

屋内定点カメラを用いたパンフレット閲覧項目推定システム*

川西康友**** 村瀬 洋**** 徐 建鋒*** 田坂和之*** 柳原広昌***

Reading Contents Estimation from an Indoor Surveillance Camera

Yasutomo KAWANISHI, Hiroshi MURASE, Jianfeng XU, Kazuyuki TASAKA and Hiromasa YANAGIHARA

In this paper, we propose a system for estimating reading content in a booklet from an indoor surveillance camera. Here, we assume that a reading content can be specified by estimating the following two; which booklet/page, and which region in the page. We implement a reading booklet/page estimation method based on image search, and propose a reading region estimation method based on the slight difference of body pose of the reader. We evaluated the method as a 44 classes classification problem, which consists of eleven pages of booklets and four regions in each pages. We confirmed that the system achieved 25.6% in accuracy of the reading content estimation in our evaluation setting.

Key words: reading contents estimation, pose estimation, gaze estimation

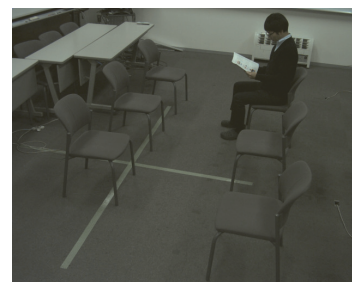
1. はじめに

人々が何に興味を持っているのかを知ることが出来れば、流行や人気商品等がわかり、それに合わせた情報やサービスの提供が可能となる。そのため、できるだけ多くの人々から、それぞれの人が興味を持っている対象を推定することは重要である。そこで本研究では、人々が何に興味を持っているのかを推定することを考える。

駅や空港、病院、各種受付窓口などには多くの人々が訪れることから待合室が併設されており、待合室にはパンフレットや書籍が置かれていることが多い。こうした場所では人々は待ち時間の間、置かれているパンフレットなどを自由に手に取り、閲覧する。人々がパンフレットなどに記載されているどの項目を閲覧したかを、そのうちの一部分の人だけでも推定できれば、その内容から人々が興味を持っている対象が何かを推定することができると思われる。

一方、このような待合室には防犯カメラが設置されていることが多い。待合室では、着席している人々が見渡せるように、上から見下ろすようにカメラが設置されていることが多い。近年、防犯カメラを、本来の防犯目的だけでなく、動線解析や混雑状況の分析など様々な用途に有効利用しようという研究が行なわれている¹⁾。本研究では、防犯カメラのように屋内に設置された高解像度な定点カメラを用いて、そこに写る人がパンフレットのどの項目を閲覧しているかを推定することを目的とする。

定点カメラで撮影した画像を用いて、対象の人がパンフレットのどの項目を閲覧しているかを推定することは、(1) 対象の人がどのパンフレットの何ページ目を開いているかの推定と、(2) そのページのどこを閲覧しているかの推定をすることで実現できると考えられる。ここで、カメラと人の位置関係によっては人が閲覧している箇所をカメラで直接観測することはできない場合も多いが、カメラで撮影した画像からパンフレッ



(i) A person reading a booklet



(ii) Example of a result

Fig. 1 Reading Contents Estimation

トの何ページ目を開いているかを推定するためには、少なくともページの一部分がカメラで撮影した画像に写っていることが必要であるため、本研究では図 1 (i) に示すように、人を斜め上方から観測し、閲覧しているパンフレットのページの一部分がカメラから観測できている状況を想定する。図 1 (ii) 中のパンフレット画像内の緑の丸印は、推定した注目対象位置を示す。これは、閲覧項目が正しく推定できている例である。

どのパンフレットの何ページ目を開いているかを推定する問題は、パンフレットの各ページの画像が既知であると仮定すると、特定物体認識の問題であると言える。そのため、本研究では特定物体認識の方法でページ推定を行うことを考える。

一方、人がパンフレットのページ内のどの位置を閲覧しているのかを知るためには、一般的に視線推定が必要である。視線計測装置を利用できるならば容易に視線を推定することが出来るが、本研究では待合室などで一般客を防犯カメラなどの定点カメラ映像を対象とするため、視線計測装置は利用できない。

一般に、パンフレットを閲覧する際には、読みやすい位置に両手で持つことが多い。この時、顔とパンフレットの距離は近いいため、パンフレットのページ内の閲覧位置に応じて、顔とパンフレットの相対的な角度は微妙に変化する。そこで本研究では、パンフレットを閲覧する際の姿勢に着目し、姿勢からパンフレットのページ内の閲覧位置を推定する手法を提案する。閲覧項目を知るという目的では、閲覧している座標値を詳細に得

