順	33
7.4	実験の結果	34

<b>8</b>	<b>Search by Snap を用いたユーザ実験</b>	<b>38</b>
8.1	実験の目的	38
8.2	予備実験	38
8.3	本実験	40
<b>9</b>	<b>議論</b>	<b>47</b>
9.1	得られた知見	47
9.2	本研究の到達点	47
9.3	改善点	48
<b>10</b>	<b>結論</b>	<b>49</b>

# 1 序論

本章では、本研究の実施に至った背景を説明し、対象とする課題を明確にする。

## 1.1 本研究の背景

### 1.1.1 街中における看板

現在、都市には多種多様な視覚情報が溢れている。都市に多く存在する視覚情報の一つとして、屋外広告物である看板が挙げられる。看板の目的は、看板を掲示する場所に存在する店舗や会社の名称の掲示が主である。店舗は不特定多数の人々に対して来店を訴求するために看板を掲示し、人々は目に入った看板に描かれているデザインや文字から何らかの影響を受けて、具体的な行動に移す[25]。例えば、人々が都市空間において店舗を探す際、看板に描かれている内容から、飲食店や衣料販売店などの店舗の種類を判別し、どの店に入るかなどの意思決定を行っている。しかし、繁華街など店舗が多数存在する地域においては、都市環境への配慮が十分でなく、「目立つこと」「インパクトがあること」「興味を引くこと」が優先され、看板やネオンによって埋め尽くされた景観が形成されている場合がある[44]。このような場合においては、看板がひしめき合うことで、逆に視覚情報が見えにくくなるという現象を引き起こしている[43]。そのため、看板が街を行き交う歩行者に対して情報を与えるという本来の役割を果たせていないという問題がある。

また、看板から得られる情報は限定的である。看板から店舗の種類などの基本的な情報は取得できるが、その店舗で享受できるサービスの詳細な内容まで取得することはできない。そのため、見慣れない看板からその店舗の詳細情報を得ることは容易ではない。

### 1.1.2 看板情報の多言語性

近年、日本を訪れる観光客は増加しつつある。日本政府観光局によると、2012年に8,358,105人であった訪日観光客は、2017年には28,691,073人と、3倍を超えて増加している[31]。こうした訪日外国人の増加にも関わらず、店舗の看板などは多言語で記載されているとは限らない。看板情報の多くは地元の言語で書かれており、それが言語障壁を引き起こしている。例えば、日本語が理解できるユーザは、図1.1のような地域においては、目の前の店舗がどのような店舗なのかを理解できることが多い。しかし、図1.2のような地域では、韓国語の文字が読めなければ、目の前にある店舗の種類や享受できるサービスなどの情報を看板から得ることは困難である。

さらに、外国人にとっては、看板に書かれている文字が母国語に翻訳されたとしても、それによって内容を理解できるとは限らない。林田らは、日本に不慣れな外国人が、街の中でどのような情報を必要としているのかを調査するために、日本人と外国人それぞれに情報要求調査実験を行った[12]。その結果、日本人からはあまり出なかった要求情報として、例えば「これが竹林ですか?」のような、自身の持っている知識と現在見ているものとの関連を尋ねるようなものがあった。日本人であれば、「これがあの有名な五条橋?」と考えることはあっても、五条橋と書いてあったり、観光客が多かったり、常識を使って考えることで判断できる場合が多い。しかし、外国の人にとっては新鮮でも、日本人にとって常識であるものが対象であると、

「これが竹林です」という立看板は中々作られない為、受動的に取得できる情報だけでは、知識との照会がスムーズにできない。店舗の看板を例に挙げると、日本語の看板に店舗名がローマ字で併記されていたとしても、日本語が分からなければ、それが店の名前であることが分からない。このように、外国人観光客は日本人の多くが常識として持っている情報を持っていないことも多く、外国人にとってはどの店がどのようなサービスを行っているか、クレジットカードでの決済はできるかなどの情報を看板から得ることは容易ではない。このように、日本人に与えられている情報をただ翻訳して伝えるだけでは不十分である。

## 1.2 街中における検索行為

人々が街中において求める条件に合致した店舗を探す際に用いる情報には、主に(1)店舗の看板や外観から取得した情報、(2)ガイドブックやガイドマップに掲載されている情報、(3)スマートフォン等の携帯端末で検索して得た情報、が挙げられる。例えば、人々が飲食店を探す場合、飲食店の看板や案内はユーザにとって必要な情報であり、それ以外の衣料店などの情報は不要な情報に分類できる。

(1)を用いる場合は、全国展開しているチェーン店などのユーザが見慣れた店舗である場合、情報は容易に取得できるが、ユーザが知らない店舗である場合、看板に書かれている店舗名や外観からのみでは店舗の詳細情報が十分に得られない可能性がある。そのため、「低価格の和食の店」などのユーザが求める条件に合う店舗なのかを素早く知る手法が求められる。

(2)を用いる場合は、選択肢がガイドブックやガイドマップに掲載されている店舗に限られてしまうという制約があるが、その中からユーザが求める条件に合致する店舗を発見し、それに掲載されている情報から、店舗の詳細情報を容易に得ることはできる。しかし、ユーザが慣れていない地域において、ガイドブック等に記載されている地図を頼りに店舗を探し出すことは容易ではない。

(3)を用いる場合は、現状ではスマートフォンなどの携帯端末を用いて、位置情報を手がかりに周辺の検索を行うことで、求める条件に合致した店舗情報を探している。Google Map<sup>1</sup>などの地理情報システムには、住所から地図上の位置を特定する機能があり、飲食店に限定すると、食べログ<sup>2</sup>やyelp<sup>3</sup>、OpenTable<sup>4</sup>などのWebサービスが数多く展開されている。このようなWebサービスには、店舗側が自身の店舗の営業時間や価格などの詳細情報を登録したり、ユーザが店舗のレビューを投稿したりでき、現在地周辺のレストランを検索する機能や、ユーザによるレストランの評価を確認する機能などを持っている。しかし、位置情報を用いるのみではユーザが実際に見ている看板や標識などの現実世界上における正確な位置まで提示できず、ユーザがいる環境と検索行為とが分断されており、その環境から目的の店舗を探す手間が残されている。

このように、現実世界において店舗を探索する際、看板はその店で享受できるサービスの種類を知る上で重要な役割を果たしているが、図1.1や図1.2のように看板が密集している地域においては、一軒一軒の店舗情報を検索する必要があり、目的とする情報の探索に少なからず

<sup>1</sup><https://www.google.com/maps> (2019/1/14 存在確認)

<sup>2</sup><https://tabelog.com> (2019/2/6 存在確認)

<sup>3</sup><https://www.yelp.com> (2019/2/6 存在確認)

<sup>4</sup><https://www.opentable.co.uk> (2019/2/6 存在確認)

時間を要してしまう。そのため、探索する看板が未知であるものや、目立たないものである場合、大量に存在する他の視覚情報に紛れて見つけることができない可能性があり、目的の店舗を発見することは容易ではない。特に、初めて訪れる場所など、慣れていない地域でユーザが看板などの視覚情報を探す場合、特定の情報を素早く見つけることは容易ではないという問題がある。

ユーザがその地域の言語を理解でき、言語障壁がない場合は時間を持つことで条件に合致する店舗を探せるが、ユーザが外国人観光客の場合、自身の目の前にある店舗が自身の求める条件に合致するかスマートフォンで検索しようとしても、言語障壁によりどのように検索して良いか分からぬ場合がある。さらに、位置情報を用いて周辺の店舗情報を検索し、条件に合う店舗が見つかったとしても、その店舗の看板に書かれている文字を読むことができなければ、多数ある店舗の中から目的の店舗を発見することは容易ではない。そのため、結局ガイドマップに記載されている店舗など、限られた選択肢から選択せざるを得ない状況にある。店舗側が看板を多言語化することや、QRコードなどを用いて多言語で情報を配信する手法を探ることでこの問題は解決できる可能性があるが、様々な言語圏からの来訪者すべてに対応するには限界があり、このような手法では店舗側の負担も大きいという問題がある。

### 1.3 本研究の目的

1.2 節で述べた問題を解決するために、本研究では、繁華街や商店街などユーザの周囲に店舗が多数存在する地域を、携帯端末のカメラを通して見た際に、(1) 視覚情報の識別性を向上させ、ユーザが目的の店舗を見つけるのに要する時間を短縮させること、(2) ユーザが慣れていない地域や周囲の文字が読めない状況であっても、目の前にある店舗の情報を直感的かつ簡単に取得できること、の2点を目的とする。

(1)を実現するために、先行研究[7]で提案されたユーザにとって不要な情報を目立たなくさせる手法を改良し、必要な情報に付加情報を重畳表示する提示手法を提案する。さらに、全天球カメラで撮影した画像内において、条件に合う店舗を探索するプロトタイプを実装する。これにより、視覚情報が密集している地域において、ユーザが迷うことなく目的の店舗を発見できるようになることが期待される。

(2)を実現するために、スマートフォンを店舗の看板にかざすことで、地理情報システムのデータベースに登録されている店舗情報を取得し、カメラ映像上に店舗情報を重畳表示するシステムを提案する。大河原らによると、店舗情報を円形ツリーマップ等を用いて視覚化することによって、店舗の全体像の把握が容易になり、素早く店舗を回遊できることが明らかとなっている[33]。看板の認識には多くの画像データが必要であるため、特定の地域を対象とした実環境で利用できるプロトタイプを実装する。これにより、言語障壁があっても、ユーザが目の前の店舗でどのようなサービスが得られるのかを知ることができるようになり、求める条件に合致した店舗を直感的に選択できるようになることが期待される。

本稿ではこれらの提案システムの優位性を検証するために従来手法と比較したユーザ実験を行い、その結果について述べる。



図 1.1: 密集する看板情報（新宿・歌舞伎町）



図 1.2: 密集する看板情報（韓国・明洞）

## 2 関連研究

本章では、関連研究について述べ、本研究の位置付けを明らかにする。

### 2.1 AR を用いた情報提示に関する研究

#### 2.1.1 拡張現実感

拡張現実感 (Augmented Reality; AR) とは、ユーザが見ている現実のシーンに仮想物体を重畳することで、ユーザがいる場所に応じた情報を直感的に提示する技術の総称である [16]。近年、AR を用いたナビゲーションシステムや付加情報提示システムが多数提案されている。AR は主に、Location-based AR と Vision-based AR に分類できる [3]。

#### 2.1.2 AR を用いたナビゲーションに関する研究

AR を用いたナビゲーションシステムに関する研究も行われている。屋外でのナビゲーションは主に Location-based AR が用いられており、それに対して屋内でのナビゲーションは主に Vision-based AR が用いられている。

Location-based AR を用いた屋外ナビゲーションシステムの例として、図 2.1 に示す Blippar 株式会社<sup>1</sup>の ARCity<sup>2</sup>が挙げられる。ARCity は、GPS を用いてユーザの絶対位置を特定し、ARKit<sup>3</sup>を用いて地形を認識することで、目的地までの道のりを示す矢印をスマートフォンの画面に重畳表示している。しかし、このような Location-based AR を用いたシステムは、地図上の目的地に到着した後、ユーザが現実世界上で目的地を探す必要がある。

Vision-based AR はさらに Marker-based AR と Markerless AR に分類できる [34]。Marker-based AR を用いた屋内ナビゲーションシステムの例として、吉野らは、迷いやすい人の特徴を考慮した上で、QR コードマーカーを目印とし、スマートフォン画面上に AR で進むべき方向の矢印を表示する屋内ナビゲーションシステム “DoCoKa” を開発している [45]。Koch らは、出口の案内板など自然なオブジェクトを AR マーカーとしたナビゲーションシステムを開発している [24]。これにより、屋内において煙が探知された際に、作動した煙探知機までの経路を提示し、AR を用いて対処方法を提示している。

Markerless AR を用いた屋内ナビゲーションシステムの研究も行われている。Rehman らは、ウェアラブルデバイスである Google Glass<sup>4</sup>を用いて、図 2.3 のように屋内の構造を 3 次元のポイントクラウドとすることにより、搭載されているカメラでポイントをトラッキングしてナビゲーションを行うシステムを提案している [36]。岩名地らは、ウェアラブルデバイスである SmartEyeglass<sup>5</sup>のジャイロセンサを用いて、歩行者の視点情報に対応した案内情報を生成し、ユーザに提示するシステムを提案している [15]。Gerstweiler らは、RGB-D カメラを用いて屋内環境を 6 自由度でトラッキングし、目的地までの最短経路を算出するアルゴリズムを提案している [8]。

<sup>1</sup><https://www.blippar.com> (2019/2/5 存在確認)

<sup>2</sup><https://itunes.apple.com/us/app/arcity-ar-navigation/id1282527727> (2019/2/5 存在確認)

<sup>3</sup><https://developer.apple.com/arkit> (2019/2/5 存在確認)

<sup>4</sup><https://x.company/glass> (2019/2/7 存在確認)

<sup>5</sup><https://developer.sony.com/ja/develop/smarkeyeglass-sed-e1> (2019/2/7 存在確認)

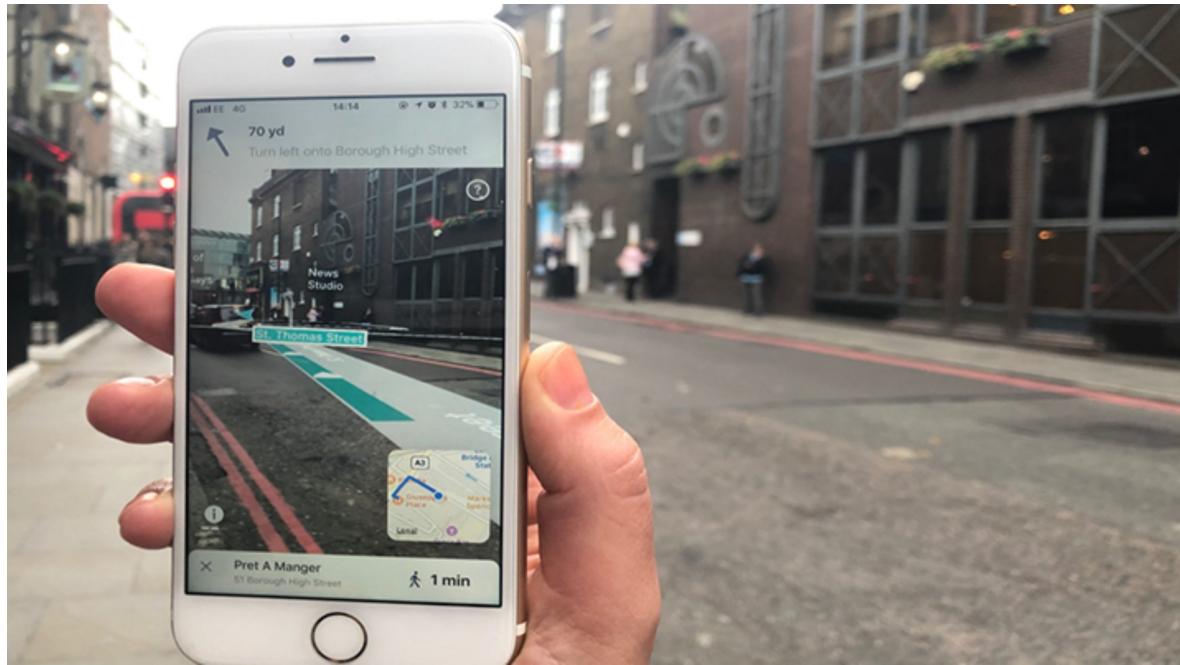


図 2.1: ARCity (脚注 2 より図引用)

しかし、これらの Vision-based AR のみを用いたシステムは、屋内など特定の場所でしか利用できない。

このような携帯端末上での AR を用いたナビゲーションは、紙の地図を用いた場合と比較して、より短い時間と少ない操作で目的地まで辿り着けることが明らかとなっている [37, 45] .

## 2.2 情報の視認性に関する研究

畠らは、画像内に高解像度領域と低解像度領域を作ることによって、ユーザに気づかせることなく視線を特定の領域に誘導させる手法を提案している [11] . 例として、図 2.4-(a) に示す画像をユーザに提示すると、ユーザの視線がどこに集中しているかを示すヒートマップは図 2.4-(b) のようになる。

拡張現実感とは対称に、隠消現実感 (Diminished Reality; DR) の研究が行われている [29] . DR とは、実世界から図 2.5-(a) のように色情報を取り除いたり、図 2.5-(b) のように物体を透視したり、図 2.5-(c) のように特定の物体を置き換えたり、図 2.5-(d) のように特定の物体を消去したりする技術の総称である。DR は主に、屋内外環境の景観シミュレーションなどに用いられている [20] .

