

## 一人称視点映像による実環境の記憶可能性推定

大泉 建人†

中澤 篤志‡

西田 豊明‡

†京都大学工学部情報学科

‡京都大学大学院情報学研究科

## 1 序論

実環境中の建物や看板、掲示物などのオブジェクトに対して、人がどのようなものを良く記憶できるかという情報は、まちづくりやマーケティングにおいて有用である。例えば、災害対策に重要な避難情報を覚えやすい所に掲示したり、広告をより記憶されやすいように設置するなどの利用方法があり、新たな地図情報として幅広い用途に活用できる。我々はこの、実環境オブジェクトの記憶されやすさを「オブジェクトの記憶可能性」と定義し、一人称視点映像を基に推定する手法を提案する。

我々はオブジェクトの記憶可能性を、対象オブジェクトを視認できる範囲を通過した人の中で一定時間後に再び思い出す事の出来る割合と定義する。これは、ある未知シーンにユーザを歩かせ、その後アンケート等を用いることで求めることができるが、実施には大きな人的コストがかかるため、広範囲の環境で実施することは困難である。そこで、本稿では一人称視点映像を基にオブジェクトの記憶可能性を推定することを考える。人は、目の前に気になる物があつた場合注視を行い、歩いている際に興味を惹かれるものを視界に捉えた場合そこに視線を留めると考えられるため、一人称視点映像中から注視行動を検出し記憶可能性を推定することは可能であると考えられる。

## 2 記憶可能性モデル

一人称視点映像  $I$  から対象オブジェクト  $O_i$  の記憶可能性  $P_{remember}(O_i, I)$  を推定するモデルを式 1 に示す。

$$P_{remember}(O_i, I) = P_{memorability}(O_i) \cdot P_{seen}(O_i, I, S) \quad (1)$$

ここで、 $P_{memorability}(O_i)$  は対象オブジェクト  $O_i$  のもつ記憶されやすさを表す項である。これは Isora らによって提案されている静的シーンでの記憶可能性である memorability[1] を表す。また、 $P_{seen}(O_i, I, S)$  は一人称視点映像  $I$  とその撮影者の状態  $S$  に関する項である。この項に関連する要素として、 $I$  における  $O_i$  の出現回数、

停留時間、映像中の位置や、 $I$  の optical flow 等が挙げられる。 $S$  には  $O_i$  への興味や  $O_i$  が  $I$  に現れている際の集中等、 $I$  からは推測し難い要素も含まれる。

モデルの妥当性 第2章に示したモデルでは、記憶可能性は  $P_{memorability}$  と  $P_{seen}$  の乗算によって表される。これは  $P_{memorability}$  が低いオブジェクトであっても  $P_{seen}$  が高い、すなわち出現回数が多く停留時間も長い場合には記憶可能性が高くなることを表している。一方で  $P_{memorability}$  が高い場合でも  $P_{seen}$  が低い場合には記憶可能性は低くなる。実際、特徴的な建物の近くを通過した場合であっても、それがほとんど視界に入っていなければ記憶していると答える可能性は低い。

## 3 記憶可能性モデルの導出

第2章で述べたモデルのうち、静的シーンの記憶可能性である  $P_{memorability}$  は Isora らの提案手法を用いて計算することができる。そこで、 $P_{seen}$  を一人称視点映像を基に推定することを考える。

本稿では、被験者に  $P_{memorability}$  が一定であると見なせる環境下で一人称視点映像を撮影しながら特定ルートを散策するタスクを課し、その後アンケートによって特定オブジェクトについての記憶の有無を確認することで、一人称視点映像と記憶可能性との関係について調査した。

## 3.1 実験概要

被験者に、一人称視点カメラを装着した状態で大学構内の指定したルートを通り 15 分程度の散策を行うタスクを課し、その後アンケートによって特定オブジェクトについての記憶の有無を確認した。ルートは被験者が普段訪れない構内にあるものを指定した。アンケートの内容は、特定オブジェクトが撮影された画像を一枚ずつ提示しその一枚ずつに対して記憶しているか否かを回答するものとし、提示画像としてルート上から視認できる 47 のオブジェクトとルート上からは見ることのできない 47 のオブジェクトの画像を用意した。アンケートに用いたオブジェクトの画像の一例を図 1(a) に例示する。撮影終了後、一人称視点映像から特定オブジェクトの出現回数、停留時間、映像中の位置の情報を取り出し、アンケート結果との関係を調査した。ここで、映像中の位置として、一人称視点映像の上下左右からそれぞれ 1/4 の

The statistical estimation of scene memorabilities from first person view  
†Kento OIZUMI ‡Atsushi NAKAZAWA ‡Toyoaki NISHIDA  
†Undergraduate school of Informatics and Mathematical Science, Faculty of Engineering, Kyoto University.  
‡Graduate school of Informatics, Kyoto University.