* 原稿受付 平成 30 年 11 月 28 日

掲載決定 平成 31 年 1 月 31 日

** 名古屋大学 大学院情報学研究所 (愛知県名古屋市千種区不老町)

*** KDDI 総合研究所 (埼玉県ふじみ野市大原 2 丁目 1-15)

る必要はないため、本研究では、ページ内をいくつかの領域に分け、それらの領域のどれを閲覧しているのかを求めるクラス分類問題として扱う。

本研究の貢献を以下にまとめる。

- 閲覧項目推定問題を、閲覧ページ推定問題と閲覧位置問題へと分解し、それらを組み合わせることで、防犯カメラなどの定点カメラ映像に写る人の、パンフレット閲覧項目推定を実現する。
- 人体姿勢（画像中の2次元座標で表現された人体の各関節位置）から、ニューラルネットワークを用いてパンフレットの閲覧位置を推定する手法を提案する。

本論文の構成は以下の通りである。まず、2.章で関連研究についてまとめる。続いて3.章で提案するシステムの構成と、閲覧ページ推定手法、閲覧位置推定手法の詳細について述べる。4.章で実験結果を示し、性能の評価を行なう。最後に、5.章でまとめと今後の課題を述べる。

2. 関連研究

本研究で提案するシステムは、閲覧ページ推定と閲覧位置推定からなる。これらは特定物体認識と視線推定に関連するため、それらの関連研究について簡潔に述べる。

2.1 特定物体認識

特定物体認識とは、与えられた画像が、認識対象とする既知の物体の画像であるか否かを判定する問題である²⁾。一般に、特定物体認識の問題では、画像中から特徴点を検出してその周辺の局所特徴量を求め^{3)~7)}、データベース内の画像から同様に得た局所特徴量との一致度や、局所特徴量の出現頻度の類似度をもとに、画像同士の類似度を算出する^{8)~10)}。

書籍などを認識するタスクでは、データベース内に類似物体が多いため、他の物体との区別がしやすい特徴点を選択するなどの工夫がなされている¹¹⁾。対象が文書画像であれば、その性質を捉えて検索する手法も提案されている¹²⁾。

多くの手法は、高解像度の入力画像を対象としているが、防犯カメラなど天井や壁に固定された定点カメラ映像中では、パンフレット領域は小さい。そのため、対象の人が閲覧しているパンフレットの文字列の性質を捉えるほどの解像度を得ることは困難であると考えられる。一方、パンフレットの場合はアイキャッチとなる写真が多数掲載されているため、本研究では一般的な画像を対象とした手法を利用する。

2.2 視線推定

一般に、人の視線を推定する場合、視線計測装置¹³⁾が利用されることが多い。しかし、前節で述べた通り、本研究では防犯カメラなどの定点カメラ映像に写る人を対象とするため、視線計測装置を各人に装着させることはできない。

もし、顔画像が正面から高い解像度で得られているならば、画像からの視線推定も可能である^{14)~17)}。例えば、Zhangら¹⁸⁾は、様々な照明環境下でも頑健に視線推定を行う手法を提案している。しかし、本研究では防犯カメラなどの定点カメラから撮影した画像を対象としているため、人物を正面から捉えることは難しく、また高解像度な顔画像を得ることは難しいため、直接こうした手法を適用することは難しい。

防犯カメラなどの定点カメラ映像から人物の視線推定を行なう手法もいくつか提案されている¹⁹⁾²⁰⁾。これらの研究では、

どちらを向いているかといった大まかな推定を行っている。本研究では、パンフレット内の閲覧位置を推定することが目的であるため、微小な違いの推定が必要である。

3. 閲覧項目推定システム

3.1 本システムの概要

パンフレット内の項目は、ページ番号及びページ内での位置で特定できることから、本研究では、1.節で述べた通り、

- どのパンフレットの何ページ目を閲覧しているか
- ページ内のどこを閲覧しているか

を推定できれば、対象の人がどの項目を閲覧しているのかを知ることが出来ると考えられる。

この考えをもとに、本システムを、以下の二つのサブシステムから構成する（提案①）。

- 閲覧ページ推定システム：
対象の人が手に持って開いているパンフレットが、どのパンフレットの何ページ目かの推定。（特定物体認識の方法で実装。）
- 閲覧位置推定システム（提案②）：
対象の人が読んでいる位置が、パンフレット紙面上のどの部分に相当するかの推定。（人体姿勢から閲覧位置を推定する手法を提案。）