## 2.3 看板認識に関する研究

看板を認識する研究や、看板に書かれてある文字を認識する研究は多数行われている [26, 40] . 主な手法としては、ニューラルネットワークを用いて看板に書かれている文字を認識する手法や、Optical Character Recognition (OCR) を用いる手法が挙げられる。He らは、シーケンスラベリング問題として背景から文字を読み取る再帰型ニューラルネットワークを開発してい

る [13] . Kavati らは , 図 2.6 のようにスマートフォンで撮影された看板や標識の写真内の文字を OCR によって認識し , 英語からテルグ語に翻訳してユーザに提示する旅行者向けの Web アプリケーションを開発している [19] . Lee らは , 特徴量を用いてストリートビュー画像から看板の文字領域を検出する手法を提案し , OCR ソフトウェアが文字認識することを容易にしている [28] . Huang らは , 置き込みニューラルネットワークを用いて看板の文字を認識し , 店舗を識別する手法を提案している [14] . しかし , 看板の中には図 2.7 のように手書き文字など崩した文字で書かれているものもあり , 人間であっても読むことが容易でない場合がある . このような場合は , OCR を用いて文字を認識することは困難である .

## 2.4 物体検出に関する研究

物体検出においては , 画像全体から特徴量を抽出できる置き込みニューラルネットワーク ( CNN ) [27] をベースとした R-CNN[9] を用いる手法が現在の主流である [30] . R-CNN では , 前処理として画像の中から物体の候補領域を予め多数取り出し , 各候補領域について CNN で物体の有無を識別することで検出を行っている . 最新の手法では , 図 2.8 のように物体候補領域の生成から識別・矩形抽出まで end-to-end で学習が行えるようになっている [35, 38] .

CNN を用いた活用方法として , ImageNet[5] などの大規模教師付き画像データセットを用いた学習済みモデルが公開されている . このような学習済みモデルを初期値に利用し , 重みデータを変更せずに特徴量抽出機として利用する転移学習 [17] や , 重みデータを一部再学習して特徴量抽出機として利用する Fine-Tuning[30] を用いることによって , 少ない画像データから良い精度で画像のクラス分類や物体検出ができる . Fine-Tuning を用いた例として , 土井は Web 上からラーメン二郎の画像を収集し , ラーメンの画像から店舗を識別するモデルを構築している [6] .

## 2.5 本研究の位置付け

本研究は携帯端末のカメラを通して見た画像の中から店舗を識別し , 情報を重複表示するため , Vision-based AR に分類される . 繁華街などの視覚情報が密集している地域においては , 先行研究 [7] で提案された DR のアプローチを用いて , 不要な視覚情報を目立たなくさせることによって視認性を向上させる . 看板領域の検出には , CNN を用いたアルゴリズムである YOLO[35] を用い , 看板画像のクラス分類には VGG16[41] を用いる .

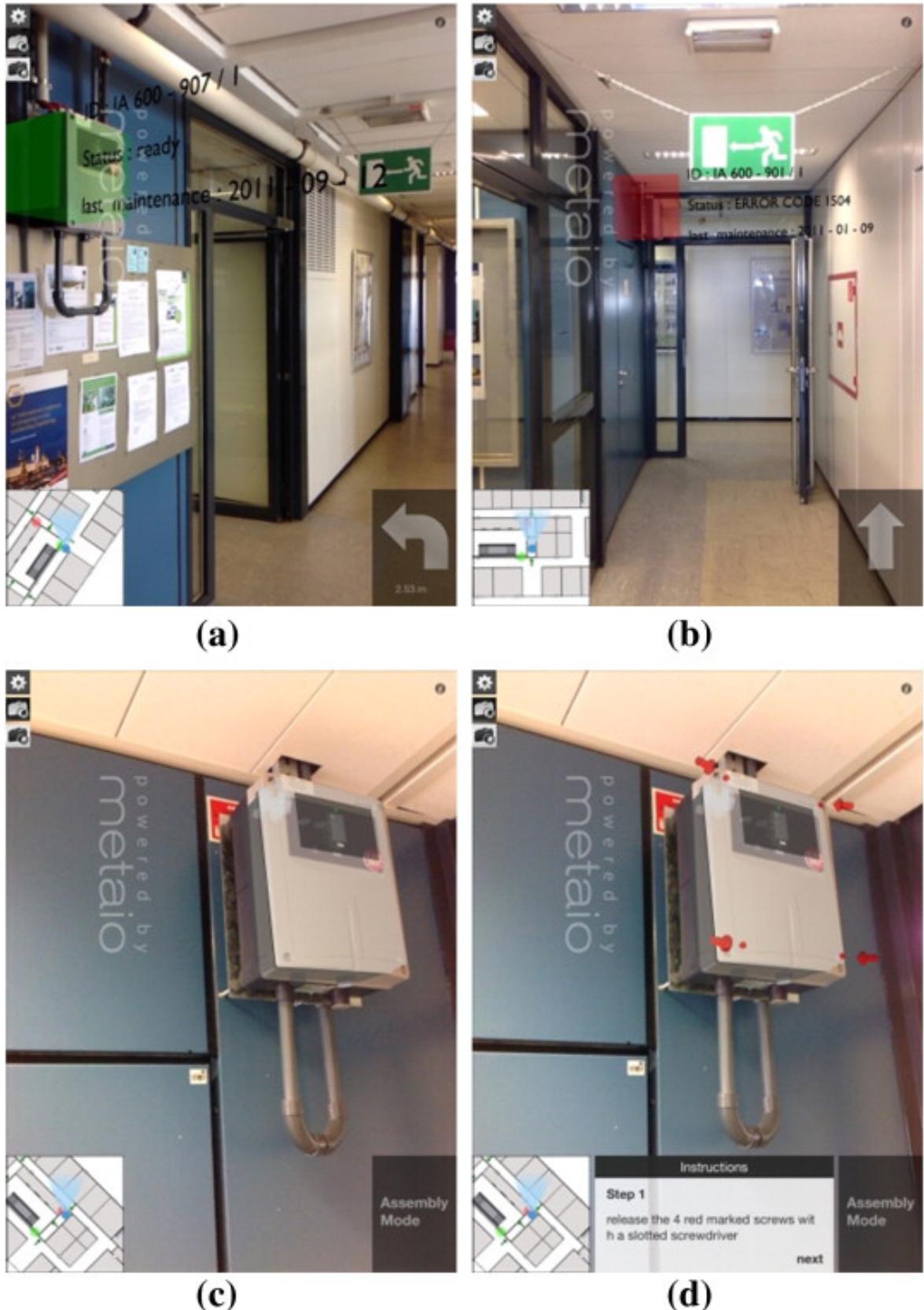


図 2.2: Marker-based AR を用いた屋内ナビゲーションの例（文献 [24] より図引用）

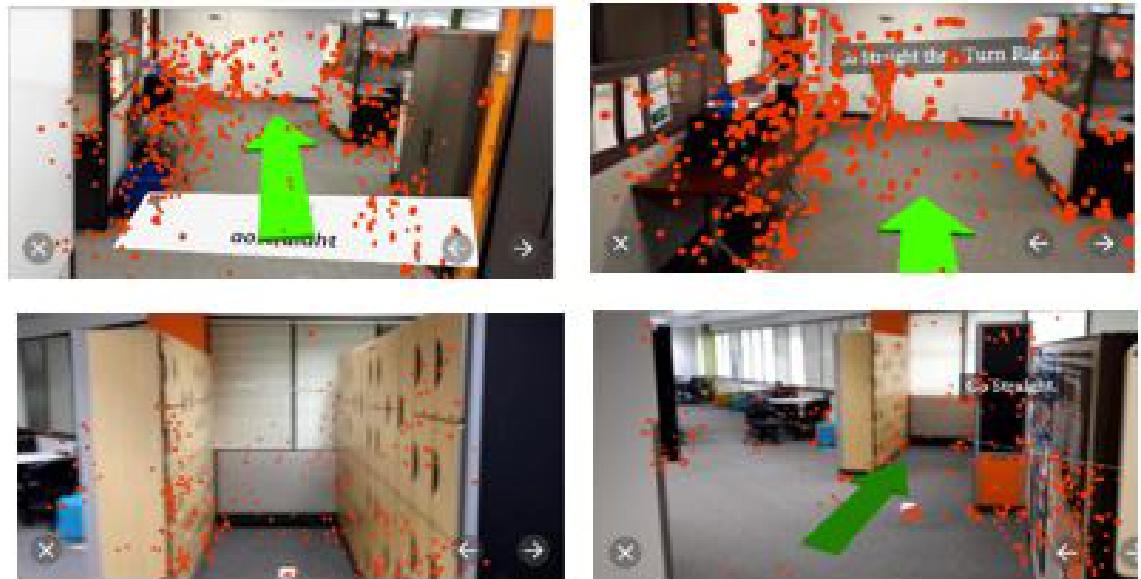
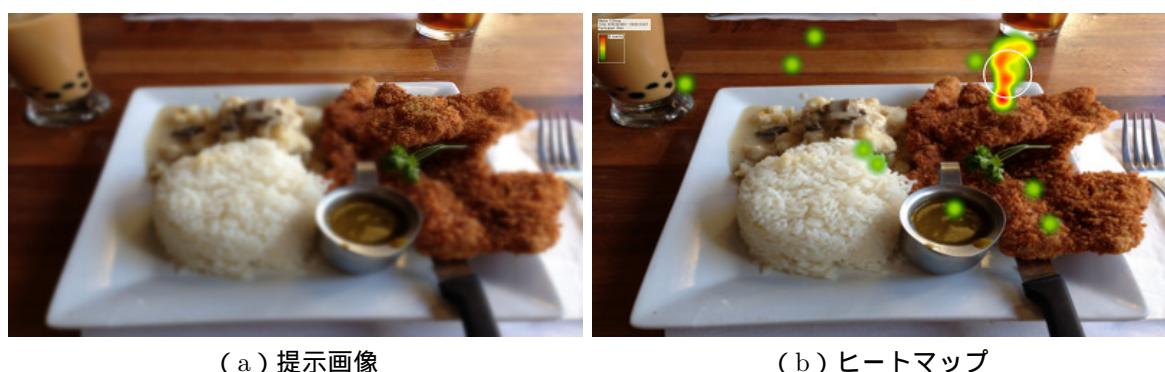


図 2.3: Markerless AR を用いた屋内ナビゲーションの例（文献 [36] より図引用）



(a) 提示画像

(b) ヒートマップ

図 2.4: 解像度制御による視線誘導（文献 [11] より図引用）

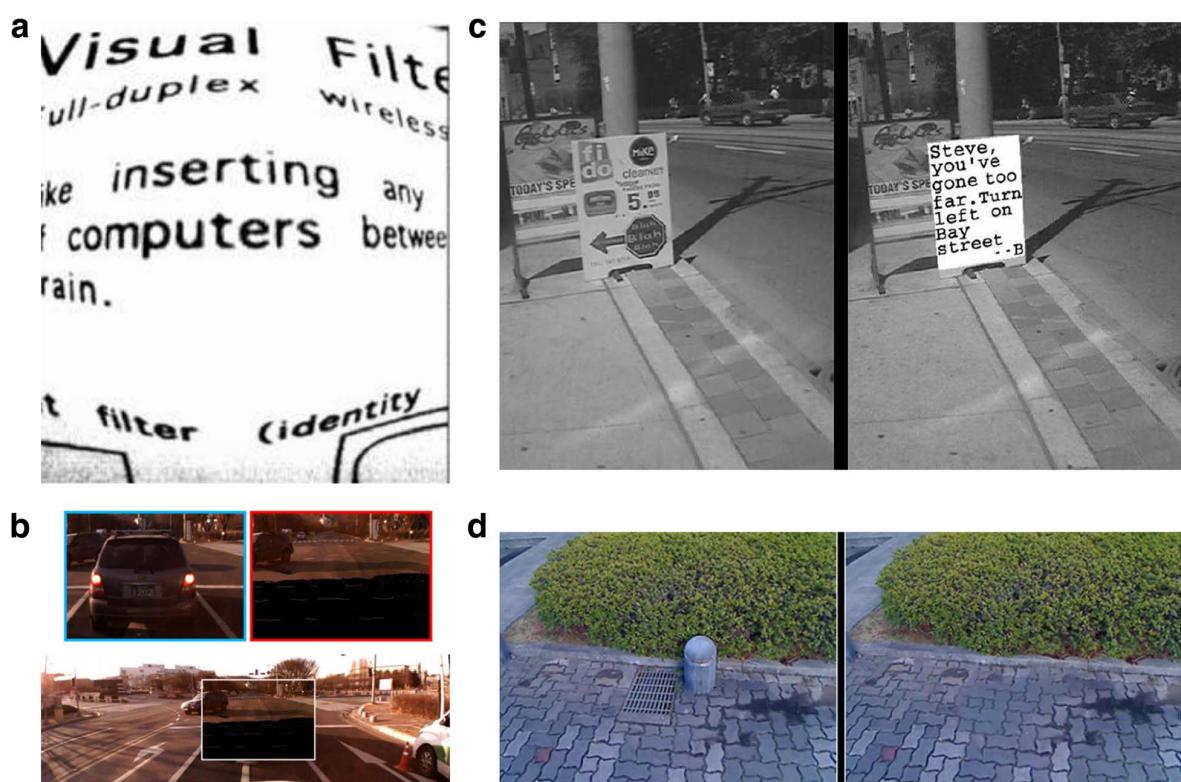
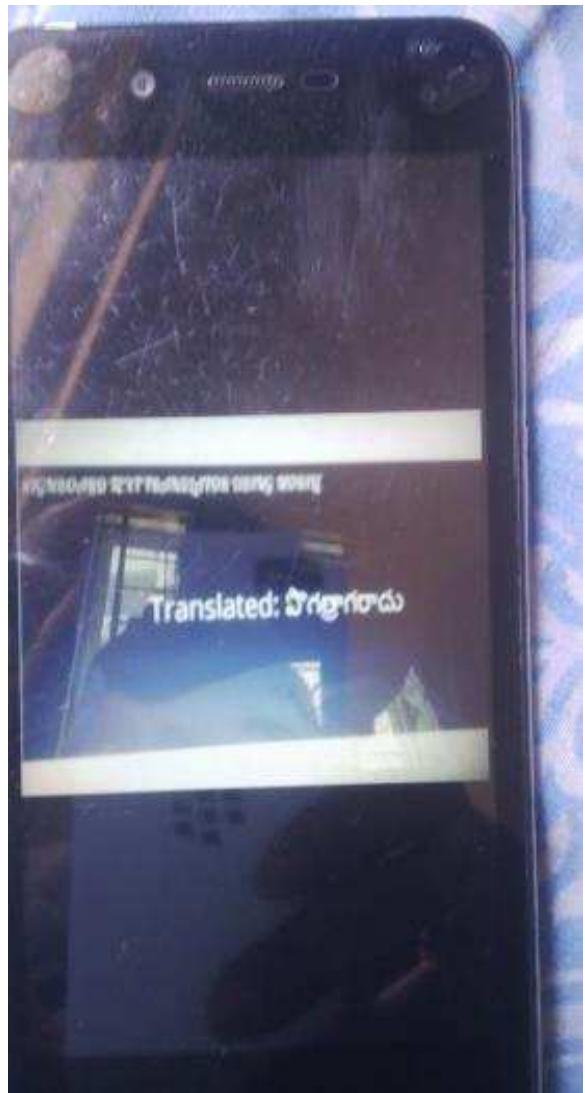


図 2.5: 隠消現実感 (文献 [29] より図引用)



(a) 英語



(b) テルグ語

図 2.6: 翻訳された看板の文字（文献 [19] より図引用）



図 2.7: 文字認識が困難な看板の例（文献 [13] より図引用）

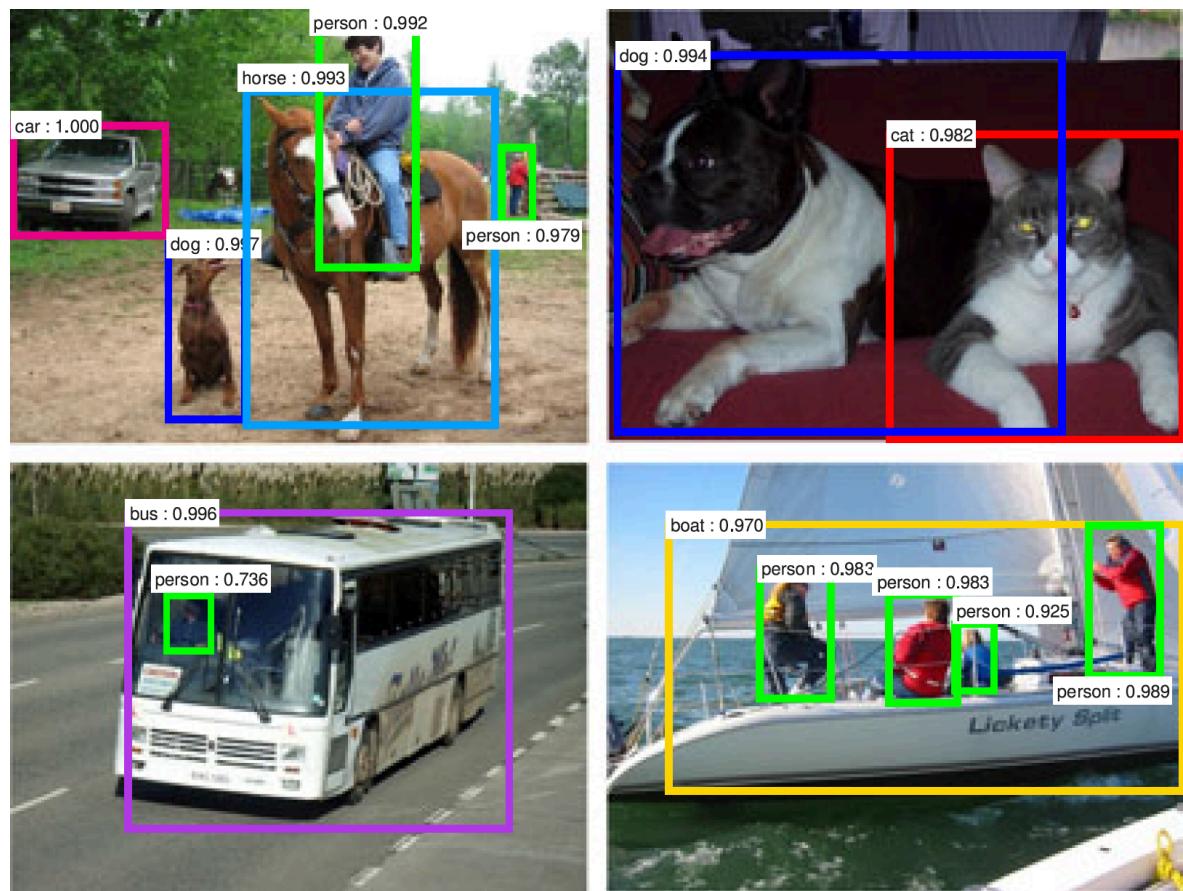


図 2.8: Faster R-CNN による物体検出の例 (文献 [38] より図引用)