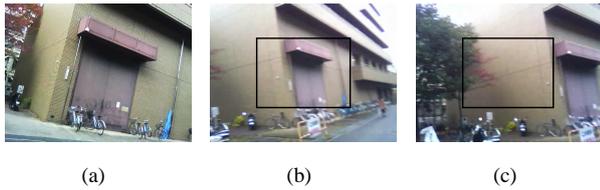


図 1: (a) アンケートに用いたオブジェクトの画像  
(b),(c) 一人称視点映像のフレーム画像

範囲を除いた部分（以下、中心部分）と、その他の部分の2種類のみを考える。一人称視点映像のフレーム画像を、図1(b),1(c)に示す。図1(b)は対象オブジェクトが一人称視点映像の中心部分に現れた際のものであり、図1(c)はその他の部分に現れた際のものである。

### 3.2 実験結果

図2にアンケート結果をプロットした図を示す。x軸は出現回数,y軸は停留時間であり,oはアンケートで記憶していると回答したもの、xは記憶していないと回答したものである。この図から、出現回数が多いほど、また停留時間が長いほど記憶可能性が高いことが分かる。そこで、このデータに対してSVMを用いた2クラス分類を行った。得られた境界面を図2に点線で示す。(b)のデータについて、このSVMを用いることで正解率70.2%、適合率73.9%、再現率68.0%の精度で分類を行うことができた。

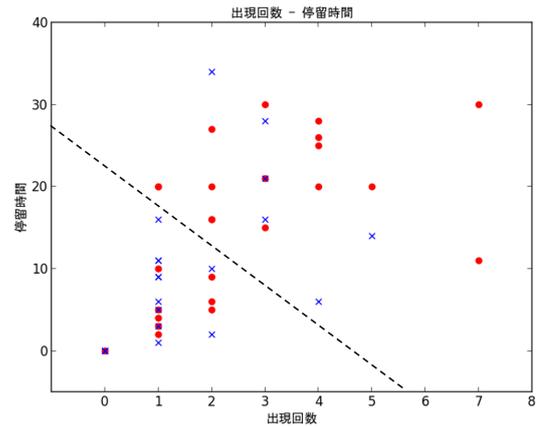
### 3.3 考察

図2より、一人称視点映像における出現回数が多く、停留時間の長いオブジェクトの記憶可能性が高いことが推測される。

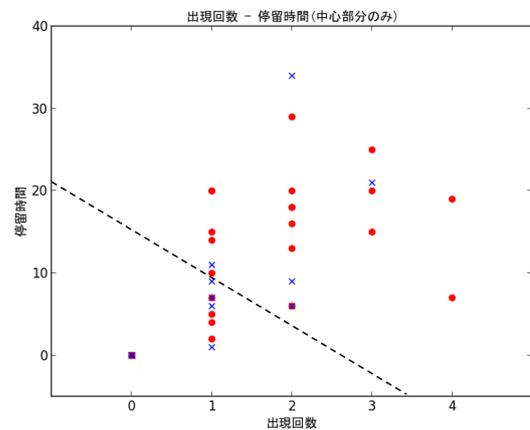
## 4 結論

一人称視点映像から対象オブジェクトの記憶可能性を推定するモデルを構築した。一人称視点映像におけるオブジェクトの出現回数、停留時間、映像中の位置の三要素を用いて、正解率70.2%、適合率73.9%、再現率68.0%の精度で撮影者の記憶の有無を推定できた。今後は、今回考慮していない要素も推定に用いることで、精度を向上させることを検討している。

謝辞本研究は科学技術振興機構さきがけ、文部科学省科学研究費助成事業研究課題番号26270058、コニカミノルタ科学技術振興財団、科学技術振興機構・研究成果展開事業「センター・オブ・イノベーション(COI)プログラム」の支援を受けた。



(a) 一人称視点映像全体



(b) 一人称視点映像の中心部分

図2: 出現回数、停留時間と記憶可能性の関係。図中のoは正しく記憶されたもの、xは記憶されなかったものを示す。(a) 一人称視点映像全体から出現回数、停留時間を計算した結果、(b) 一人称視点映像の中心部分から出現回数、停留時間を計算した結果

## 参考文献

[1] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *Computer Vision and Pattern Recognition (CVPR)*, pp. 145–152. IEEE, 2011.