全体の処理の流れを図2に示す。以下、この二つの処理について詳述する。

3.2 閲覧ページ推定システム

待合室などに置いてあるパンフレットは有限であり、内容を予め知ることが出来るため、事前に全ての見開きページの画像を登録したデータベースを構築しておく。そして、その中から対象の人が読んでいるページと最も類似する見開きページの画像を検索することで、閲覧ページ推定を実現する。

撮影した画像全体から、対象の人が読んでいるパンフレットの見開きページを直接検索することは困難であるため、まずパンフレットの大まかな位置を特定し、その領域に絞り込んで画像検索を行う。領域の絞り込みには人とパンフレットの位置関係に着目し、姿勢推定により得た顔と手の位置からパンフレットの位置を推定する。

3.2.1 ページデータベースの構築

まず、全てのパンフレットの各ページの見開き画像からなる集合 $I = \{i_1, i_2, \dots, i_N\}$ を用意する。ただし、 N はパンフレットの見開きページ数である。ここで、パンフレットが複数冊ある場合は、各パンフレットの見開きページ数の合計とする。各画像 i_n に対し、特徴点 $\{\mathbf{k}_{n,j}\}_{j=1}^{J_n}$ を検出し、各特徴点から局所特徴量 $\{\mathbf{f}_{n,j}\}_{j=1}^{J_n}$ を算出する。ただし、 J_n は、画像 i_n から検出した特徴点の個数である。そして画像 i_n ごとに、特徴点と局所特徴量の組の集合 $\{(\mathbf{k}_{n,j}, \mathbf{f}_{n,j})\}_{j=1}^{J_n}$ を保存し、ページデータベースとする。本稿では、特徴点の抽出、特徴量の記述方法として Scale Invariant Feature Transform (SIFT)²¹⁾ を利用する。

3.2.2 パンフレット位置推定

パンフレットを閲覧する際、人は読みやすいように手でパンフレットを持つ。そのため、パンフレットは画像内で対象の人の手の周辺に写るはずである。このことに着目し、人の姿勢推