### 3 デザイン指針

本章では、提案システムのデザイン指針について述べる。

#### 3.1 これまでの取り組み：減算型表示

1.1.1 節で述べたように、繁華街には看板などの視覚情報が氾濫しているため、どの情報がユーザにとって必要なものかを見分けることが難しい。加藤らは、複数の視覚表示が同時に提示される際、表示数が多くなるほど情報を取得する時間が増加することを示している[18]。このため、視覚情報が溢れている環境の中においては、ユーザが同時に見る情報量を減らすことによって、求める条件に合致する店舗の探索が容易になることが期待される。一方、看板など店舗の外観から得られる情報量には限度があるため、ARを用いて詳細情報を重畠表示することによって、ユーザが看板から取得できる情報量が増加することが期待される。

これまでの研究動向として、藤田らは、ARを用いて情報を重畠することで必要な情報を目立たせる手法を「加算型の情報提示」、DRのアプローチを取り入れることで不要な情報を目立たなくさせる手法を「減算型の情報提示」と位置付け、減算型情報提示手法を提案した[7]。減算型情報提示手法において適切な情報削除の方法を検証するために、藤田らは以下に示す3種類の手法を用いて画像から不要な情報を減算し、情報探索の所要時間を計測する実験を行った。

- 白黒：不要な情報の彩度をなくしたもの
- ぼかし：不要な情報の輪郭情報をなくしたもの
- 白黒とぼかし：不要な情報の彩度と輪郭情報をなくしたもの

実験は、ユーザが街の中で減算型情報提示手法を用いたシステムを使うことを想定し、上記の手法で情報を減算した画像を携帯端末に表示して行った。実験開始前に探索する看板の店舗名のみを実験参加者に教示し、その看板を見つけるよう指示を出した。実験開始から教示した看板を発見するまでの時間を計測した。実験の結果、探索時間は削減方法の差による影響を受けないことが示唆された。また、アンケートで最も良かった減算手法について質問したところ、白黒が過半数を超えた。そこで本研究では、情報の削減手法に白黒を用いて実験を行う。藤田らの実験では、看板の多い環境において減算型情報提示手法の優位性を確認するために、既存の情報にARを用いて情報を加算する従来の手法（加算型情報提示手法）と、DRのアプローチを用いて不要な情報を減算する提案手法（減算型情報提示手法）との比較を行い、それぞれの探索時間の比較を行った。加算型情報提示手法を適用する画像は、画像の中にある探索対象の看板に緑色の吹き出しを追加し、その中に店舗名を表示した。一方、減算型情報提示手法を適用する画像には、探索対象の看板以外の情報を白黒にした。実験から得られた探索時間の結果を比較したところ、減算型情報提示手法に優位性は見られなかった。このため、加算型情報提示手法と減算型情報提示手法には差がないことが示唆された。

減算型情報提示手法の問題点として、不要な情報を減算する際、対象となる看板の色が白黒であり、かつその周辺の景色の彩度が低い場合がある。例として、新日本新地ビル<sup>1</sup>付近で撮影

---

<sup>1</sup>大阪府大阪市北区曾根先新地 1-7-8

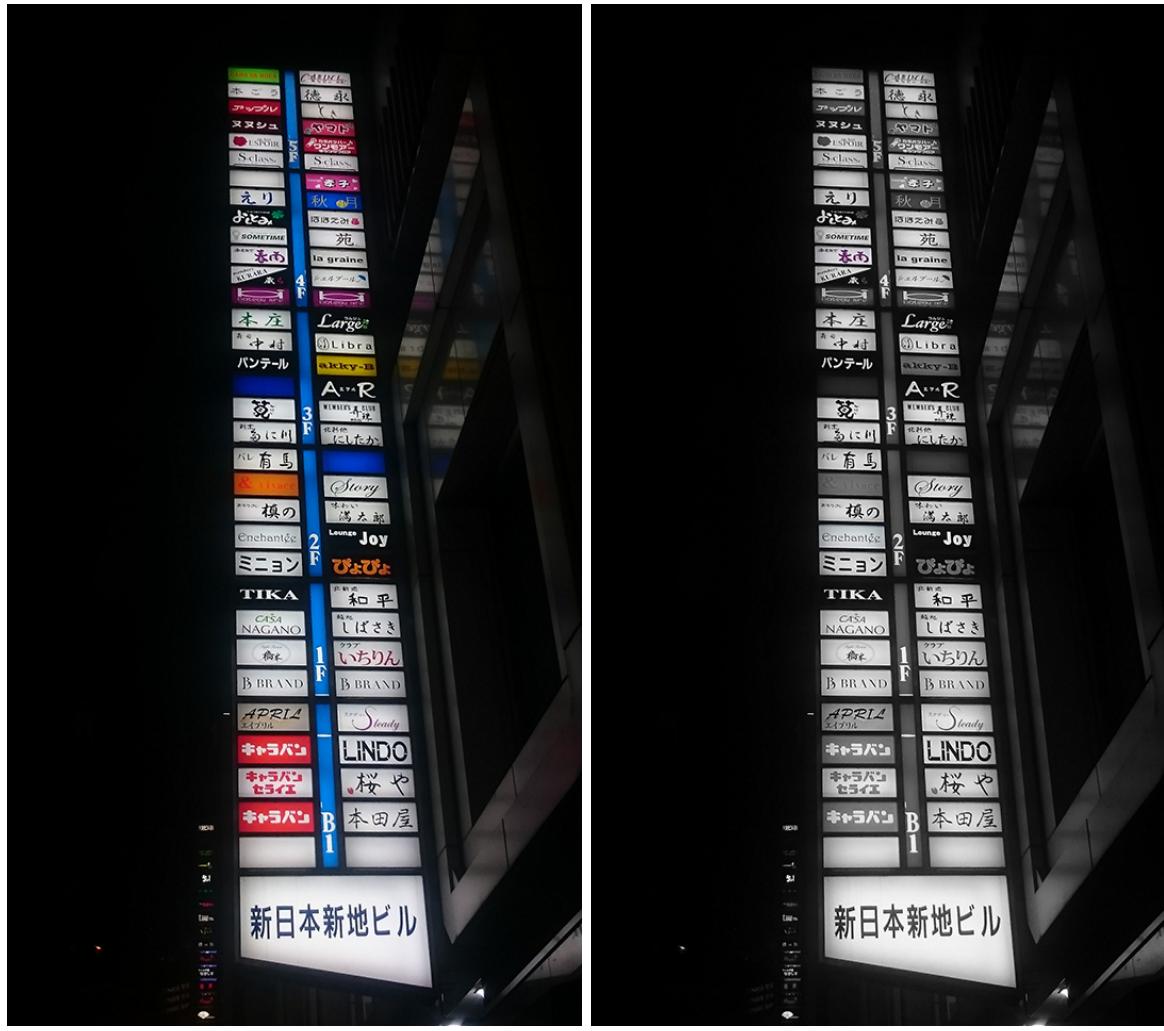


図 3.1: 彩度が低い看板（文献 [21] より図引用）

した図 3.1-(a) の写真から，“春雨”以外の看板情報を減算すると，図 3.1-(b) のようになる。このような場合，探索対象である看板の視覚情報と，減算された周囲の視覚情報をユーザが混同し，減算の効果が減少するという問題が生じる。この問題を解決するために，先行研究 [21, 23] では，藤田らが提案した減算型情報提示手法を加算型で拡張し，加算型と減算型のハイブリッド型情報提示手法を提案した。

### 3.1.1 減算型表示のインターフェース

減算型表示のインターフェースは，文献 [7] で提案された減算型情報提示手法をベースとする。しかし，不要な情報を減算するのみでは，周辺の景色の彩度が低い場合など，環境条件によっては減算の効果が得られないことがある。そこで先行研究 [21, 23] で提案された，AR を用いて文字情報を付加したハイブリッド型情報提示手法用いたインターフェースとする。文字色は背景の色に合わせて変化させる。これにより，ユーザが求める情報のみが分かりやすく提示されるため，上記の問題の解決が期待される。提案システムでは，ユーザが目的の店舗名や店舗の種類をクエリとして入力する。ユーザが入力したクエリに一致する視覚情報はユーザにとって

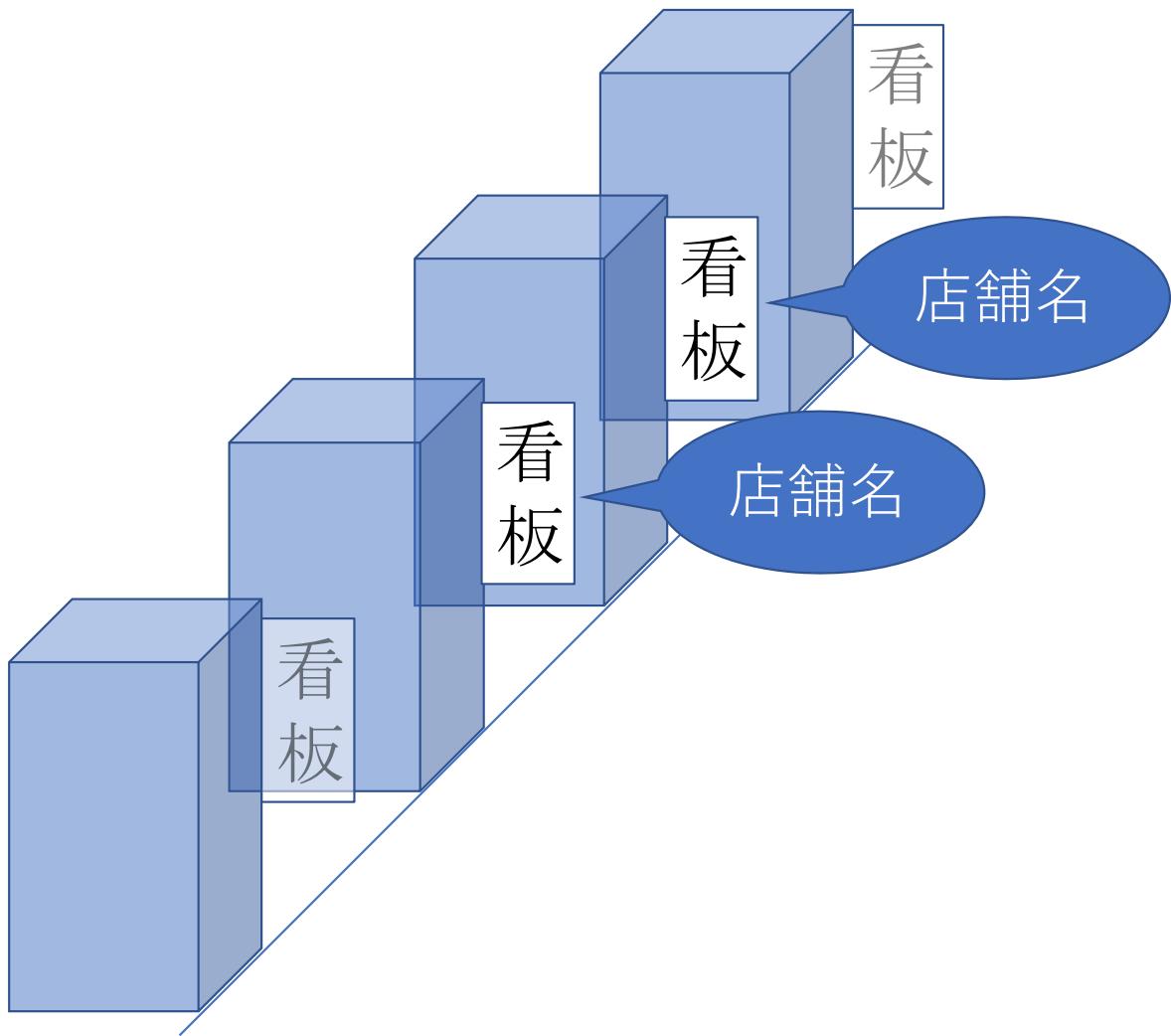


図 3.2: 減算型表示の出力イメージ（文献 [21] より図引用）

必要な情報であり，それ以外の視覚情報は不要な情報である．そのため，情報の識別性向上を目的として，図 3.2 に示すような出力をユーザに提示する．これは，不要な視覚情報をグレースケール化することによって減算し，必要な視覚情報には店舗名などの文字情報を看板の横に重畳して表示したものである．これにより，ユーザは不要な視覚情報を完全に失うことなく，必要な視覚情報のみを得られることが期待される．

### 3.2 対象とする状況

本研究では，1.3 節で述べた目的を実現するために，都市部や繁華街などの店舗が多数存在している地域において，観光客などその地域に慣れていないユーザが携帯端末を手に持って情報を探索する状況を対象とする．本稿で対象とする状況は以下の 2 つである．

- (1) 看板などの周囲の視覚情報が多すぎるため，目的地の看板は目に入っているものの，周辺の過剰な視覚情報に埋没して，素早く見つけることが困難である状況
- (2) 目の前にある店舗の情報が知りたいが，その地域に慣れていないことや言語障壁などの

問題によって、その店舗の詳細情報を知ることが困難である状況

- (1) 対しては、過剰に存在する視覚情報の中から、ユーザにとって不要な情報を低減することによって、情報の視認性の向上を実現する。この手法を“減算型表示”と定義し、視覚情報が密集している地域においてユーザが必要とする情報を素早く発見できるシステムを実装する。
- (2) 対しては、看板画像を認識し、オープンデータを用いて看板から店舗情報を取得することによって、目の前にある店舗の詳細情報を直感的かつ容易に得られるようにする。この手法を“Search by Snap”と定義し、ユーザが店舗の看板に携帯端末のカメラを向けるだけでその店舗の情報が得られるシステムを実装する。

### 3.3 提案システムの要件

3.2 節で述べた問題を解決し、素早く目的の情報にアクセスするために必要な要件として、以下の4点を定義する。

- (1) ユーザが求める店舗情報の探索が直感的にできる操作方法であること
- (2) ユーザがカメラを通して見た店舗の看板と端末の画面上に表示される付加情報とのシームレスな連携が可能であること
- (3) 周囲に店舗が密集している場合においても、ユーザが必要とする看板情報を迷わずに探索できること
- (4) 天候や時間帯などの環境条件に左右されず、ユーザに迷わせることなく情報の提示ができること

本稿で提案するシステムは、モバイルデバイス上で実行することを想定している。システムを実現するにあたり、上述した(1)と(2)を満たすために、

- ユーザが店舗名や店舗の種類をクエリとして検索するために、システムを利用する地域における店舗の種類、名称及びその看板画像の組がデータベースとして用意されていること
- リアルタイムで看板認識を行うために、モバイルデバイスは動画の撮影が可能であり、端末上またはサーバ上でリアルタイムで画像処理が行える環境が存在すること
- サーバ上にあるデータベースにアクセスするために、Long Term Evolution (LTE) や Wi-Fiなどの方法でインターネットにアクセスできること
- 歩きながら端末を操作することは危険であり、ユーザは立ち止まって片方の手でモバイルデバイスを持ち、もう片方の手でシステム操作を行うことが望ましいため、片手で操作できるサイズの端末であること

