

ウェアラブルカメラを利用した展示会場での見学履歴取得 Recording Experience in Exhibition using Wearable Camera

畑田 晃希[†] 山崎 俊彦[†] 相澤 清晴^{‡,†}

Koki Hatada Toshihiko Yamasaki Kiyoharu Aizawa

1. はじめに

個人の行動履歴の取得を試みる研究は、近年盛んに行われている[1]. Rekimoto らは Place Engine と呼ばれる Wi-Fi の電波強度を利用した測位システムを用いて、屋外における個人の位置履歴を記録・分析することを試みた[2].

ところが、屋内での位置情報を取得・蓄積し、分析する試みは、屋外ほど多く行われていない。屋内における測位の困難さ、測位システムの規模の大きくなりやすさ、より高い精度の要求が、その一因であると考えられる。

その中で de Silva らのユビキタスホームは、家の床に圧力センサを敷き詰めることによって、家で生活する人の位置情報の取得を行った[3]. また、ウェアラブル端末を用いた、屋内での位置取得の最も初期の研究は、Aoki らによって行われた[4]. 彼らはノート PC に接続されたウェアラブルカメラで取得した動画をもとにした移動経路の探索に、動的計画法を適用することで、リアルタイムで位置の同定を行った。

本研究では、対象を展示会場に限定することにより、ウェアラブル端末で取得された画像のみを用いて、行動履歴を取得するシステムを開発した。HMM でモデル化することにより、画像のぼけに対して頑健性を高めた。鉄道博物館[5]において、見学者の行動履歴を取得する実験を行い、高い精度で行動履歴が取得できることを示した。

2. ウェアラブルカメラ

展示会場にセンサを設置するかわりに、見学者がウェアラブルカメラを身につけて行動することにより、展示会場での見学履歴の取得を行う。本研究では図 1 に示す Flip Recorder Camera をウェアラブルカメラとして利用した。

図 1 のようにユーザの首からストラップでぶら下げて利用し、一定の時間間隔で自動的に画像が撮影される。撮影される画像の解像度は 1280 × 1024 pixel であり、撮影時間間隔は 1 秒である。

3. 見学履歴の推定

3.1 展示物画像とのマッチング

本研究では、ウェアラブルカメラで取得した画像の時系列中の、それぞれの画像について、撮影されたときに見学者が注目していた展示物を推定することを目指す。つまり画像の時系列を入力とし、注目している展示物の時系列（これを見学履歴と呼ぶことにする）を推定結果とする。

そのため、あらかじめ撮影した展示物の画像と、見学者が身につけたウェアラブルカメラの画像のマッチングを行



図 1 Flip Recorder Camera

うことにより注目展示物を推定する。画像のマッチングには、Speed-Up Robust Features (SURF) [6]と呼ばれる、局所特徴量を利用する。

SURF とは、スケール変化、回転変化に不変な特徴量で、Haar-wavelet 変換から算出される 64 次元の特徴量である。SURF を用いた画像同士のマッチングは、異なる画像から抽出された特徴点の SURF 特徴量を比較して、対応点探索をすることにより行う。SURF 特徴量の比較には、ユークリッド距離を利用する。ある特徴点 1 点について、異なる画像に含まれる全特徴点と特徴量の距離を計算し、距離が最小となる点同士を対応点として検出する。

3.2 HMM

時間的なコンテキストを利用して、位置の認識を行う場合、隠れマルコフモデル (HMM) が利用される[7][8]. 本研究では、ウェアラブルカメラで取得した画像の時系列を観測 y とし、注目している展示物を隠れた状態 x とする left-to-right 型 HMM を利用して、見学履歴の推定を行う。

この HMM の状態 x は、展示物の総数を n 個とすると、注目している展示物を表わす $e_i (i=0, \dots, n-1)$ 、および「展示物 e_i の付近にいるが、注目していない」状態を表わす $w_i (i=0, \dots, n-1)$ の $2n$ 個の値を取る。初期状態は、一様な確率分布と仮定し、

$$p(x_{t=0}) = \frac{1}{2n}$$

と設定する。状態 x_t からの遷移は、 x_t および x_{t+1} のみに依存し、展示物の位置情報から算出される。

$$p(x_{t+1}|x_t) = \alpha \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{l_{t,t+1}^2}{2\sigma^2}\right)$$