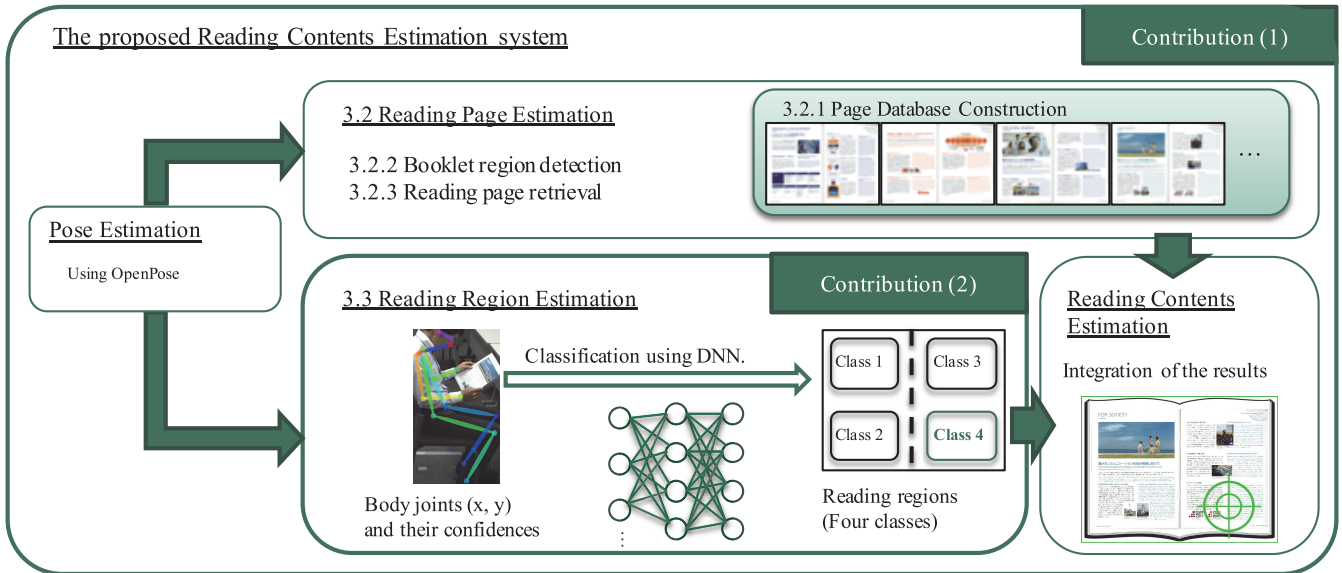


Fig. 2 The process flow of the proposed Reading Contents Estimation system

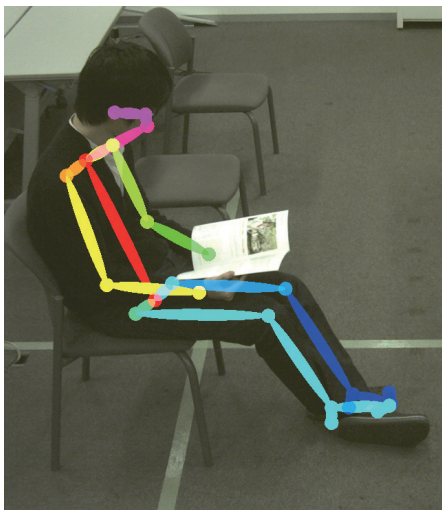


Fig. 3 A pose estimation result



Fig. 4 Keypoint matching to find the reading page

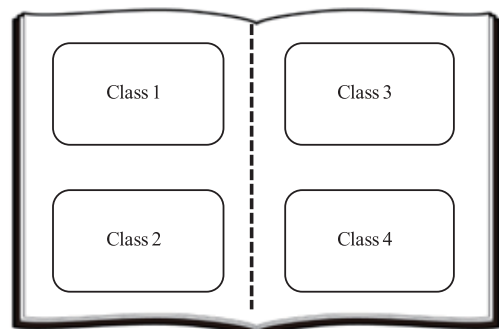


Fig. 5 Positions of the four classes

定結果からパンフレットの位置を推定する。

姿勢推定には、高速で高精度に動作する Cao ら²²⁾ の手法 (OpenPose) を利用する。この手法は、体の関節点 25 個の、画像上での座標及び尤度が得られる (図 3)。入力画像 i から得られた頭部 (目, 鼻, 耳) の座標および左右の腕 (肘, 手首) の座標から、それらを囲む矩形を決定する。その矩形を、頭部から見て手首側に 2 倍に拡張した矩形をパンフレットの推定位置とする。

3.2.3 閲覧ページ推定

前節の処理によって得られたパンフレット推定位置から画像を切り出し、パンフレット領域画像 i^p とする。

パンフレット領域画像 i^p から 3.2.1 節と同様に特徴点検出及び局所特徴量の算出を行い、特徴点と特徴量の組の集合 $\{(\mathbf{k}_j^p, \mathbf{f}_j^p)\}_{j=1}^{J^p}$ を得る。パンフレット領域画像から得た特徴点と、ページデータベース内の特徴点との対応付けを行い、最も多くの対応点が得られた画像を、閲覧ページ推定結果として得る。特徴点の対応付の際には、近似最近傍探索²³⁾ を用いて高速化を図る。図 4 に対応付けの例を示す。

3.3 閲覧位置推定システム

パンフレットを閲覧する際、閲覧する項目のページ内での位置に応じて、人の姿勢は微妙に変化すると考えられる。そのため、人の姿勢推定結果を元に、人がどの位置を閲覧しているのかを推定する。

本稿では閲覧位置推定を、図 5 に示すように、見開きページ内の、左右各ページの上下どちらかの 4 クラス分類問題とする。

入力する姿勢は、3.2.2 節と同様に Cao らの手法²²⁾ を利用して観測画像から獲得する。こうして得られる 25 個の関節点の画像中での 2 次元座標及び、その検出尤度を全て並べた 75 次元ベクトル \mathbf{v} を入力とし、4 つのクラスのうちのどれかを出力する推定器 $f(\mathbf{v})$ を構築する。関節点の座標に加えて、尤度も入力することで、各関節点の推定値が、関節点の欠損や誤検出

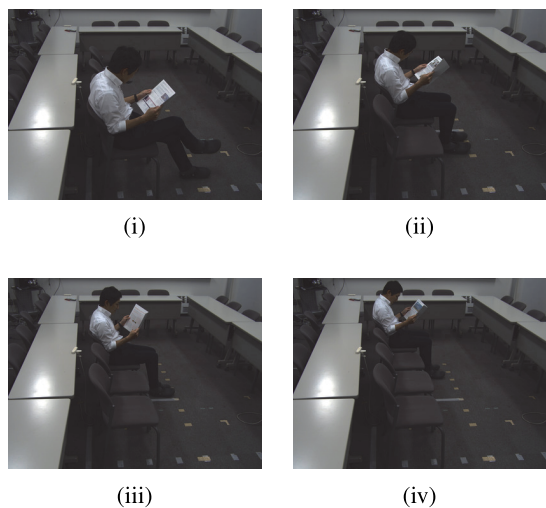


Fig. 6 Evaluation settings for evaluation against distance variation

によるものであるか否かが判別できると期待できる。カメラに対する人物の位置の違いによる移動・拡大縮小に対応するため、25個の関節点の座標は、その25個を囲む最小矩形の幅と高さがそれぞれ1になるよう正規化しておく。