の4点を前提条件として想定する。



図 3.3: Search by Snap の出力イメージ

### 3.4 提案手法 : Search by Snap

1.2 節で述べたように、ユーザが慣れていない地域や言語障壁のある地域において、目の前にある店舗で受けられるサービスの詳細やクレジットカードでの支払いの可否や、店内での喫煙の可否等の詳細情報を瞬時に把握することは困難である。これらの情報は、店舗の Web サイトや飲食店の情報が掲載されている Web サービス等から取得する必要がある。この問題を解決するために、ユーザが携帯端末に搭載されているカメラを通して目の前の店舗の看板を見ることで、求める情報に効率良くアクセスできるようにする。

#### 3.4.1 提案インターフェースのデザイン

3.3 節で述べた(3)を満たすために、飲食店における予算や提供されている料理の種類、クレジットカードでの決済の可否など、ユーザが求めている条件をクエリとして入力でき、減算型表示のアプローチを用いてユーザにとって不要な情報は目立たなく表示するインターフェースとする。3.3 節の(4)を満たすために、減算型表示と同様に、条件に合致しない看板情報の色調を低減させ、条件に合う看板のみを強調して表示するインターフェースとする。提案インターフェースの出力イメージを図 3.3 に示す。

## 4 減算型表示を用いたプロトタイプの実装

本章では、減算型表示を用いて実装したプロトタイプについて述べる。

### 4.1 システムの概要

減算型表示を用いたシステムのプロトタイプは、モバイルデバイス上で実行できるようUnity (ver 5.6.0f3) を用いてC#言語で実装した。実環境に近づけるために、RECOH THETA<sup>1</sup>を用いて新日本新地ビル付近で撮影した昼と夜の全天球画像（図4.1, 4.2）を用意し、モバイルデバイスを通して見回せるようにした。全天球画像では看板が鮮明に撮影できないため、看板は別に撮影した画像をオブジェクトとして設置した。

実験に用いる看板情報は、図3.1に示す大阪府北新地の新日本新地ビルに掲示されている61枚の看板を用いた。店舗の種類は株式会社キックによる北新地総合情報サイト<sup>2</sup>を参考に、「スナック・ラウンジ」、「バー」、「和食」、「居酒屋」、「クラブ」、「寿司」、「中華料理」、「ドレスショップ」の8種類に分類した。サイトに記載がなかった店舗は、「その他」として分類した。データベースはシステムから容易に参照できるようにするために、リレーションナルデータベースであるSQLite<sup>3</sup> (ver 3.8.10.2) を用いて、画像内の看板の位置、店舗の名前、店舗の種類、看板画像をそれぞれ格納した。

### 4.2 システムの動作

プロトタイプでは、ユーザはモバイルデバイスを全天球画像内で全方向に向けることができる。システムは画面中央のローカル座標と看板のワールド座標を対応させることにより、画面中央と看板を重ね合わせると特定の看板が選択されたことを認識できる。情報提示手法に関しては、

- 加工なし（以下、通常型と記す）
- 看板の横に店舗名を表示（図4.3-(a)、以下、加算型と記す）
- 特定の看板以外の情報をグレースケール化（図4.3-(b)、以下、減算型と記す）
- 特定の看板以外の情報をグレースケール化し、看板の横に店舗名を表示（図4.3-(c)、以下、ハイブリッド型と記す）

の4種類を切り替えられるようにし、探索対象の看板は、店舗の種類別に切り替えられるようにした。プロトタイプのインターフェースを図4.4に示す。また、ユーザがスタートボタンをタップしてから特定の看板を見つけるまでの所要時間を計測できる機能を用意した。ユーザが特定の看板に画面の中央を1秒間合わせると、看板を見つけたと判定するよう実装した。

<sup>1</sup><https://theta360.com/> (2017/4/27確認)

<sup>2</sup><http://kgnet.jp/> (2017/4/27確認)

<sup>3</sup><https://www.sqlite.org/> (2017/4/27確認)



図 4.1: プロトタイプで用いる環境(昼)



図 4.2: プロトタイプで用いる環境(夜)



(a) 加算型

(b) 減算型

(c) ハイブリッド型

図 4.3: プロトタイプの出力画面 (文献 [21] より図引用)



図 4.4: プロトタイプのインターフェース

## 5 リアルタイム看板認識 API の実装

本章では、リアルタイムで看板認識を行うために実装した API について述べる。

### 5.1 システムの概要

カメラを通して見た映像に店舗情報を付与するために、街を撮影した画像から看板を認識し、その結果を返却する API を実装した [22]。システムは(1)看板領域の検出、(2)看板の分類、の2段階の処理を行うことで看板を認識し、画像内の看板の座標及び店舗を JSON 形式で返却する。まず、システムは YOLOv2 [35] を用いて画像の中から看板を囲む長方形の枠線であるバウンディングボックスを検出する。次に、YOLOv2 によって切り出された看板を VGG16 [41] を用いて店舗毎にクラス分けを行う。これにより、少ない教師データから高精度での看板認識が可能になる。リアルタイムで画像認識を行うためには、ハイスペックな GPU を搭載したマシンが必要であるため、スマートフォン等のモバイルデバイスでこれを行うことは困難である。そのため、GPU を搭載したマシン上に Web API サーバを構築し、モバイルデバイスから一定ミリ秒毎にこの API を呼び出すことによって、リアルタイムでの看板認識を実現する。

### 5.2 対象とするデータ

本稿では、大阪府高槻市の JR 高槻駅と阪急高槻市駅の間に位置する商店街の一区域を対象としたプロトタイプを実装する。プロトタイプでは、図 5.1 の赤線部で示されている高槻本通の 100m 区間において、オープンソースの地理情報システムである OpenStreetMap (以下、OSM と記す) [10] に 2018 年 10 月 1 日時点でデータが存在した、(1) 赤から 高槻店、(2) メサベルテ 高槻店、(3) 高槻ちやぶちやぶ、(4) 炭焼酒場 森田屋 高槻店、(5) 全席個室居酒屋 北海の恩返し 高槻店、(6) 肉丼専門店 高槻肉劇場、(7) 磯丸水産 高槻店、(8) おだいどこはなれ 高槻店、(9) ぢどり亭 高槻店、(10) ビリヤード・ダーツ & Food Bar OzBuddy、(11) 小だるま JR 高槻駅前店、(12) 焼肉・しゃぶしゃぶ食べ放題 ふくふく 高槻店、(13) 甲南チケット 高槻本通店、(14) セブン-イレブン 高槻高槻町店、(15) 駿河屋、の 15 店舗を対象とする。

### 5.3 看板領域の抽出

YOLOv2 は 9,000 種類を超えるオブジェクトの領域推定と分類が可能であるため、YOLOv2 のみを用いても看板領域の検出及び分類は可能である。しかし、将来的に様々な商店街に対応させるためには、対象とする全ての店舗の看板画像を学習させ 1 つのモデルを生成する必要があるため、汎用性に欠けるという問題がある。そのため、学習させるオブジェクトのクラスは “signboard” の 1 クラスとし、YOLOv2 は看板領域の検出のみに用いる。

#### 5.3.1 アノテーション

看板領域をアノテーションするために、画像内に存在するオブジェクトのバウンディングボックスを容易にラベリングできるツールである BBox-Label-Tool<sup>1</sup> と Python<sup>2</sup> 2.7.14 を用いる。

<sup>1</sup><https://github.com/puzzledqs/BBox-Label-Tool> (2019/1/26 存在確認)

<sup>2</sup><https://www.python.org> (2019/1/26 存在確認)



図 5.1: 対象とする区域

学習に用いる画像は，対象とする地域において複数の店舗が写るよう様々な角度から撮影した650枚の写真を用いる。各画像について，対象とする店舗の看板の左上と右下の座標，クラス“signboard”をそれぞれ人手でアノテーションした。アノテーションの例を図5.2に示す。その後，アノテーションしたデータをYOLOv2で用いられているフォーマットに，Python 3.7.0とスクリプト<sup>3</sup>を用いて変換する。これらのデータの90%の585枚をトレーニングセットとして，10%の65枚をテストセットとして無作為に分類した。

### 5.3.2 学習

YOLOv2では，入力画像を $S \times S$ のグリッドに区切り，グリッド毎に，ある一定のアスペクト比を持つ $B$ 個の矩形領域の中心座標 $(x, y)$ ，及び矩形内に物体が存在する確率を予測する。さらに，各矩形に何らかの物体が存在するとき，その物体がどのクラスに属するかを示す事後確率も予測する。矩形領域，物体の存在確率，クラスの事後確率の予測を，YOLOv2では一つの損失関数に統合している。矩形領域の中心座標と大きさに関する損失関数を $Loss_{box}$ ，存在確率に関するものを $Loss_{conf}$ ，クラスの事後確率のものを $Loss_p$ としたとき，統合した損失関

<sup>3</sup><https://github.com/Guanghan/darknet/blob/master/scripts/convert.py> (2018/1/27 存在確認)



図 5.2: アノテーションの例

数を次式に示す .

$$Loss = Loss_{box} + Loss_{conf} + Loss_p \quad (5.1)$$

また , それぞれの損失関数は次式で定義される .

$$Loss_{box} = \lambda_{coord} \sum_i^{S^2} \sum_j^B \mathbb{1}_{ij}^{obj} (x_{ij}^{pred} - x_{ij}^{truth})^2 \quad (5.2)$$

$$+ (y_{ij}^{pred} - y_{ij}^{truth})^2 + (w_{ij}^{pred} - w_{ij}^{truth})^2 + (h_{ij}^{pred} - h_{ij}^{truth})^2$$

$$Loss_{conf} = \lambda_{conf}^{obj} \sum_i^{S^2} \sum_j^B \mathbb{1}_{ij}^{obj} \{conf_{ij}^{pred} - iou(box_{ij}^{pred}, box_{ij}^{truth})\}^2 \quad (5.3)$$

$$+ \lambda_{conf}^{no\_obj} \sum_i^{S^2} \sum_j^B \mathbb{1}_{ij}^{no\_obj} (conf_{ij}^{pred})^2$$

$$Loss_p = \lambda_{prob} \sum_i^{S^2} \sum_j^B \mathbb{1}_{ij}^{obj} \{p_{ij}^{pred}(c) - p_{ij}^{truth}(c)\}^2 \quad (5.4)$$

ここで , 添字  $pred$  ,  $truth$  が付与された変数がそれぞれ予測値 , 正解を示す . また ,  $\mathbb{1}_i^{obj}$  は , 何らかの物体がセル  $i$  に存在するかどうかを示し , 存在すれば 1 , 存在しなければ 0 である .  $\mathbb{1}_{ij}^{obj}$  は , セル  $i$  内の  $j$  番目の矩形候補が , その物体の予測の「担当」であることを示す .  $\mathbb{1}_i^{no\_obj}$  は ,

表 5.1: 看板画像の学習と検出に用いたマシンのスペック

要素	スペック
CPU	Intel (R) Core (TM) i7-8700K @ 3.70 GHz
RAM	16 GB
GPU	NVIDIA GeForce GTX 1080 Ti
VRAM	12 GB
OS	Ubuntu 17.10

セル  $i$  に物体が存在すれば 0 , 存在しなければ 1 である . また ,  $IoU(A, B)$  は , 矩形  $A$  と  $B$  の Intersection over Union ( IoU ) を表し ,  $p(c)$  はセル  $i$  内の  $j$  番目の矩形候補内に物体が存在するとき , それがクラス  $c$  である事後確率を表す . また ,  $\lambda_{coord}^{obj} = 1.0$  ,  $\lambda_{conf}^{obj} = 5.0$  ,  $\lambda_{conf}^{no\_obj} = 1.0$  ( ただし ,  $IoU$  が閾値未満のとき ) ,  $\lambda_{prob} = 1.0$  である [32] .

上記の学習データを , 1 クラスの R-CNN として YOLOv2 で学習させる . 学習データは  $416 \times 416$  の正方形にリサイズし , パラメータとしてバッチサイズは 64 , 分割数は 8 , 学習率は  $10^{-3}$  , 慣性は 0.9 , 荷重減衰は 0.0005 を与え反復回数は 1,000 回とする . 少ない学習データからモデルを作成するため , 22 層の畳み込み層と 5 層のプーリング層からなる学習済みモデルである Darknet19 モデル<sup>4</sup>を用いる . Darknet19 モデルの最後の畳み込み層を取り除き , フィルタの数が 1024 である 3 層の  $3 \times 3$  の畳み込み層と , フィルタサイズが 30 である 1 層の  $1 \times 1$  の畳み込み層を加え , 学習を行う . 学習に用いたマシンのスペック表 5.1 に示す .

### 5.3.3 評価

看板領域抽出の精度を評価するために , 領域の一致具合を評価する手法である IoU を用いる . 正しい看板の領域を  $GT$  , 看板と推定された領域を  $PR$  とすると , IoU 値は ,

$$IoU = \frac{GT \cap PR}{GT \cup PR} \quad (5.5)$$

と定義される . 1,000 回の反復学習の時点において , average loss は 0.526 , 8 分割されている IoU 値の平均は 0.850 であった . 損失のグラフを図 5.3 に示す .

## 5.4 抽出した看板の分類

YOLOv2 で抽出した看板領域を分類するために , VGG16 を用いる . VGG16 とは , 画像データベースの ImageNet<sup>5</sup>に登録されている 120 万枚の画像を 1,000 カテゴリに分類した CNN のモデルである . VGG16 は 13 層の畳み込み層と 3 層の全結合層の計 16 層からなるモデルであり , この全結合層を取り除き , 新たに全結合層を追加する転移学習を用いることによって , 少ないデータから短時間で高精度のモデルを構築できる . 本稿では , ニューラルネットワークのライブラリである Keras[4] で実装された VGG16 を用いて学習を行う .

<sup>4</sup><https://pjreddie.com/darknet> ( 2019/1/27 存在確認 )

<sup>5</sup><http://www.image-net.org> ( 2019/1/29 存在確認 )

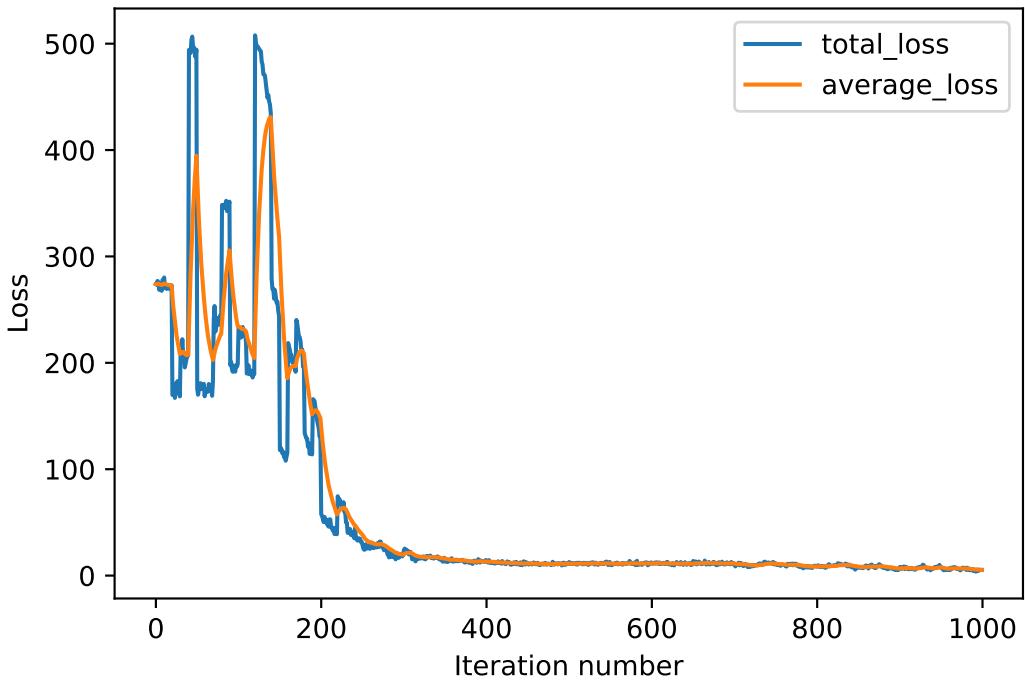


図 5.3: YOLOv2 による学習誤差曲線

#### 5.4.1 学習データ

学習に用いるデータは、5.2 節で述べた 15 店舗の看板を、YOLOv2 に用いたものとは別に各店舗 100 枚ずつ様々な角度から撮影した写真を計 1,500 枚容易し、人手で看板部分のみを切り出した。切り出した写真を、1 店舗につきトレーニングセットとして 50 枚、バリデーションセットとして 25 枚、テストセットとして 25 枚に無作為に分類した。

#### 5.4.2 学習

上記の学習データを、サイズは  $224 \times 224$ 、バッチサイズは 16、学習率は  $10^{-4}$ 、慣性は 0.9、エポック数は 50 として、表 5.1 のマシンを用いて学習させる。1 エポックにつき 24 回反復して学習を行う。学習データが少ないため、トレーニングセットを Keras の ImageDataGenerator<sup>6</sup> クラスを用いてせん断変換したり、拡大縮小したり、水平方向に反転したりすることによって十分な量となるまで拡張しながら学習を行う。損失関数には、次式で示す 15 クラスの交差エントロピーを用いる。

$$E = -\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^{15} t_{nk} \log y_{nk} \quad (5.6)$$