ここで、 $l_{t,t+1}$ は時刻 t の展示物と時刻 $t+1$ の展示物の展示会場における距離 (m) であり、 α は規格化定数である。このように遷移確率を設定することにより、見学者は近くにある展示物に注目しやすく、遠くにある展示物には注目しにくいという経験則を HMM でモデル化できる。

出力確率は、SURF による対応点探索の結果を用いて算出する。展示物 e_i の画像と、見学者のウェアラブルカメラ画像 y_j の間で SURF の対応点が多いほど、その画像 y_j は展示物 e_i に注目しているときに、撮影されたと考えることは自然である。そこで、出力確率は、

[†] 東京大学大学院 情報理工学系研究科

Department of Information and Communication Engineering,
The University of Tokyo

[‡] 東京大学大学院 情報学環・学際情報学府

Interfaculty Initiative in Information Studies, Interdisciplinary
Information Studies, The University of Tokyo

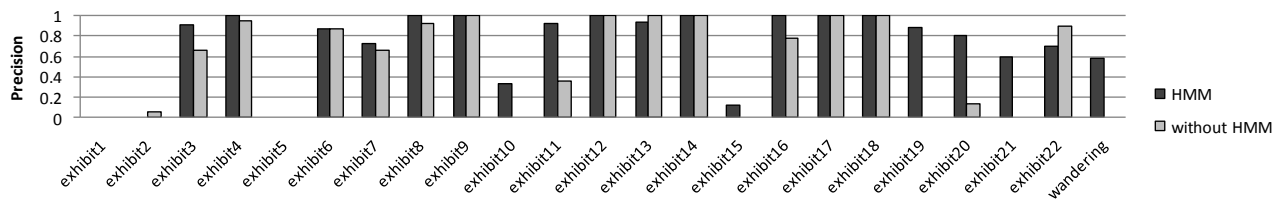


図2 見学履歴の推定精度 (exhibit i は注目展示物の種類, wandering は”注目していない”状態を表す。HMM でモデル化した場合(HMM), 画像のマッチングのみの場合(without HMM)の結果を示す。)

$$p(y_t | x_t = e_i) \propto p(x_t = e_i | y_t) p(y_t) = \beta C_{i,t}$$

とする。ここで, $C_{i,t}$ は時刻 t にウェアラブルカメラで撮影した画像 y_t と, 展示物 e_i の画像の SURF 対応点の数, β は規格化定数である。

3.3 ぼけ画像による非注目状態の推定

ウェアラブルカメラで撮影した画像にはモーションブラーのある画像が多数含まれる。特に展示会場の中は, しばしば照明を暗くしてあり, 露光時間の増加によりモーションブラーが生じやすい。

ぼけ画像は, SURF の特徴点の検出や, 特徴量の算出の際, 意図通りの結果が期待できない。そのため, 展示物画像とのマッチング結果も誤る可能性が高く, 3.2 節の HMM の出力確率の算出に利用するのは適当でないと考えられる。

そこで, ぼけ画像が観測されたときは, 見学者は”展示物 e_i の付近にいるが, 注目していない” $w_i (i=0, \dots, n-1)$ の状態のいずれかであると仮定する。ところで, ぼけ画像ほど検出される SURF 特徴点が多く, 鮮明な画像ほど検出される SURF 特徴点が多いという傾向が経験的に存在する。そこで, SURF 特徴点の数が少ないほど w_i である確率が大きいとする。すると, HMM の出力確率は,

$$p(y_t | x_t = w_i) = \gamma (C_{\max} - C_t) \quad (i = 0, \dots, n-1)$$

となる。ここで, C_t は画像 y_t に含まれる SURF 特徴点の数, C_{\max} は C_t の最大値, γ は規格化定数である。

また, 前述の $p(y_t | x_t = e_i)$ も以下のような例外処理で, SURF 特徴点が少ない画像(つまり, ぼけ画像)と展示物画像とのマッチング結果を無視する。