推定器 $f(\mathbf{v})$ の学習時の教師信号は、閲覧している項目のクラス番号である。そのため、指定した4箇所うちのどれを閲覧しているのが既知の画像を用いて $f(\mathbf{v})$ の学習を行なう。

推定器 $f(\mathbf{v})$ の実装として、4層からなる深層学習モデルを用いる。出力層以外の各中間層のユニット数は実験的に50とした。また、Batch Normalizationによる正規化を施した。

4. 実験

4.1 データセット

推定システムの評価を行なうため、データセットを構築した。

まず、ページデータベースとして、パンフレットの各ページをスキャンし、見開き11ページ分のデータベースを構築した。

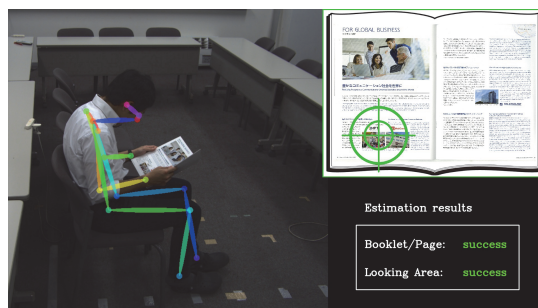
次に、実験室環境において1人がパンフレットを閲覧している状況を再現し、PointGrey社製カメラGrasshopper 3 (GS3-U3-28S5C-C; 1926 × 1440画素)を用いて約2.2mの高さから撮影した。カメラからの距離の違いに対する頑健性を評価するため、図6に示す4箇所に被験者を配置して撮影を行った。

被験者は、4箇所の被験者位置において、指定されたページを開き、図5に示す4箇所を順に閲覧した。閲覧するページは、ページデータベースに含まれる見開き11ページのうちの4ページから選択した。

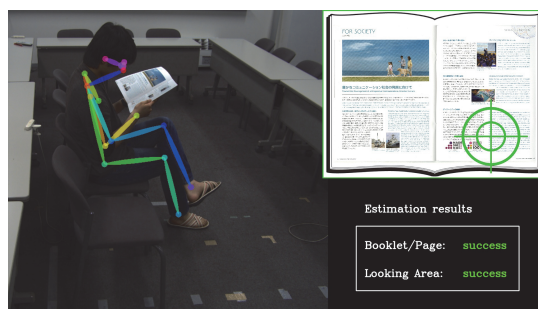
被験者3人に対し、被験者位置、ページ内の閲覧位置の組み合わせそれぞれについて、200枚、合計 $3 \times 4 \times 4 \times 200 = 9,600$ 枚の画像を撮影した。

4.2 実験方法

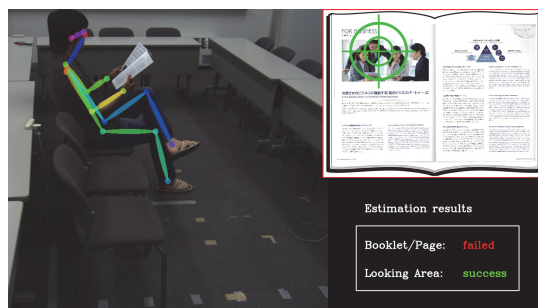
閲覧項目の推定精度を評価するため、見開き11ページのパンフレットの各ページ4箇所のうちのどれかを答える、44クラス分類問題として正解率を算出した。また、実際に実験に用いた4ページの各4箇所の閲覧位置のうちのどれかを答える、16クラス分類問題としても評価した。さらに、提案手法の構成要素である、閲覧ページ推定と閲覧位置推定の2つの処理のそれぞれについても個別に評価をした。



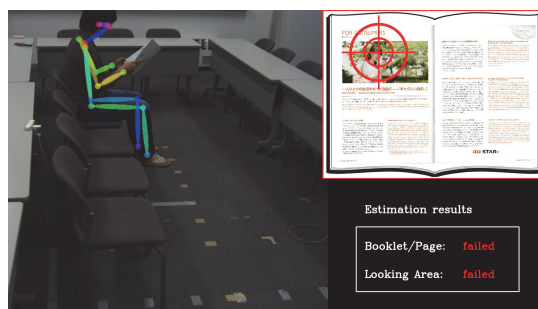
(i)