ここで、 $N$  はデータ数、 $y_k$  はニューラルネットワークの出力、 $t_k$  は正解ラベルを表す。

<sup>6</sup><https://keras.io/preprocessing/image> ( 2019/1/29 存在確認 )

### 5.4.3 評価

学習過程におけるトレーニングデータとバリデーションデータの精度と損失のグラフを図 5.4 に示す。50 エポック時点でのトレーニングデータの精度は 0.987，損失は 0.097，バリデーションデータの精度は 0.963，損失は 0.156 であった。評価指標に関しては，真に正であり，かつ予測結果も正であるデータ数を True Positive (TP)，真に正であるが，予測結果が負であるデータ数を False Negative (FN)，真に負であるが，予測結果が正であるデータ数を False Positive (FP)，真に負であり，かつ予測結果も負であるデータ数を True Negative (TN) とすると，正と予測した場合の中で実際にそれが正しい割合である適合率 (Precision)，再現率実際に正である場合の中で正であると予測されたものの割合である再現率 (Recall)，適合率と再現率の調和平均である F 値は次式で定義される [39]。

$$Precision = \frac{TP}{TP + FP} \quad (5.7)$$

$$Recall = \frac{TP}{TP + FN} \quad (5.8)$$

$$F = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision} \quad (5.9)$$

テストデータを用いて各クラス毎に上記の評価指標を計算した結果を表 5.2 に示す。各評価指標について，全クラスの結果の平均であるマクロ平均を算出した。その結果，適合率は 0.960，再現率は 0.955，F 値は 0.956 であり，このモデルは看板画像のクラスを分類するために十分な精度があるといえる。

## 5.5 API の実装

モバイルデバイスでリアルタイムに看板を認識するために，画像を送信すると，画像内の看板領域と，そのクラスを返却する API を実装する。API は，Python 3.6.5 と Web API フレームワークである Falcon<sup>7</sup>を用いて構築した。サーバには，表 5.1 に示す，学習に用いたものと同一のマシンを用いる。まず，クライアントが API に画像を POST リクエストとして送信すると，サーバは Tensorflow 1.0[1] で実装された YOLOv2 により画像内の看板領域が検出され，切り出される。次に，切り出された看板画像を VGG16 によってクラス分けを行い，どの店舗の看板なのかを識別する。最後に，画像内に存在する全ての看板領域の左上と右下の座標，店舗のクラスを格納した JSON データを生成し，クライアントに返却する。この API を 500 ミリ秒毎に呼び出すことによって，モバイルデバイスでのリアルタイムな看板認識が可能になる。

## 5.6 フィールド実験

構築したモデルを用いて，実環境でも店舗を識別できるか確認するために，スマートフォンのカメラで撮影した画像を API に送信し，結果を表示する Android アプリケーションを Kotlin<sup>8</sup> を用いて実装した。実装したアプリケーションを用いて，各店舗の前で写真を撮影し，10 回ず

<sup>7</sup><https://falconframework.org> (2019/2/4 存在確認)

<sup>8</sup><https://kotlinlang.org> (2019/1/29 存在確認)

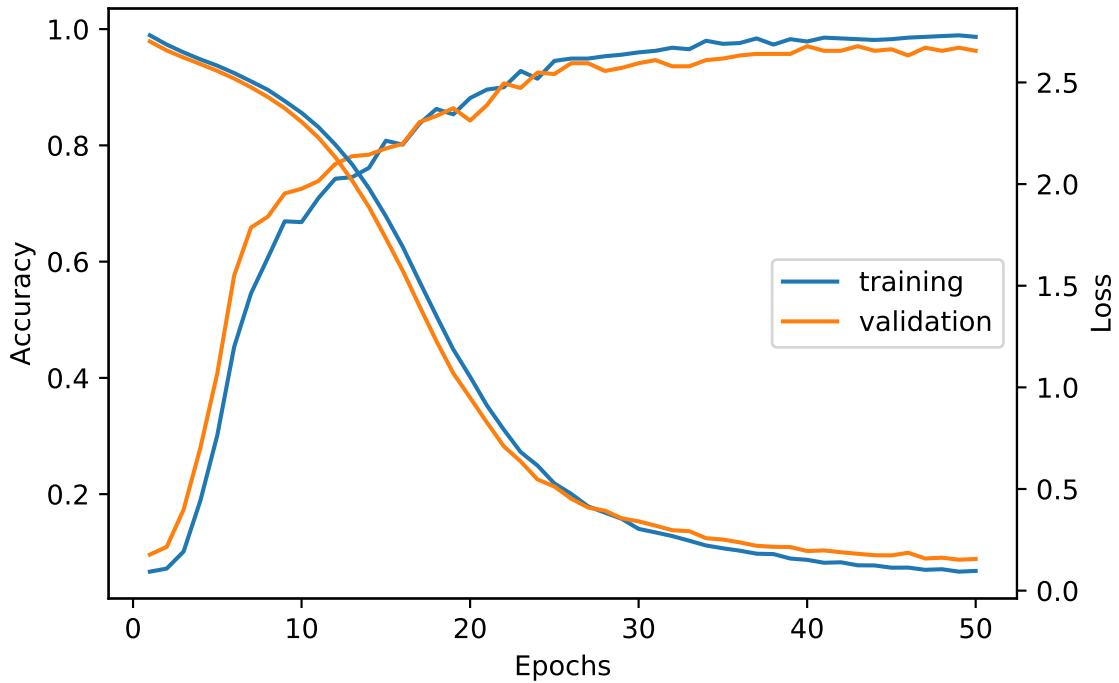


図 5.4: VGG16 による学習誤差曲線

つ API に画像を送信するフィールド実験を行った。実験では、(a) 看板領域が正しく検出された回数、(b) 看板領域が正しく検出された場合の内、看板画像が正しくクラス分けされた回数、をそれぞれ測定した。実験結果を表 5.2 に示す。看板領域検出の再現率は 0.947、看板領域が正しく検出された場合の内、看板画像のクラス分けの再現率は 0.944 であり、全体の再現率は 0.893 であった。500ms 毎に API を呼び出すことによって、リアルタイムでの認識も可能であった。考察として、図 5.5 に示す「おだいどこはなれ」のように、同一の店舗が類似していない複数の看板を掲示している場合、本稿では 1 つの看板のみを学習させていたため、テストデータの精度は高かったものの、実世界での認識の精度は低くなる傾向が見られた。

表 5.2: モデルの評価及びフィールド実験の結果

店舗	看板画像	適合率	再現率	F 値	(a)	(b)
(1)		1.000	1.000	1.000	9	9
(2)		0.962	1.000	0.980	10	10
(3)		1.000	1.000	1.000	9	9
(4)		0.767	0.920	0.836	10	10
(5)		0.962	1.000	0.980	10	10
(6)		0.800	0.960	0.873	9	9
(7)		1.000	0.960	0.980	10	10
(8)		1.000	1.000	1.000	8	3
(9)		0.955	0.840	0.894	10	7
(10)		0.962	1.000	0.980	10	10
(11)		1.000	0.840	0.913	10	10
(12)		1.000	0.880	0.936	10	10
(13)		1.000	1.000	1.000	8	8
(14)		1.000	0.960	0.980	9	9
(15)		1.000	0.960	0.980	10	10



図 5.5: 認識精度が低かった看板の例

## 6 Search by Snap のプロトタイプ

本章では、Search by Snap を用いて実装したプロトタイプについて述べる。

### 6.1 システムの概要

3.4 節で述べたデザイン指針に基づき、5 章で用いた店舗を対象としたプロトタイプを実装する。OSM 上では、飲食店などの店舗のオブジェクトは位置を定義された単一の点であるノードとして扱われ、各オブジェクトには“OSM ID”という一意の番号が付与されている。各店舗の看板画像と OSM ID を紐付け、Overpass API<sup>1</sup>を用いて店舗情報を取得する。これにより、5 章で実装した API を用いてリアルタイムで看板を認識し、店舗の看板から店舗情報へのアクセスができる。各店舗の店舗名とノードは表 6.1 に示す通り対応しており、表 6.2 に示すそれぞれのノードには複数のタグが設定されている。“name”は店舗の名称が代入されている。“shop”タグは、店舗が販売している商品を記述するために使用される。飲食店の場合、施設の種類を表す“amenity”タグにバーやレストランのような店舗の種類が代入されている。“opening\_hours”タグには、店舗の営業時間が代入されている。年中無休で 24 時間営業の場合は“24/7”が、曜日によって営業時間が異なる場合は、セミコロン区切りで複数の値が代入されている。例えば、小だるま JR 高槻駅前店は、月曜日から木曜日までは 11 時 30 分から 14 時 00 分と 17 時から 23 時 30 分まで、金曜日、土曜日は 11 時 30 分から 14 時 30 分と 17 時 00 分から 24 時 30 分、日曜日は 11 時 30 分から 23 時 30 分が営業時間である。この場合、“opening\_hours”タグの値は“Mo-Th 11:30-14:00, 17:00-23:30; Fr-Su 11:30-14:30, 17:00-24:30; Su 11:30-23:30”となる。支払いにクレジットカードが利用可能かどうかは、“american\_express”, “payment:diners\_club”, “payment:jcb”, “payment:master”, “payment:visa”, のタグに“yes”が代入されていれば利用可能である。表 6.2 には “payment:visa” タグのみを掲載している。

### 6.2 システムのインターフェース

クライアントサイドは、ユーザが OS を問わずに携帯端末で実行できるようにするために、HTML5 と Javascript を用いて Web アプリケーションとして実装した。インターフェースはデフォルトで端末の背面カメラ画面となっており、ユーザが前面カメラと切り替えられるようになっている。ユーザがカメラを通して店舗の看板を見ると、その店舗の店舗名、営業時間、使えるクレジットカードの種類などの情報が表示される。表示される店舗の種類や表示する情報の種類はユーザが選択できる。視覚情報の視認性を上げるために、減算型表示を用いてユーザにとって不要な情報は目立たなくする。Javascript ではカメラ映像をリアルタイムでの表示に十分な速度でグレースケール化できないため、不要部分に半透明の黒をマスクし、明度を下げるこことによって情報を減算している。実装したユーザインターフェースを図 6.1 に示す。

---

<sup>1</sup><http://overpass-api.de> ( 2019/1/22 確認 )



図 6.1: システムのインターフェース

### 6.3 システムの動作

提案システムにおいて、ユーザが取る行動とシステムが行う処理を以下に示し、図で表したもののが図 6.2 に示す。

- (1) システムは携帯端末の GPS 機能により、ユーザの位置情報を取得する
- (2) ユーザは店舗の種類や価格帯などのクエリを入力する
- (3) システムは Overpass API に位置情報とクエリを送信し、クエリに一致する周辺の店舗の OSM ID を取得する
- (4) ユーザは携帯端末のカメラを通して街から画像情報を取得する（図 6.2 中①）
- (5) システムは画像をサーバに送信する（図 6.2 中②）
- (6) サーバは YOLO を用いて画像の中から看板領域を検出する（図 6.2 中③）

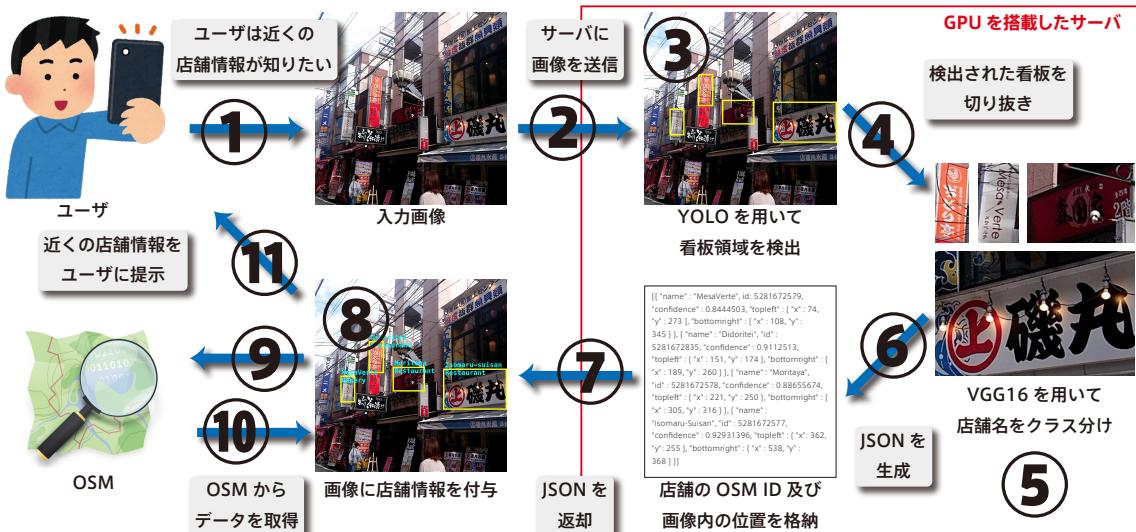


図 6.2: システムの動作

- ( 7 ) サーバは検出された看板領域を切り抜く ( 図 6.2 中④ )
- ( 8 ) サーバはそれぞれの看板画像を VGG16 を用いて店舗名でクラス分けする ( 図 6.2 中⑤ )
- ( 9 ) サーバは画像内の看板の左上と右下の座標と検出した看板の店舗と関連付けられた OSM のノードの id をそれぞれ格納した JSON データを生成する ( 図 6.2 中⑥ )
- ( 10 ) サーバは生成された JSON をシステムに返り値として返却する ( 図 6.2 中⑦ )
- ( 11 ) システムはユーザが求めていない視覚情報の色調を低減させ、ユーザが求めている店舗の情報を看板の横に吹き出しとして重畳表示する ( 図 6.2 中⑧ )
- ( 12 ) システムは認識した看板と関連付けられている OSM ID を Overpass API に送信する ( 図 6.2 中⑨ )
- ( 13 ) Overpass API は OSM ID と一致する店舗のノードをシステムに返却する ( 図 6.2 中⑩ )
- ( 14 ) システムは出力結果をユーザに提示する ( 図 6.2 中⑪ )

デバイスが画像を API に送信してからサーバが JSON データを返却するまでに要した時間は、上り 1Mbps、下り 1.25Mbps の通信速度、980 × 1307 の解像度で平均 359ms であった。Thrope らによると、人間が画面中央のオブジェクトを認識するまでのリアクションタイムの中央値は 445ms であるため [42]、十分にユーザが満足できる速度であるといえる。

表 6.1: 店舗名と OSM ID の対応

店舗名	OSM ID
セブン-イレブン 高槻高槻町店	2391866925
全席個室居酒屋 北海の恩返し 高槻店	5279265335
小だるま JR 高槻駅前店	5281672553
肉丼専門店 高槻肉劇場	5281672555
おだいどこはなれ 高槻店	5281672556
ビリヤード・ダーツ&Food Bar Ozbuddy	5281672557
磯丸水産 高槻店	5281672577
炭焼酒場 森田家	5281672578
メサベルテ 高槻店	5281672579
駿河屋	5281672580
甲南チケット 高槻本通店	5281672581
焼肉・しゃぶしゃぶ食べ放題 ぱくぱく 高槻店	5281672739
赤から 高槻店	5281672740
ちどり亭 高槻店	5281672835
高槻ちゃんこ	5281672942

表 6.2: 対象とする店舗のノード (2018 年 12 月 1 日時点の OSM のデータベースを基に作成)

id	shop	amenity	opening_hours	payment:visa	...
2391866925	convenience		24/7	yes	
5279265335		bar	17:00-24:00	yes	
5281672553		bar	Mo-Th 11:30-14:00; ...	yes	
5281672555		restaurant	11:00-23:00	no	
5281672556		bar	Mo-Th 11:30-24:00; ...	yes	
5281672557		pub	Mo-Th 14:00-26:00; ...	no	
5281672577		restaurant	24/7	yes	
5281672578		restaurant	17:00-26:00	yes	
5281672579	bakery		Mo-Sa 07:00-20:30; ...	no	
5281672580	hobby		10:00-23:00	no	
5281672581	ticket		Mo-Sa 10:00-19:00; ...	no	
5281672739		restaurant	Mo-Fr 16:30-25:00; ...	yes	
5281672740		bar	Mo-Fr 17:00-24:00; ...	yes	
5281672835		restaurant	17:00-25:00	yes	
5281672942		bar	17:54-06:00	yes	

## 7 減算型表示を用いた評価実験

本章では、減算型表示を用いて実施した評価実験と、その結果について述べる。

### 7.1 実験の目的

看板密集地域において特定の看板を探索する際、加算型の情報提示手法と減算型の情報提示手法を用いたそれぞれの探索時間は、3.1節で述べたように差がないことが示唆されている。また、減算型情報提示手法には、看板や背景の彩度が低い場合において、無加工の状態と比較して情報の視認性が変わらない問題点がある。

そこでこの実験では、看板が密集している地域において、加算型情報提示手法と減算型情報提示手法を組み合わせた本稿の提案手法であるハイブリッド型情報提示手法を用いる。この提示手法による探索時間を、従来の加算型情報提示手法、減算型情報提示手法と比較することにより、探索時間に関して提案手法が優位であるかを検証する。

