

## 行動認識のための複数カメラを用いた環境情報の自動獲得法

増田 宗一† 顧 優輝‡ 真部 雄介† 菅原 研次†

†千葉工業大学情報科学部 ‡千葉工業大学大学院情報科学研究科

### 1 はじめに

人間の多様な行動の自動認識を実現するためには、行動が行われている周囲の環境や道具などの認識も含めた、広範囲の認識が必要不可欠であると考えられる。本研究では、行動認識に必要な不可欠な周囲の情報を環境情報と定義し、複数のカメラから得られる画像情報を用いて環境情報を推定する手法を提案する。提案手法は、対象とする環境に関する情報の入力を事前に必要としない方法であり、人物の動作が行われている周囲の環境まで含めた高度な行動認識を実現するために必要不可欠な技術となると考えられる。

### 2 提案手法

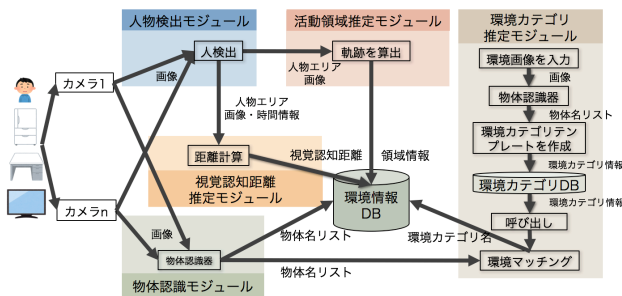


図1 全体の処理の流れ

図1に全体の処理の流れを示す。本手法の主なモジュールは人物検出モジュール、活動領域推定モジュール、視覚認知距離推定モジュール、物体認識モジュール、環境カテゴリ推定モジュールの5つである。

各モジュールの詳細について述べる。人物検出モジュールはOpenPose[1]を使用する。OpenPoseとは深層学習に基づく骨格推定技術である。また、ここで人物の検出された時間とカメラを記録する。この情報は後に視覚認知距離推定モジュールで使用する。

物体認識モジュールはカメラが設置されている環境内の物体を識別するために、深層学習を利用した

YOLOv2[2]を使用する。

活動領域推定モジュールでは、人物が活動すると思われる領域を推定する。初めに、カメラの画角を144マス(水平16マス×垂直9マス)に分割する。次にOpenPoseから推定された骨格座標データをもとに、検出された骨格座標の平均を中心点とし、中心点が144マスのどこに当たるか計算することで活動領域を推定する。

視覚認知距離推定モジュールは、同一人物が複数のカメラで検出される際の時間差を利用しカメラ間のおおよその距離を推定するモジュールであり、推定される距離を視覚認知距離と呼ぶ。視覚認知距離 $D_v$ は、以下の式(1)により算出される。

$$D_v = S \times \frac{\sum_{i=0}^n t_i}{n} \quad (1)$$

ここで、 $t_i$ はカメラ間の移動時間、 $n$ はカメラ間を移動した回数、 $S$ は人物の平均歩行速度( $m/min$ )をそれぞれ示す。本研究では $S = 80$ とした。

環境カテゴリ推定モジュールは、はじめに取得する環境名の一覧を作成する。環境名をキーとして検索した画像をWebから収集し、物体検出モジュールを用いて、物体を検出する。数の多い上位 $k$ 種類の物体数を $N_i(i = 1, 2, \dots, k)$ とし、環境 $E$ に対するテンプレート $T_E$ を式(2)により算出する。

$$T_E = (P_1, P_2, \dots, P_k) \quad (2)$$

$$P_i = \frac{N_i}{\sum_{j=1}^k N_j}$$

テンプレート $T_E$ は環境内に存在する $k$ 種類の物体のヒストグラムを表す。ここで作成されたテンプレートは環境カテゴリ推定に使用される。

その後環境カテゴリを推定したい画像から物体認識モジュールを使用して物体と物体数を検出し、テンプレート作成時と同じようにヒストグラムを求めるが、ここではテンプレートに含まれる物体の種類のみで構成されるヒストグラムを算出する。そして、式(3)に基づく類似度計算を行う。

$$Z_k = Sim(T_i, T_k) \quad (3)$$

$$k = (1, \dots, N_c)$$

ここで、 $T_i$ は環境推定を行う対象画像から算出したヒス

A Proposal of environmental information estimation method using multiple cameras  
 †Shuichi MASUDA †Yuki KAERI †Yusuke MANABE  
 †Kenji SUGAWARA  
 †Graduate School of Information and Computer Science, Chiba Institute of Technology

トグラム,  $T_k$  は事前に作成した環境カテゴリテンプレート,  $k$  は環境カテゴリテンプレの ID,  $N_c$  は環境カテゴリ数,  $Sim$  は類似度関数を示す. 本研究では環境カテゴリを 52 種類取得したため,  $N_c = 52$  とした. 類似度関数は, ユークリッド距離,  $\cos$  類似度, KL ダイバージェンス, Tanimoto 係数の 4 つである.