(ii)



(iii)



(iv)

Fig. 7 Examples of output of the proposed system

評価では、3人の被験者、4箇所の被験者位置の12パターンに対し、1つを評価用を選び、それ以外のデータで学習をする、12分割交差検証を行った。そして、閲覧項目推定、閲覧ページ推定、閲覧位置推定について、12回の分割における推定結果の正解率の平均を算出した。

4.3 結果

まず、提案システムによる推定結果例を図7に示す。図中の右上のパンフレット画面、右下の結果表示画面において、緑色

Table 1 The result of reading contents estimation for the chair position (i) in Fig. 6

	Evaluation	Accuracy %
Overall	Contents (44 classes)	49.0
	Contents (16 classes)	56.9
Individual	Page (11 classes)	64.9
	Page (4 classes)	75.0
	Region (4 classes)	76.0

Table 2 The result of reading contents estimation for all the chair positions

	Evaluation	Accuracy %
Overall	Contents (44 classes)	25.6
	Contents (16 classes)	37.9
Individual	Page (11 classes)	41.1
	Page (4 classes)	56.5
	Region (4 classes)	66.3

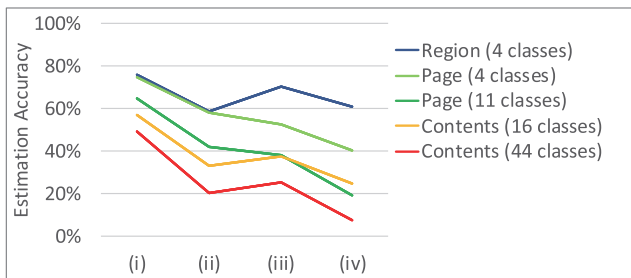


Fig. 8 Estimation accuracy by changing the distance between the camera and a subject. The horizontal axis corresponds to (i), (ii), (iii), and (iv) in Fig. 6

は推定が正しくできたことを、赤色は推定が失敗したことを表す。

次に、人が最も大きく写っている図 6 (i) の被験者位置での閲覧項目推定結果を表 1 に示す。表 1 は、閲覧項目推定精度、閲覧ページ推定精度、閲覧位置推定精度について、12 回の評価のうち、図 6 (i) に関するもの 3 回の平均をまとめたものである。提案手法は、閲覧項目の 44 クラス分類問題において、49.0% の正解率を達成した。

表 2 の結果は、全ての被験者位置での推定結果の平均である。全ての位置の平均としては、閲覧項目の 44 クラス分類の正解率は 25.6% であった。

カメラからの距離に応じて閲覧ページ推定、閲覧位置推定精度は変化すると考えられる。そこで、各推定精度について、カメラと被験者の距離ごとの分析を行なった。結果を図 8 に示す。図 8 の横軸は、カメラからの距離の違いを表し、図 6 (i) から (iv) に相当する。特徴点検出は画像の解像度に依存するため、カメラからの距離が遠ざかるにつれ、閲覧ページ推定の精度が低下することが確認できた。一方、閲覧位置の推定精度の変化は小さい。これは、12 分割交差検証において、閲覧位置推定器の学習データに同距離での他人のデータが含まれているからであると考えられる。

なお、本システムは GTX1080 を搭載した PC 上で、およそ 10 fps での閲覧項目推定処理を実現した。本システムの処理のボトルネックはデータベース内の画像との照合にあるため、もし照合を高速化することができれば、高速な閲覧位置推定が実現できる。