### 7.2 実験の概要

実験参加者は情報系の学部に通う大学生 12 名である。本実験で比較する提示手法は、4 節で述べた通常型、加算型、減算型、ハイブリッド型の 4 種類である。実験は 4 節で述べたプロトタイプを ASUS 社<sup>1</sup>の Nexus 7 (2013, Android 6.0.1) にインストールして行った。

実験参加者図は図 7.1 に示すように立った状態で端末を持ち、端末を全方向に向けることで全天球画像を見回し、提示された看板を探索する。ユーザが実験前に看板を記憶することを防止するために、必ず看板に背を向けた状態で実験を始めるよう指示を出した。実験の条件は、表 7.1 に示す 8 通りとした。ここで探索対象の看板数に関して、単体である場合と複数である場合の 2 通りに区別した。これは探索対象の看板が複数である場合、不要な情報を減算する方が情報を加算することに比べてより容易に情報を探索でき、探索時間が短くなるという仮説に基づいている。

### 7.3 実験の手順

初めに、端末の画面が図 7.2 - (a) の状態で実験参加者に端末を渡す。この画面には探索対象の看板画像が表示されている。ユーザが開始ボタンをタップすると、図 7.2 - (b) に示す探索画面に遷移する。次に、ユーザが振り返ると、図 7.2 - (c) に示すビルの看板があり、ユーザはその中から探索対象の看板を探す。ユーザが探索対象の看板に 1 秒間照準を合わせると、図 7.2 - (d) に示す終了画面に遷移する。探索対象が複数である場合は、全ての看板を見つけた後に遷移する。

実験では、全ての実験参加者に表 7.1 の条件で探索してもらった。探索対象が単体である条件（表 7.1, 1~4）、探索対象が複数である条件（表 7.1, 5~8）の順に、昼間の画像と夜の画像でそれぞれ探索してもらった。

<sup>1</sup><https://www.asus.com/> (2017/4/27 確認)

表 7.1: 実験の条件

番号	提示手法	対象
1	通常型	単体
2	加算型	単体
3	減算型	単体
4	ハイブリッド型	単体
5	通常型	複数
6	加算型	複数
7	減算型	複数
8	ハイブリッド型	複数

## 7.4 実験の結果

実験参加者が開始ボタンをタップしてから、指示された看板を全て見つけるまでの所要時間を計測し、通常型、加算型、減算型各々の探索時間とハイブリッド型の探索時間の平均値を比較した。その結果を以下に述べる。

### 7.4.1 時間帯が昼、探索対象が単体の場合

時間帯が昼、探索対象が単体の場合の探索時間を図 7.3 - (a) に示す。ハイブリッド型を用いた探索時間は、(1) 通常型より有意に短い ( $t(22) = 5.729, p < .05$ ) こと、(2) 加算型より有意に短い ( $t(22) = 2.852, p < .05$ ) ことが確認されたが、(3) 減算型との間に有意差は見られなかった ( $t(22) = 1.478, n.s.$ )。

### 7.4.2 時間帯が昼、探索対象が複数の場合

時間帯が昼、探索対象が複数の場合の探索時間を図 7.3 - (b) に示す。ハイブリッド型を用いた探索時間は、(1) 通常型より有意に短い ( $t(22) = 5.702, p < .05$ ) こと、(2) 加算型より有意に短い ( $t(22) = 6.144, p < .05$ ) こと、(3) 減算型より有意に短い ( $t(22) = 3.318, p < .05$ ) ことがそれぞれ確認された。

### 7.4.3 時間帯が夜、探索対象が単体の場合

時間帯が夜、探索対象が単体の場合の探索時間を図 7.4 - (a) に示す。ハイブリッド型を用いた探索時間は、(1) 通常型より有意に短い ( $t(22) = 4.325, p < .05$ ) こと、(2) 加算型より有意に短い ( $t(22) = 2.103, p < .05$ ) こと、(3) 減算型より有意に短い ( $t(22) = 2.485, p < .05$ ) ことがそれぞれ確認された。

### 7.4.4 時間帯が夜、探索対象が複数の場合

時間帯が夜、探索対象が複数の場合の探索時間を図 7.4 - (b) に示す。ハイブリッド型を用いた探索時間は、(1) 通常型より有意に短い ( $t(22) = 5.512, p < .05$ ) こと、(2) 加算型との間に有意差はない ( $t(22) = 1.857, n.s.$ ) こと、(3) 減算型より有意に短い ( $t(22) = 3.129, p < .05$ ) こと



図 7.1: 実験風景（文献 [21] より図引用）

ことがそれぞれ確認された。

#### 7.4.5 アンケート結果

実験終了後に、最も対象の看板を見つけやすかった提示手法についてアンケートを実施したところ、実験参加者の過半数が提案手法であるハイブリッド型情報提示手法を選択した。



(a) 開始画面



(b) 探索画面 (1)



(c) 探索画面 (2)



(d) 終了画面

図 7.2: 実験の手順 (文献 [21] より図引用)

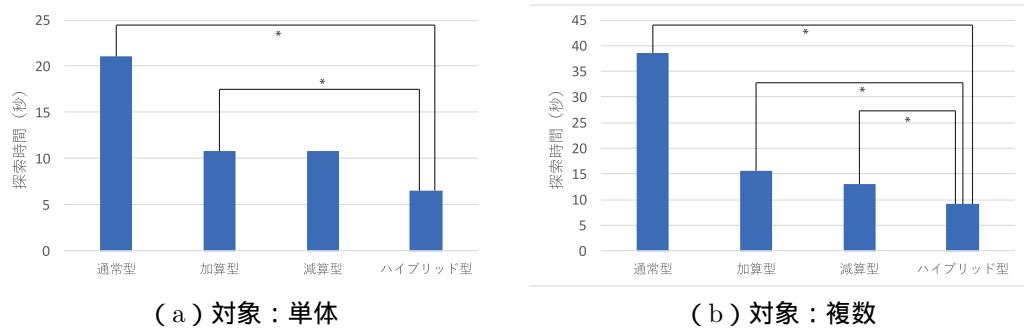


図 7.3: 実験結果 : 昼 (\* :  $p < .05$ ) (文献 [21] より図引用)

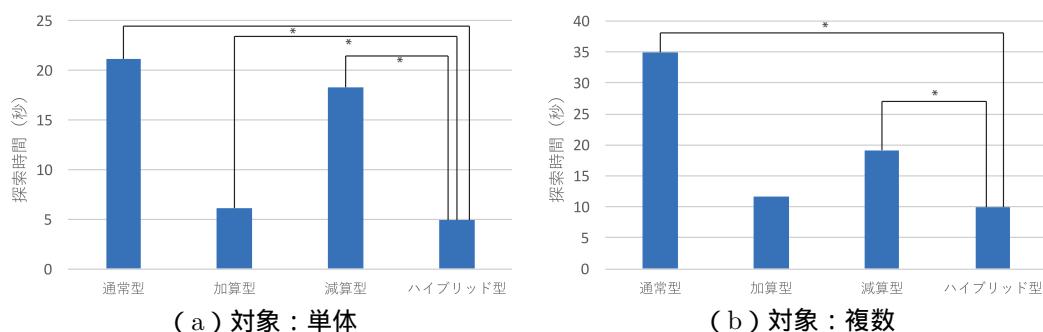


図 7.4: 実験結果 : 夜 (\* :  $p < .05$ ) (文献 [21] より図引用)

## 8 Search by Snap を用いたユーザ実験

本章では、Search by Snap を用いて実施したユーザ実験と、その結果について述べる。

### 8.1 実験の目的

1.2 節で述べたように、慣れていない地域や周囲の文字が読めない状況において、ユーザが目の前にある店舗の情報や条件に合致する店舗の情報を瞬時に取得することは容易ではない。位置情報を手がかりに周囲の検索を行い、条件に合った店舗が見つかったとしても、ユーザ自身が居る環境と検索行為とが分断されているため、ユーザの現在時点から検索対象の店舗を探す手間が残るという問題点がある。そのため、提案システムを用いることによって、位置情報を利用したサービスと比較して店舗を簡単かつ直感的に探索できることを確認する。本稿では、地域に慣れていないユーザを対象に評価を行う前段階として、提案システムのユーザビリティを評価するために、地元の大学生を対象にユーザ実験を実施する。

### 8.2 予備実験

#### 8.2.1 実験の概要

実験は大阪府高槻市にある高槻本通りにおいて、6 章で実装したシステムを用いて店舗を探査する際のユーザの行動を観察するために、ユーザが商店街において条件に合う店舗を探している状況を想定して行った。実験参加者は情報系の学部に通う大学生 4 名である。実験は昼間を行い、天候は曇りであった。

表 6.2 に示す 15 店舗を対象とし、実験参加者にはタスク (1) として、支払いに VISA カードが使える店舗を探すこと、タスク (2) として、月曜日の 24 時以降に営業している店舗を探すこと、の 2 点を課した。それぞれ提案システムを用いた場合と Web 検索エンジンを用いた場合とで探索に要した時間及び正確さを測定した。Web 検索エンジンには、World Wide Web 上で最も多く使われている検索エンジン [2] である Google 検索<sup>1</sup>を用いる。実験に用いる端末は 7 章と同じ Nexus 7 である。

#### 8.2.2 実験の手順

探索に用いる手法の順序効果によるデータの偏りを排除するため、実験は 2 人同時に実行し、タスク (1) は 1 人が提案システムを用いて行い、もう 1 人が Google 検索を用いて行う。2 人がタスク (1) を終えたら、実験参加者はタスク (2) を異なる手法を用いて行う。

実験の手順として、初めに、実験参加者に回答用フォームの URL を伝え、参加者は自身のスマートフォンでフォームを開く。次に、提案システムの操作方法を実験参加者に説明する。その後、2 人の実験参加者が同時に探索を開始し、フォーム上で条件に合致する店舗のチェックボックスを選択する。実験参加者が探索を開始してから送信ボタンをタップするまでの時間を測定する。

実験終了後には、参加者に「求めていた情報の見つけやすさ（簡便性）」、「情報探索の直感性（直感性）」について、提案システムと Google 検索との間で 5 段階のリッカート尺度を用い、「1」

<sup>1</sup><https://www.google.com> (2019/2/1 存在確認)

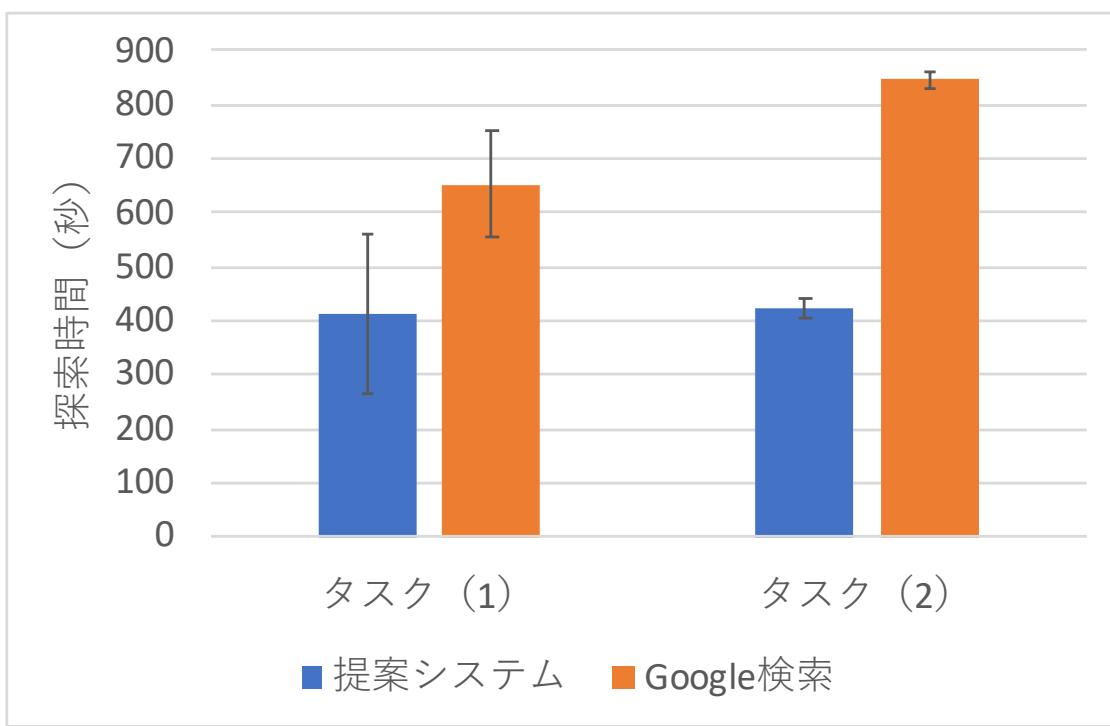


図 8.1: 予備実験の結果（探索時間）

を「Google 検索」、「5」を「提案システム」として 5 段階評価のアンケートへの回答を求めた。

### 8.2.3 実験の結果と考察

実験の結果を図 8.1, 8.2 に示す。探索時間に関して、システムを用いた場合の探索時間は Google 検索を用いた場合と比較して短くなる傾向が見られた。正確さに関して、システムを用いた場合の正確性は Google 検索を用いた場合とで差が見られない傾向があった。アンケートの結果を表 8.1 に示す。アンケートの結果から、提案システムを用いることによって、Google 検索を用いた場合より直感的かつ簡易に探索ができるようになることが示唆された。

表 8.1: 予備実験のアンケート結果

実験参加者	簡便性	直感性	コメント
A	5	5	カメラを向けるだけですぐ情報が見えるのでよかったです
B	5	5	看板を映してパッと求めている情報が出てくるのはとても良いなと思いました。情報が見つけやすい。
C	4	5	同じ地域に何店舗もある店だと検索サイトで検索した際、どの店かわかりづらく時間がかかりました。
D	5	4	

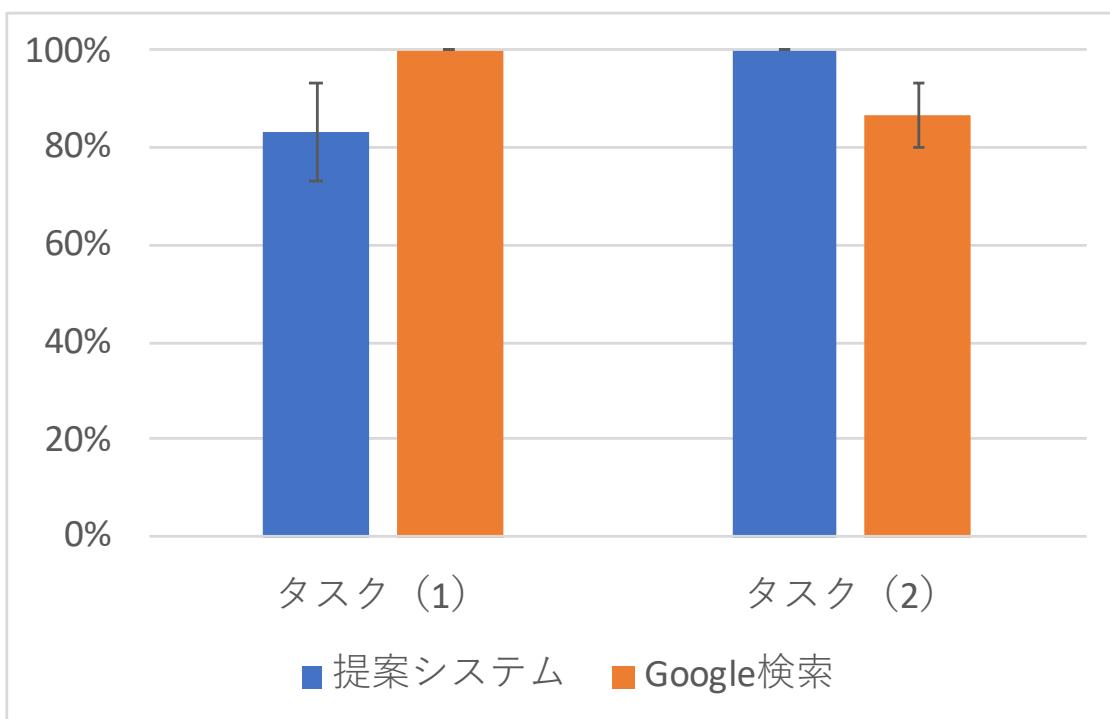


図 8.2: 予備実験の結果（正解率）

#### 8.2.4 ユーザ観察

実験中のユーザを観察した結果、システムを用いて探索する際、至近距離で看板を認識しようとする行動が見られた。端末のカメラが店舗の看板の大部分の領域を捉えられない場合、正常に看板を認識できない場合が多いため、本実験では看板から一定の距離を置いて端末のカメラを向けるよう指示を与える。Google 検索を用いて探索する際、多くのユーザがタスク(1)においては「(店舗名) 高槻」というクエリ、タスク(2)においては「(店舗名) 高槻 営業時間」というクエリで検索を行っていた。対象がチェーン店である場合、位置に関する情報をクエリとして与えなければ目的の情報に素早く辿り着けず、探索に時間を要している傾向が見られた。多くのユーザは、対象が飲食店である場合において、Google 検索結果に表示された食べログのサイトへ移動し、そこから情報を得ていた。そのため、本実験では、探索対象を飲食店に限定し、食べログを用いて提案システムとの比較を行う。