得られた類似度を用いて, 本研究では 2 つの手法により環境カテゴリを推定する. 手法 1 では, カメラが捉えている画像 1 枚に対して環境カテゴリを推定する. 1 枚ごとの類似度計算で最も類似度の高いものを環境カテゴリとして推定する. 手法 2 では, 同一環境カテゴリの画像が複数枚得られる場合は, 各画像の類似度に応じたスコア (最も類似度が高い環境カテゴリには 3 ポイント, 2 番目には 2 ポイント, 3 番目には 1 ポイント) を合算し, 最もスコアが高かった環境カテゴリを推定結果とする.

### 3 実験

本研究の実験方法について述べる. 実験はモジュールごとに行う. 人物検出モジュールと物体認識モジュールについては, 既存の技術を使用しているため, 評価は行わないものとする.

#### 3.1 概要

活動領域推定モジュールでは, 人物が歩行している動画像を計測し, 活動領域を推定する. 動画像の計測は, 2 台のカメラで行う. 各カメラの設置条件は俯角を 45 度, 高さを 2m80cm に固定する. 歩行を行う人物は 1 名とし, 歩行回数は 10 回行い, 計測は 3 回行う. 計測した動画像に対し, 提案手法によって推定された活動領域と被験者が推定する, カメラが捉えている画角内の活動領域と比較を行い, 再現率と適合率で評価を行う. 再現率は  $C/H \vee C$  適合率は  $C/M \vee C$  で表現できる. ここで,  $M$  は活動領域推定モジュールが推定した領域,  $H$  は人間が推定した活動領域,  $C$  は人とモジュールがともに推定した領域のことを指す.

視覚認知距離推定モジュールは, 活動領域推定モジュールと同じ動画像を使用し実験を行う. 式 (1) に基づく推定値と実空間での距離を比較する.

環境カテゴリ作成モジュールのテンプレート作成部は初めに被験者に物体リストを見させる. その後表示された選択肢のうち, もっともらしい環境を選択させる. 選択肢は 5 つ用意し, 正解の環境カテゴリ名を除き, 残りの 4 つの選択肢は無作為に選択した選択肢を使用する.

環境カテゴリ推定モジュールの推定部は, 2-fold cross validation で検証を行う. 各環境カテゴリを 2 分割し,

環境カテゴリテンプレート作成, 検証で使用する.

#### 3.2 結果と考察

まず, 活動領域推定モジュールの実験結果について述べる. モジュールの推定した活動領域と被験者が推定した活動領域を比較した結果, 再現率は平均 73% で適合率が平均 50% となった. 本実験で人とモジュール間で推定に差が生じた. 適合率が低くなった理由としては, 人が推定した活動領域の範囲が広く, 実験で使用した動画像では歩行していないエリアがあり, 適合率の低下に繋がったと考えられる.

次に, 視覚認知距離推定モジュールの実験結果を述べる. 3 回計測した動画を実験に使用した結果は平均誤差 15.2% となった. 誤差が少なく, 高い精度で視覚認知距離を推定すること可能であることがわかった.

そして, 環境カテゴリ推定モジュールのテンプレート作成の実験結果について述べる. アンケートを行った結果, 平均一致率は 56% となった. アンケートで正解率にばらつきが生じた原因としては正解率の低い環境カテゴリは特徴量が少なく, 他のテンプレートと間違えやすいと考えられる.

最後に, 環境カテゴリ推定の実験結果を述べる. 手法 1 では平均 16% で, 手法 2 では平均 40% となった. 環境カテゴリ推定モジュールの推定部では 1 枚ごとの類似度計算での正解の確率より, 検証用の画像をすべて使用し, 3 位まで考慮したもののほうが認識率が良いことから, 環境の認識には 1 枚の画像ではなく, 1 つの環境内を写した複数の画像が必要であると考えられる.

### 4 おわりに

本研究では複数のカメラで計測した画像から環境情報を推定する手法を提案した. 視覚認知距離の推定では, 比較的高い推定精度が得られた. 活動領域の推定では, 人間と提案手法の間で差が生じた. また, 環境カテゴリの推定では, 低い推定精度となった. 今後の課題として各推定手法の実験回数を増やし, 評価を行う必要がある.

#### 参考文献

- [1] Cao, Z., Simon, T., Wei, S.-E. and Sheikh, Y.: Real-time multi-person 2d pose estimation using part affinity fields, arXiv preprint arXiv:1611.08050 (2016).
- [2] Redmon, J. and Farhadi, A.: YOLO9000: Better, Faster, Stronger, arXiv preprint arXiv:1612.08242 (2016).