4.4 閲覧位置推定の分析

閲覧位置推定のさらなる評価をするため、図 9 に示す 12 箇所に被験者合計 7 人配置してさらに撮影を行った。ただし、実

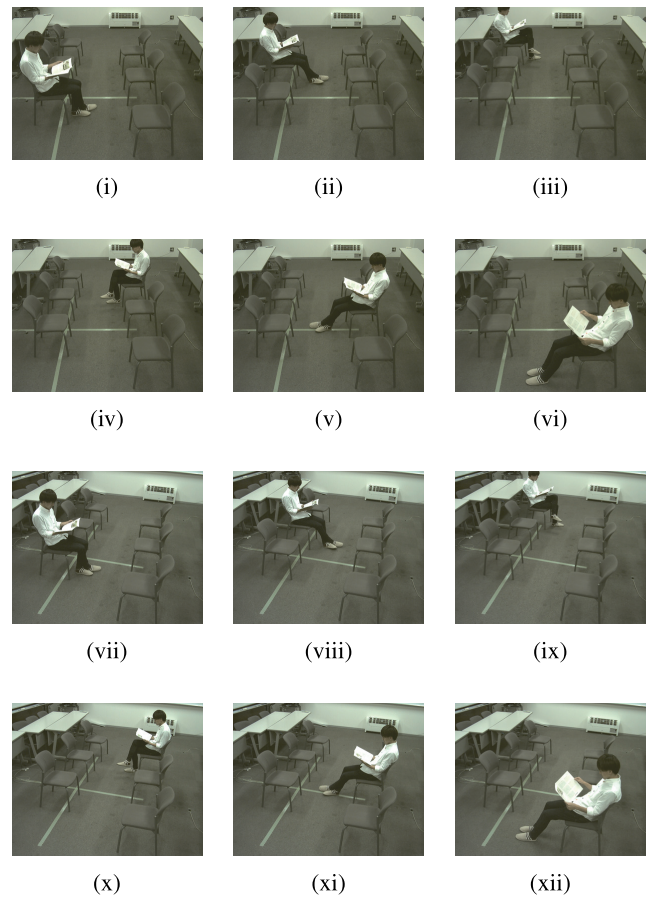


Fig. 9 Different settings (chair positions): distance variation between the camera and subjects

Table 3 Confusion matrix of the reading contents estimation

	Class 1	Class 2	Class 3	Class 4
Class 1	0.56	<u>0.26</u>	0.12	0.06
Class 2	<u>0.22</u>	0.58	0.06	0.14
Class 3	0.09	0.10	0.55	<u>0.25</u>
Class 4	0.04	0.15	<u>0.26</u>	0.55

際の座席は 6 箇所であるが、同時に 2 台のカメラで撮影した。以降では、この 2 台のカメラで撮影した同一の座席を、異なる被験者位置として扱う。

被験者は、12 箇所の被験者位置において、指定されたページを開き、図 5 に示す 4 箇所を順に閲覧した。閲覧するページは、ページデータベースに含まれる見開き 11 ページのうちの 4 ページから選択した。

被験者 7 人に対し、被験者位置、ページ内の閲覧位置の組み合わせそれぞれについて、40 枚、合計 $7 \times 12 \times 4 \times 40 = 13,440$ 枚の画像を撮影した。

4.2 節と同様の評価方法である、被験者 7 人、被験者位置 12 箇所の 84 分割交差検証の結果、認識率は 56.29% であった。4.2 節の実験結果と比較すると、反対向きの被験者位置が追加された分、全体的な認識率が低下したと考えられる。

閲覧位置推定結果の分析として、認識結果の混同行列を表 3 に示す。実験結果から、左右の分類よりも、上下の分類の方が困難であることが確認できた。

これまでの実験では全身の関節点 25 点の座標と検出尤度を

Table 4 Reading region estimation

Body parts	Accuracy %
(A) Full body	56.29
(B) Upper body	55.82
(C) Head and Arms	55.04

全て入力したが、直感的には、足など下半身の関節点は閲覧位置推定に影響しないと考えられる。そこで、(A) 全身 25 点、(B) 上半身のみ 14 点、(C) 頭部と前腕部のみ 10 点、を入力として閲覧位置推定精度を比較した。その結果を表 4 に示す。本実験では使用する関節点を変更しても、あまり結果に差はないことが確認できた。これは、閲覧項目の推定器の学習時に特徴選択が行なわれることにより、不要な関節点の影響が低減されたことによると考えられる。

5. おわりに

本論文では、屋内に設置された定点カメラを用いて、観測対象の人が、パンフレットのどの項目を閲覧しているのかを推定するシステムを提案した。このシステムを実現するため、閲覧項目の推定システムを、閲覧ページ推定と閲覧位置推定によって構成した(提案①)。閲覧ページの推定は、定点カメラで撮影した画像中から、対象の人の姿勢に基づいてパンフレット位置を切り出し、画像検索することで実現した。一方、画像から対象の人の閲覧位置を推定するため、パンフレット閲覧時の姿勢に着目し、姿勢から閲覧位置を推定する手法(提案②)を提案した。実験で用いた環境において、閲覧項目の 44 クラスを、25.6% の精度で推定できることを確認した。また、対象の人がカメラに近ければ、49.0% の精度で推定できることも確認した。