### 8.3 本実験

#### 8.3.1 実験の概要

実験は大阪府高槻市にある高槻本通りにおいて、昼間にユーザが飲食店を探している状況を想定して行った。実験参加者は情報系の学部に通う大学生 10 名である。

実験参加者にはタスク(A)として、「小だるま JR 高槻駅前店」、「肉丼専門店 高槻肉劇場」、「磯丸水産 高槻店」の3店舗について、使用できるクレジットカードの種類を“Visa”, “Mastercard”, “JCB”, “American Express”, “Diners Club” の5つのチェックボックスから全て選択すること



図 8.3: 実験の実施地点

と、タスク(B)として、「おだいどこはなれ 高槻店」、「高槻ちやぶちやぶ」、「赤から 高槻店」の3店舗について、月曜日の営業開始時刻を数値で入力すること、の2点を課した。タスク(A)は図8.3中①で示した位置に立った状態、タスク(B)は図8.3中②で示した位置に立った状態で行った。実験に用いた携帯端末はクアンタ・コンピュータ社<sup>2</sup>のVA-10J( Android 5.0.2 )である。提案システムとの比較対象とする位置情報を用いたサービスは、ランキングやユーザの口コミ・写真をもとにレストランの検索が行えるサービスである食べログ<sup>3</sup>を用いた。

### 8.3.2 実験の手順

初めに、提案システムと食べログの使い方を実験参加者に説明する。次に、実験参加者に回答用フォームのURLを伝え、参加者は自身のスマートフォンでフォームを開く。実験に用いるシステムを起動した状態で参加者に実験用のスマートフォンを渡す。実験参加者はシステムを用いて求められている情報を探索し、フォームに入力して送信する。実験参加者がシステムの操作を始めてから送信ボタンをタップするまでの時間を計測する。実験の風景を図8.4に示す。

実験参加者をグループ(1)とグループ(2)に2分割し、グループ(1)にはタスク(A)を提案システム、タスク(B)を食べログを用いて行うよう指示を出し、グループ(2)にはタス

<sup>2</sup><http://www.quantatw.com> (2019/2/8 存在確認)

<sup>3</sup><https://tabelog.com> (2019/2/8 存在確認)



図 8.4: 実験の風景

ク (A) を食べログ , タスク (B) を提案システムを用いて行うよう指示を出した . 実験終了後には , 参加者に「求めていた情報の見つけやすさ (簡便性)」, 「情報探索の直感性 (直感性)」について , 食べログと提案システムとの間で 5 段階評価のアンケートへの回答を求めた .

### 8.3.3 実験の結果

実験参加者がシステムに操作を初めてから指示された情報を全て収集し , 送信ボタンをタップするまでの時間を計測し , タスク (A) とタスク (B) に関して提案システムを用いた場合と食べログを用いた場合とにおいて , 探索時間の平均値を比較した . その結果を図 8.5 に示す . タスク (A) において , 提案手法を用いた場合の探索時間は食べログを用いた場合よりも有意に短い ( $t(8) = 2.343, p < .05$ ) ことが確認された . タスク (B) においても , 提案手法を用いた場合の探索時間は食べログを用いた場合よりも有意に短い ( $t(8) = 4.370, p < .05$ ) ことが確認された .

各タスクにおいて , ユーザが正確に情報を収集できたかを測定するために , 対象とした 3 店舗のうち , 正確に情報を取得できた店舗数の割合を正解率として測定し , 提案システムを用い

た場合と食べログを用いた場合において、正解率の平均値を比較した。その結果を図 8.6 に示す。タスク (A)において、提案手法を用いた場合と食べログを用いた場合とで、正解率に有意差は見られなかった ( $t(8) = 1.000, n.s.$ )。タスク (B)においても、提案手法を用いた場合と食べログを用いた場合とで、正解率に有意差は見られなかった ( $t(8) = 1.633, n.s.$ )。

予備実験と同様に、実験終了後に情報探索の簡便性と直感性について 5 段階のリッカート尺度を用い、「1」を「食べログ」、「5」を「提案システム」として 5 段階で回答してもらった。簡便性に関する平均値は 4.4、直感性に関する平均値は 4.5 であった。そのアンケート結果を表 8.2 に、アンケート結果の分布を図 8.7 に示す。

表 8.2: 本実験のアンケート結果

実験参加者	群	簡便性	直感性	コメント
1	A	5	5	看板が多いところだと情報の表示が重なり見にくいところがあった。
2	A	4	4	どちらの検索機能も分かりやすいと思います。
3	A	5	5	リアルタイムな情報が検索できたらもっと便利かと思いました。
4	A	2	4	システムを使った時に店の情報が違う店の情報に被ってしまって見にくいときがあったこと、情報が一定の場所に出続けないでなくて毎回情報と、店名とを見比べる必要がありました
5	A	4	4	食べログはUI次第でどうにかなると思いました。システムの方は、文字を色分けなどしていただいた方が目につきやすいのかとおもいました。
6	B	5	5	タップ数やかかる時間でもシステムの方が有利に感じた。
7	B	5	4	システムでは、欲しい店舗の情報を常に表示し続けるためには かざし続けなければいけなく、ズレてしまった場合 違う店舗情報が出てしまったため 慣れが必要なのかなと思いました。一方、食べログ検索では店舗によって載ってる情報がバラバラなため、必要な情報がなかった際に 本当に情報が無いのかのダブルチェックを必要とすることが億劫であった。
8	B	5	5	
9	B	5	4	
10	B	4	5	すぐに情報を知れて便利だと思いました。

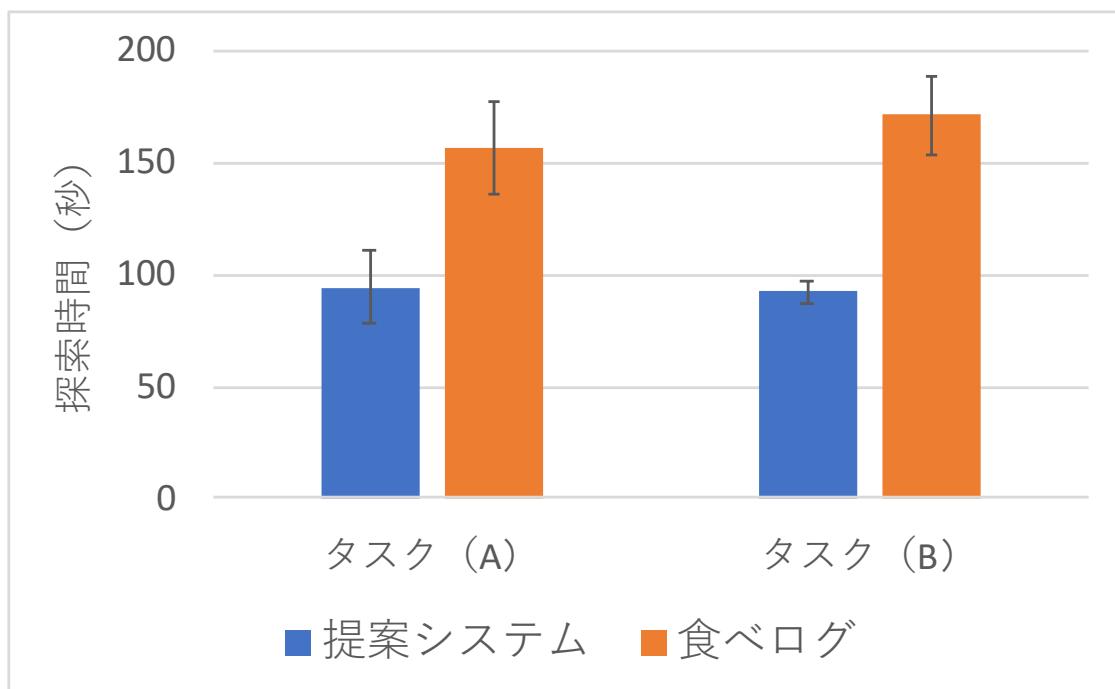


図 8.5: 本実験の結果（探索時間）

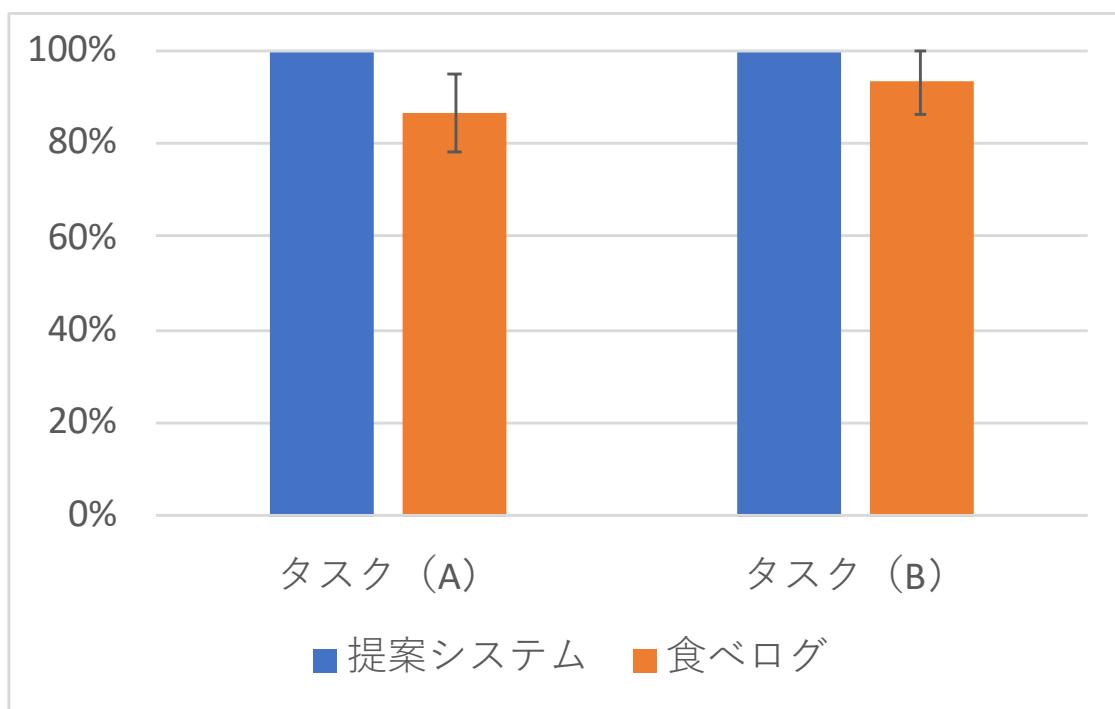


図 8.6: 本実験の結果（正解率）

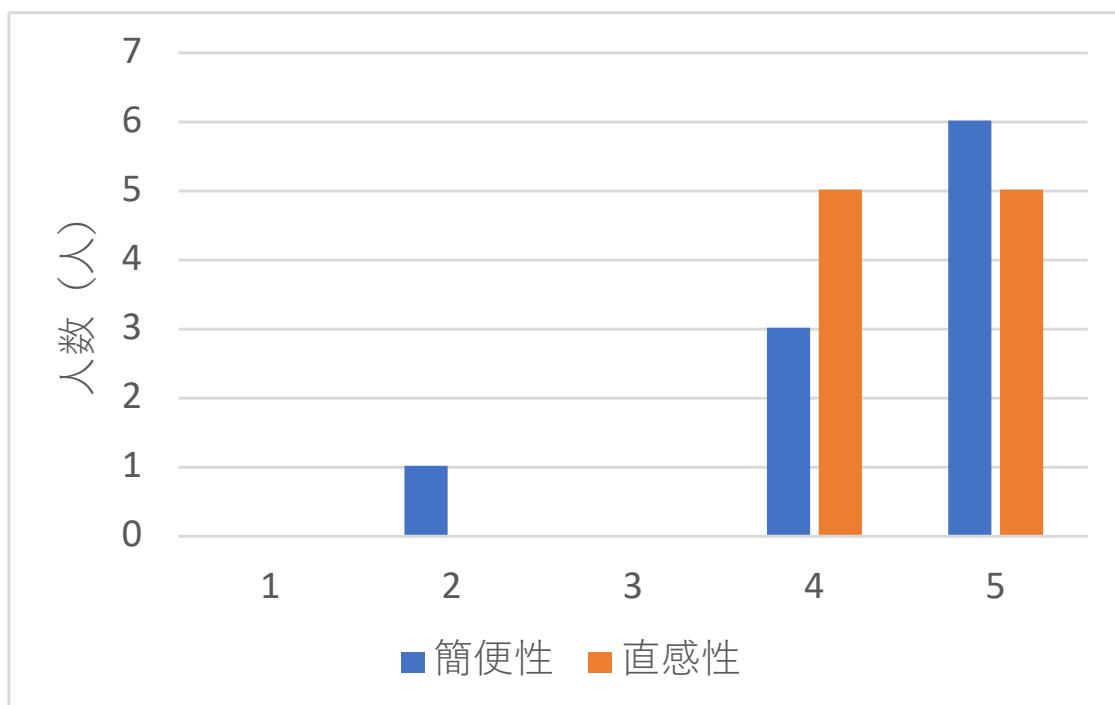


図 8.7: 本実験のアンケート結果の分布 (1: 食べログ ~ 5: 提案システム)

## 9 議論

本章では、7章と8章で述べた実験の結果に基づき、本研究の到達点と改善点、今後の展望について述べる。

### 9.1 得られた知見

7章で述べた実験では、探索対象が単体であり、時間帯が昼の場合、提案手法であるハイブリッド型情報提示手法は文献[7]で提案された減算型情報提示手法と探索時間に関して有意差は見られなかった。しかし、探索対象が複数の場合及び時間帯が夜で探索対象が単体の場合において、提案手法は減算型情報提示手法よりも探索時間が有意に短いことが確認された。このことから、背景の彩度が低い場合や複数の看板を探索する場合において、提案手法は有効であると考えられる。また、時間帯が夜で探索対象が複数である場合、加算型情報提示手法と比較して減算型情報提示手法の方が探索時間が長くなる傾向が見られた。これは実験に用いた写真的背景の彩度が低かったこと、実験に用いたビルの看板に白と黒から構成されているものが多く含まれていたことにより、減算の効果が減少したことが原因と考えられる。

8章で行なった実験の結果から、情報探索時間に関しては、本稿で扱ったいずれの場合においても、提案システムを用いた場合は食べログを用いた場合よりも探索時間が有意に短くなることが明らかとなった。このことから、提案システムを用いることによって、位置情報のみを用いた場合と比較してユーザはより素早く求めている情報を取得することが可能になるといえる。情報探索の正確さに関しては、本稿で扱ったいずれの場合においても、提案システムを用いた場合と食べログを用いた場合とでは正解率に有意差は見られなかった。このことから、言語障壁がなければ位置情報のみを用いても正確に情報を探索することが可能であり、提案システムを用いた場合においても正確な情報探索が可能であるといえる。さらに、アンケート結果から、提案システムを用いることによって、位置情報のみを用いた場合と比較して、より簡単かつ直感的に情報が探索できるようになることが示唆された。

### 9.2 本研究の到達点

7章で実施した実験では、文献[7]で提案された減算型情報提示手法に文字情報を追加した加算型と減算型のハイブリッド型情報提示手法を提案した。提案手法を用いたシステムのプロトタイプを実装することにより、減算型情報提示手法の問題点であった、不要な情報を減算する際に対象となる看板が白黒であり、かつその周辺の景色の彩度が低い場合に減算の効果が減少するという点が解消されたと考えられる。これにより、彩度が低い環境においても分かりやすい情報提示が可能となった。また、実験結果から看板を探索する際、提案手法を用いることによって探索時間が短くなることが確認された。以上により、1.3節で述べた看板密集地域における視覚情報の識別性の向上、及び探索時間の短縮が可能になった。

提案システムを用いることによって、本研究の目的である、ユーザの目の前にある店舗の情報を直感的かつ簡単に取得できることが達成されたと考えられる。これにより、1.2節で述べた、慣れていない地域においても、ユーザが求める条件に合致する店舗を探索できるようになることが示唆された。

## 9.3 改善点

### 9.3.1 減算型表示の実環境における実験

4章で実装したシステムは特定の位置で撮影した全天球画像内でのみ店舗の探索が行えるため，7章での実験から得られた知見は，人工環境内における結果にとどまる．そのため，実環境においても本稿で提案した情報提示手法が有用であるかを検証する必要がある．5章で実装したリアルタイム看板認識 API を用いることにより，実環境での実験が可能になるため，看板が密集している地域において OSM のデータベースと看板データベースを構築し，実環境において提案手法の優位性を検討する．