$$\text{if } C_t < C_{th} \text{ then } p(y_t | x_t = e_i) = \beta \epsilon$$

ここで C_{th} は特徴点の数の閾値, ϵ は微小な定数である。

以上のようにパラメータを設定した HMM で, Viterbi アルゴリズムにより, もっともらしい状態 x の系列 (Viterbi パス) を推定する。この推定結果の系列を, 推定した見学履歴とする。

4. 実験

4.1 実験条件

鉄道博物館のヒストリーゾーンの一部において, ウェアラブルカメラを装着して見学し, その見学履歴を推定した。見学には 12 分程度時間を要し, その間にウェアラブルカメラは, 約 700 枚の画像を撮影した。展示物画像は, 同一のウェアラブルカメラを用いて, あらかじめ撮影したものを利用した。展示物の数は, $n=22$ であった。

3.2 節で述べた HMM を利用し, 見学履歴を推定した。遷移確率のパラメータ σ は $2m$ とした。また, HMM を利

用せず画像の SURF によるマッチング結果のみでも, 見学履歴を推定し比較した。ただし, マッチングのみの場合, ”注目していない”状態は推定していない。

4.2 結果

図2に推定した見学履歴の推定精度を展示物ごとに集計して示した。 ”注目していない”状態 $w_i (i=0, \dots, n-1)$ は, ひとつに合せて集計した。図2によると, 展示物ごとに推定精度のばらつきが大きいが, 80%以上の精度で推定できるものが多い。推定の誤りは主に SURF によるマッチングの誤りに起因していた。また, 同一の展示物を撮影したにも関わらず, 視点の違いにより全く異なる画像となり, マッチングが不能となるケースも見受けられた。SURF のマッチングのみの場合と比較すると, HMM の利用は効果的であるといえる。

”展示物 e_i の付近にいるが, 注目していない”状態 w_i の推定精度は, 注目している展示物がある場合に比べて低い。画像にモーションブラーがあるとき, この状態であると仮定したが, SURF の特徴点の数を利用するという, 今回のぼけ画像の検出方法に課題があったと考えられる。

5. まとめ

本研究では, ウェアラブルカメラを用いて展示会場において見学履歴を取得する手法を開発した。撮影画像のマッチングに SURF を, 位置・時間の関係のモデル化に HMM を利用することで, 精度良く見学履歴を取得できることを実証した。

参考文献

- [1] 相澤清晴, “体験の記録とコンテキストに基づく効率的な検索”, システム制御学会誌 システム/制御/情報, Vol. 50, No. 1, pp. 7-11 (2006)
- [2] J. Rekimoto, et al, “LifeTag: WiFi-based Continuous Location Logging for Life Pattern Analysis,” Proc. 3rd International Symposium on Location- and Context-Awareness, pp.35-49 (2007)
- [3] G. C. De Silva, et al, “Ubiquitous Home: Retrieval of Experiences in a Home Environment,” IEICE Trans. Information and Systems, V.91-D, No.2, pp.330-340 (Feb. 2008)
- [4] H. Aoki, et al, “Realtime personal positioning system for a wearable computer,” The Third International Symposium on Wearable Computers, pp.37-43 (1999)
- [5] 鉄道博物館, <http://www.railway-museum.jp/>
- [6] H. Bay, et al, “SURF: Speed-Up Robust Features,” 9th European Conference on Computer Vision, pp.404-417 (2006)
- [7] A. Torralba, et al, “Context-based vision system for place and object recognition,” Computer Vision, 2003. Proceedings. 9th IEEE International Conference on Computer Vision, vol.1, pp.273-280 (Oct. 2003)
- [8] J. Kosecka, et al, “Vision based topological Markov localization,” Proceedings. IEEE International Conference on Robotics and Automation, vol.2, pp. 1481-1486 (2004)