構築したシステムの閲覧ページの推定方法は、単純なキーポイントマッチングに基づくため、限られた状況でしか使えないと考えられる。そのため、オクルージョンが大きいときなど様々な状況でも利用可能な閲覧ページ推定方法へと改善する余地がある。また、本論文では、学習データ収集の関係上、パンフレット紙面を 4 分割し、4 クラス分類として評価を行なったが、視線推定装置を用いて学習データを収集することにより、4 クラスよりも詳細な分類や、回帰による連続的な閲覧位置推定へと拡張することも考えられる。

謝 辞

本研究の一部は、JSPS 科研費 JP17H00745 の助成を受けたものである。

参 考 文 献

- 1) I. Mitsugami, M. Mukunoki, Y. Kawanishi, H. Hattori, and M. Minoh: Privacy-protected camera for the sensing web, Proc. the 13rd Information Processing and Management of Uncertainty in Knowledge-Based Systems, (2010) 622.
- 2) 黄瀬浩一: Bag-of-Features に基づく物体認識 (1) 特定物体認識, CVIM チュートリアルシリーズコンピュータビジョン最先端ガイド, アドコム・メディア (2010).

- 3) D.G. Lowe: Object recognition from local scale-invariant features, Proc. of the 7th IEEE International Conference on Computer Vision, **2**, (1999) 1150.
- 4) H. Bay, T. Tuytelaars, and L. Van Gool: SURF: Speeded up robust features, Proc. 9th European Conference on Computer Vision, (2006) 404.
- 5) Y.-G. Jiang, C.-W. Ngo, and J. Yang: Towards optimal bag-of-features for object categorization and semantic video retrieval, Proc. the 6th ACM International Conference on Image and Video Retrieval, (2007) 494.
- 6) S. Leutenegger, M. Chli, and R.Y. Siegwart, BRISK: Binary robust invariant scalable keypoints, Proc. 13th International Conference on Computer Vision, (2011) 2548.
- 7) E. Rublee, V. Rabaud, K. Konolige, and G. Bradski: ORB: An efficient alternative to sift or surf, Proc. 13th International Conference on Computer Vision, (2011) 2564.
- 8) H. Jégou, M. Douze, C. Schmid, and P. Pérez: Aggregating local descriptors into a compact image representation, Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2010) 3304.
- 9) X. Zhou, K. Yu, T. Zhang, and T.S. Huang: Image classification using super-vector coding of local image descriptors, Proc. 11th European Conference on Computer Vision, (2010) 141.
- 10) F. Perronnin, Y. Liu, J. Sánchez, and H. Poirier: Large-scale image retrieval with compressed fisher vectors, Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2010) 3384.
- 11) 渡邊之人, 入江 豪, 黒住隆行, 杵渕哲也: 弁別的キーポイントの選択に基づく類似物体群からの特定物体認識, 映像情報メディア学会誌, **71**, 2, (2017) J93.
- 12) T. Nakai, K. Kise, and M. Iwamura: Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval, Proc. 7th IAPR International Workshop on Document Analysis Systems, (2006) 541.
- 13) A.T. Duchowski: Eye tracking methodology: Theory and practice, Springer, (2007).
- 14) F. Lu, Y. Sugano, T. Okabe, and Y. Sato: Adaptive linear regression for appearance-based gaze estimation, IEEE Transaction on Pattern Analysis and Machine Intelligence, **36**, 10, (2014) 2033.
- 15) H. Park and D. Kim: Gaze classification on a mobile device by using deep belief networks, Proc. 3rd IAPR Asian Conference on Pattern Recognition, (2015) 685.
- 16) L. Fridman, P. Langhans, J. Lee, and B. Reimer: Driver gaze region estimation without use of eye movement, IEEE Intelligent Systems, **31**, 3, (2016) 49.
- 17) K. Krafska, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba: Eye tracking for everyone, Proc. 2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2016) 2176.
- 18) X. Zhang, Y. Sugano, M. Fritz, and A. Bulling: Appearance-based gaze estimation in the wild, Proc. 2015 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2015) 4511.
- 19) K. Smith, S.O. Ba, J.M. Odobez, and D. Gatica-Perez: Tracking the visual focus of attention for a varying number of wandering people, IEEE Transactions on Pattern Analysis and Machine Intelligence, **30**, 7, (2008) 1212.
- 20) S. Duffner and C. Garcia: Unsupervised online learning of visual focus of attention, Proc. 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, (2013) 25.
- 21) D.G. Lowe: Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, **60**, 2, (2004) 91.
- 22) Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh: Realtime multi-person 2d pose estimation using part affinity fields, Proc. 2017 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2017) 7291.
- 23) M. Muja and D.G. Lowe: Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration, Proc. 4th International Conference on Computer Vision Theory and Applications, **2**, (2009) 2.