### 9.3.2 看板認識手法

5章で実装したシステムの改善点として，(1) OSM のノードと看板画像を手作業で関連付けなければならない点，(2) インターネット上の情報から多種多様な店舗の看板画像を大量に集めることは困難であるため，手作業で看板画像を 1 店舗につき 100 枚程度集めなければならない点，が挙げられる．(1) に対しては，看板画像を提示してユーザに店舗名を回答するシステムを実装することで解決でき，(2) に対してはユーザに店舗の看板画像を提示し，それと同じ写真を撮影して投稿するシステムを実装することで解決できると考えられる．これらのシステムにゲーミフィケーションを利用し，ユーザの行動に対して報酬を与えることによって，多数のデータを効率よく収集できると考えられる．

今後の展望として，看板認識をサーバ上で行うのではなく，Tiny-YOLO 等を用いて携帯端末上で行うこと検討する．これにより，サーバへ画像を送信する必要がなくなるため，通信量の大幅な軽減が期待される．

### 9.3.3 対象地域の拡張

6章で用いた，OSM のデータは誰もが編集可能であるため，その地域に慣れている地元のユーザが自身でデータを収集し，OSM を通して活用できるようになる枠組みの構築を目指す．OSM のノードには“cuisine” タグが存在し，“burger”，“noodle”，“japanese”，“chinese” など，飲食店で提供される食品の種類を表す値を追加できる．他にもベジタリアン向けのメニューが提供されていることを表す“diet:vegetarian” や，イスラム教の戒律で許されている食品のみを使用したメニューが提供されていることを表す“diet:halal” などのタグも存在する．さらに，店舗名の英語またはローマ字表記を表す“name:en” タグや中国語表記の“name:zh” タグ，韓国語表記の“name:ko” タグなどを充実させることによって，ユーザインタフェースを多言語に対応させることが可能となる．これらのデータを活用することにより，その地域に慣れていない人や，地元の文字が読めない外国人観光客に対して，求めている情報を容易に取得でき，アレルギーや宗教的制約などの理由による食事制限にも対応可能なナビゲーションを実現できる．

本稿における実験参加者は地元の大学生であるため，地域に慣れていないユーザや非漢字圏など地元の文字が読めないユーザを対象としたユーザ実験を実施する．

## 10 結論

本研究の目的は、ユーザが慣れていない地域や周囲の文字が読めない状況であっても、目の前にある店舗の情報を直感的かつ簡便に取得できるシステムの実現である。本稿では「周囲に看板が多数存在する状況」と「周囲の文字が読めない状況」を対象としたプロトタイプを実装し、ユーザ実験を行うことで、既存のシステムとの差異を明らかにした。以下の本稿の内容を纏める。

1章では、本研究に至った背景と問題点を述べ、解決すべき課題を明らかにした。街中における店舗の看板は、人々が求める条件に合致した店舗を探す際に重要な役割を果たしている。しかし、街中における情報探索の問題点として(1)繁華街など看板が密集している地域においては、大量に存在する他の視覚情報に紛れて目的の看板を見つけることが出来ない可能性があること、(2)外国人観光客の場合、言語障壁により目の前にある店舗が自身の求める条件に合致するか検索することが困難である点、の2点を挙げた。(1)に対しては、ユーザにとって不要な情報を目立たなくさせ、必要な情報には追加情報を重畠表示すること、(2)に対しては、スマートフォンのカメラを通して店舗の看板を見るとカメラ映像上に店舗の詳細情報が得られるシステムを実装することにより、これらの問題を解決すると述べた。これらの提案手法の優位性を検証するために、従来手法と比較したユーザ実験を行うことを述べた。

2章では、本研究で提案するシステムの実現に関連する研究を述べ、本研究の位置付けを明らかにした。まず、ARを用いたナビゲーションシステムについて説明し、ARを用いることで紙の地図を用いる場合と比較してより少ない時間で目的地まで辿り着けることについて述べた。次に、情報の視認性に関する研究として、解像度制御による視線誘導と、隠消現実感について述べた。続いて、看板認識に関する研究として、OCRを用いて看板に描かれている文字を認識する研究や、物体検出に関する研究として、CNNを用いて画像の中から物体領域を検出する研究について述べた。最後に、本研究の位置付けとして、Vision-based ARを用いることと、CNNを用いて看板を認識することを述べた。

3章では、これまでの取り組みとして、減算型表示について説明し、本研究で対象とする状況及び提案システムの要件について述べた。加えて、提案システムの要件を満たすための前提条件と、本稿の提案手法であるSearch by Snapについて説明し、そのインターフェースのデザインについて述べた。

4章では、減算型表示を用いたプロトタイプの実装について述べた。プロトタイプでは、ユーザはモバイルデバイスを全天球画像内で全方向に向けることができ、選択されていない種類の店舗の視覚情報を減算すると述べた。

5章では、リアルタイムで看板認識を行うために実装したAPIについて述べた。(1)YOLOv2を用いて看板領域を検出し、(2)VGG16を用いて看板を分類することにより、画像内の看板を認識することを述べた。リアルタイムで看板認識を行うために、GPUを搭載したマシン上にWeb APIサーバを構築し、モバイルデバイスからAPIを呼び出すことでリアルタイム認識を実現した。

6章では、3章で述べたデザイン指針に基づき、Search by Snapを用いて実装したプロトタイプについて述べた。5章で実装したAPIと、OpenStreetMapのデータベースに格納されている店舗情報を用いることによって、ユーザがモバイルデバイスのカメラを通して店舗の看板

を見ると、その店舗の情報が得られるシステムを実装した。

7章では、4章で実装したプロトタイプを用いて実施した評価実験について述べた。実験参加者には、全天球画像内で指示された看板を探索するタスクを課した。実験の結果、ユーザにとって不要な視覚情報を減算し、必要な情報には付加情報を重畳表示することによって、探索時間が有意に短くなることを示した。

8章では、6章で実装したプロトタイプを用いて実施したユーザ実験について述べた。実験参加者には、実際の商店街において、ユーザが条件に合う店舗を探している状況を想定し、使えるクレジットカードの種類と月曜日の営業開始時刻を調べるタスクを課した。提案システムと位置情報を用いた検索サービスとを比較した結果、探索時間に関しては提案システムを用いることで有意に短くなることを示し、正解率に関しては有意差がなく、どちらも正確に探索ができる事を示した。

9章では、7章及び8章で実施した実験から得られた知見と本研究の到達点、提案システムの改善点を述べた。本稿で提案した2つのシステムを用いることによって、ユーザが慣れていない地域においても、ユーザが求める条件に合致する店舗を探索できるようになることが示唆された。改善点としては、(1) 4で実装したプロトタイプを実環境でも利用できるよう新たに実装を行う点、(2) 看板認識に必要なデータセットを効率よく収集する手法を確立する点、(3) OSMのデータベースの充実やインターフェースを多言語化する点、の3点が挙げられる。

10章では、本稿の要点をまとめ結論づけた。

## 謝辞

本研究を纏めるにあたり，関西大学総合情報学部総合情報学科の松下光範教授には多大なご指導，ご鞭撻を賜りました．教授には国内や海外での学会発表の機会を頂きました．研究以外においても，研究室に配属されてから約4年の間，ハッカソンやインターンシップへの参加の機会，IUI や SIGGRAPH などの国際学会での Student Volunteer の機会を頂きました．これらの経験は今後の活動においても大きな自身になると思います．深く感謝いたします．

本研究で用いたデータを作成するにあたり，喜多皓太氏には，炎天下の中多数の看板写真を収集して頂きました．心より感謝いたします．

本研究の実験にご参加頂いた皆様には，寒い中高槻本通りまでご足労いただき，ご協力頂きました．心より感謝いたします．

研究生活を送るにあたり，研究に関する議論や他愛もない会話に付き合って下さった，松下研究室の安尾萌氏，赤星俊平氏，大杉隆文氏，朴炳宣氏，岩崎有基氏，鈴木雄登氏，中西聖氏，濱田もえ氏，後輩たち，卒業された先輩方に心より感謝いたします．岩崎有基氏にはさらに日々の雑談や食事にお付き合いいただきました．また，住友梨花氏，返町周氏，OB の盛山将広氏には，ヤフー株式会社主催のハッカソン “Hack Day 10th Aniv.” に共に参加して頂きました．心より感謝いたします．

小林研究室，インフラ構築

明治大学中村研究室の

株式会社 Stroly の上田洋氏には，

大阪市立西高等学校の

最後に，大学入学から大学院までの6年間，生活を支えてくれた家族，友人に心より感謝いたします．

## 参考文献

- [1] Abadi, M., Barham, P., Chen, J., Davis, Z. C. A., Dean, J., Matthieu Devin, S. G., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y. and Zheng, X.: TensorFlow: A System for Large-Scale Machine Learning, *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation*, pp. 265–283 (2016).
- [2] Alexa Internet, Inc.: The top 500 sites on the web, <https://www.alexa.com/topsites>. (2019/2/1 存在確認).
- [3] Chatzopoulos, D., Bermejo, C., Huang, Z. and Hui, P.: Mobile Augmented Reality Survey: From Where We Are to Where We Go, *IEEE Access*, Vol. 5, pp. 6917–6950 (2017).
- [4] Chollet, F. et al.: Keras, <https://keras.io>. (2019/1/29 存在確認).
- [5] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L.: ImageNet: A large-scale hierarchical image database, *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255 (2009).
- [6] 土井賢治: ディープラーニング活用事例と使いこなしの勘所:[ 画像処理分野 ] 2 . ラーメン画像からの全店舗識別, *情報処理*, Vol. 59, No. 11, pp. 971–973 (2018).
- [7] 藤田一秀, 山本真也, 篠木良, 松下光範: 情報密集地域における情報識別性の向上を目指した提示手法の検討, *情報処理学会研究報告*, Vol. 2013-HCI-154, No. 3, pp. 1–6 (2013).
- [8] Gerstweiler, G., Platzer, K. and Kaufmann, H.: DARGS: Dynamic AR Guiding System for Indoor Environments, *Computers*, Vol. 7(1), No. 5, pp. 1–19 (2018).
- [9] Girshick, R., Donahue, J., Darrell, T. and Malik, J.: Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, *The IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587 (2014).
- [10] Haklay, M. and Weber, P.: OpenStreetMap: User-Generated Street Maps, *IEEE Pervasive Computing*, Vol. 7, No. 4, pp. 12–18 (2008).
- [11] Hata, H., Koike, H. and Sato, Y.: Visual Guidance with Unnoticed Blur Effect, *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 28–35 (2016).
- [12] 林田尚子, 石田亨: 街中における機械翻訳を介したコミュニケーションの支援, *情報科学技術フォーラム一般講演論文集*, Vol. 4, No. 3, pp. 429–430 (2005).

- [13] He, P., Huang, W., Qiao, Y., Loy, C. C. and Tang, X.: Reading Scene Text in Deep Convolutional Sequences, *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pp. 3501–3508 (2016).
- [14] Huang, C., Li, H., Li, W., Wu, Q. and Xu, L.: Store classification using Text-Exemplar-Similarity and Hypotheses-Weighted-CNN, *Journal of Visual Communication and Image Representation*, Vol. 44, pp. 21–28 (2017).
- [15] 岩名地良太, 新田知之, 石川和明, 柳澤政生, 戸川望: 歩行者の視点情報に基づく屋内経路案内, マルチメディア, 分散協調とモバイルシンポジウム 2016 論文集, Vol. 2016, pp. 1748–1756 (2016).
- [16] 神原誠之: 基礎 1: 拡張現実感 (Augmented Reality: AR) 概論, 情報処理, Vol. 51, No. 4, pp. 367–372 (2010).
- [17] 神鳴敏弘: 転移学習, 人工知能学会誌, Vol. 25, No. 4 (2010).
- [18] 加藤未佳, 関口克明: 多数の視覚表示が混在した状況下における効率的な情報取得の為の提示手法に関する研究, 日本建築学会環境系論文集, Vol. 73, No. 631, pp. 1085–1091 (2008).
- [19] Kavati, I., Kumar, G. K., Kesagani, S. and Rao, K. S.: Signboard Text Translator: A Guide to Tourist, *International Journal of Electrical and Computer Engineering*, Vol. 7, No. 5, pp. 2496–2501 (2017).
- [20] Kawai, N., Sato, T. and Yokoya, N.: Diminished Reality Based on Image Inpainting Considering Background Geometry, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 22, No. 3 (2016).
- [21] 北村茂生, 松下光範: 看板密集地域における情報識別支援手法の改良, 情報処理学会研究報告, Vol. 2017-HCI-173, No. 12, pp. 1–7 (2017).
- [22] Kitamura, S., Kita, K. and Matsushita, M.: Real-Time Recognition of Signboards with Mobile Device using Deep Learning for Information Identification Support System, *Proceedings of the Symposium on Spatial User Interaction*, p. 178 (2018).
- [23] Kitamura, S. and Matsushita, M.: Information Identification Support Method for Areas with Densely Located Signboards, *Adjunct Publication of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 159–160 (2017).
- [24] Koch, C., Neges, M., König, M. and Abramovici, M.: Natural markers for augmented reality-based indoor navigation and facility maintenance, *Automation in Construction*, Vol. 48, pp. 18–30 (2014).
- [25] 小山雅明, 高橋由樹, 植塙久雄: “そぞる看板” デザインの基礎的考察, 日本感性工学会論文誌, Vol. 15, No. 1, pp. 65–73 (2016).

- [26] Krishna, O., Aizawa, K. and Reimerth, S.: Signboard Saliency Detection in Street Videos, *2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1917–1921 (2018).
- [27] Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P.: Gradient-Based Learning Applied to Document Recognition, *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324 (1998).
- [28] Lee, C. and Shen, S.: Feature Point Based Text Detection in Signboard Images, *2016 International Conference on Applied System Innovation*, pp. 1–4 (2016).
- [29] Mori, S., Ikeda, S. and Saito, H.: A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects, *IPSJ Transactions on Computer Vision and Applications*, Vol. 9, No. 17, pp. 1–14 (2017).
- [30] 中山英樹: 深層畳み込みニューラルネットワークによる画像特徴抽出と転移学習, 電子情報通信学会技術研究報告, Vol. 115, No. 146, pp. 55–59 (2015).
- [31] 日本政府観光局: 年別 訪日外客数 ,出国日本人数の推移, [https://www.jnto.go.jp/jpn/statistics/marketingdata\\_outbound.pdf](https://www.jnto.go.jp/jpn/statistics/marketingdata_outbound.pdf). (2018/1/14 存在確認) .
- [32] 西川由理, 佐藤仁, 小澤順: 一般物体検出 YOLO の分散深層学習による性能評価, 情報処理学会研究報告, Vol. 2018-HPC-166, No. 12, pp. 1–6 (2018).
- [33] 大河原一輝, 平野廣美, 益子宗, 星野准一: ショッピングモール型 EC サイトのための店舗情報視覚化システム, 情報処理学会論文誌, Vol. 56, No. 3, pp. 847–855 (2015).
- [34] Rabbi, I. and Ullah, S.: A Survey on Augmented Reality Challenges and Tracking, *Acta graphica: znanstveni časopis za tiskarstvo i grafičke komunikacije*, Vol. 24, No. 1–2, pp. 29–46 (2013).
- [35] Redmon, J. and Farhadi, A.: YOLO9000: Better, Faster, Stronger, *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6517–6525 (2017).
- [36] Rehman, U. and Cao, S.: Augmented Reality-Based Indoor Navigation Using Google Glass as a Wearable Head-Mounted Display, *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1452–1457 (2015).
- [37] Rehman, U. and Cao, S.: Augmented-Reality-Based Indoor Navigation: A Comparative Analysis of Handheld Devices Versus Google Glass, *IEEE Transactions on Human-Machine Systems*, Vol. 47, No. 1, pp. 140–151 (2017).
- [38] Ren, S., He, K., Girshick, R. and Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137–1149 (2017).
- [39] Richert, W. and Coelho, L. P.: *Building Machine Learning Systems with Python*, Packt Publishing Ltd. (2013).

- [40] 佐々木隆行, 陳建軍, 高木昇: Toggle Mapping を用いた情景画像からの看板領域抽出, 日本知能情報ファジィ学会 ファジィ システム シンポジウム 講演論文集, Vol. 30, pp. 476–479 (2014).
- [41] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *Computer Vision and Pattern Recognition. arXiv e-prints*, arXiv:cs.CV/1409.1556v6 (2015).
- [42] Thorpe, S., Fize, D. and Marlot, C.: Speed of processing in the human visual system, *Nature*, Vol. 381, No. 6582, pp. 520–522 (1996).
- [43] 渡辺聰, 後藤春彦, 三宅論, 李彰浩: 商業地街路における歩行者の看板注視傾向に関する研究 : 銀座中央通りにおける歩行実験の分析, 日本建築学会計画系論文集, Vol. 68, No. 574, pp. 113–120 (2003).
- [44] 横川昇二: メディア景観の考察—都市環境を形成する広告・看板・情報の課題—, デザイン学研究特集号, Vol. 8, No. 1, pp. 22–27 (2000).
- [45] 吉野孝, 奥村賢悟: 迷いやすい人の特徴を考慮した屋内ナビゲーションシステム DoCoKa の開発と評価, 情報処理学会研究報告, Vol. 2013-GN-88, No. 18, pp. 1–8 (2013).