2G1-5

一人称視点映像による実環境の記憶可能性推定

The estimation of scene memorability using a first person view video

大泉 建人　　中澤　篤志　　西田　豊明
Kento OIZUMI　Atsushi NAKAZAWA　Toyoaki NISHIDA

*1京都大学大学院情報学研究科
Graduate scool of Informatics, Kyoto University.

Information of whether the person can well remember the objects in real environment such as buildings, signs and notices, are useful in city planning and marketing. We defined the stored ease of object in real environment as "environmental memorability". In recent years, attention is focused on a first-person video because the capacity of the recording medium is increased. First-person video data is very good as the data for recording the status of an individual and it is associated with easy with objects. Then we propose a method to estimate the environmental memorability with first-person video. As a result, we found that the more times and the longer an object appear, the higher environmental memorability is. Improving the process for estimating the head movement by using the optical flow or the difference in luminance value between frames is considered as a future issue.

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 [1][2]. 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.

実環境中のオブジェクトの記憶可能性は，その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人の割合と定義できる. 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. しかし，アンケートに基づく方法は大きなコストがかかるため，自動化する必要がある. そこで，本稿では人の前方向を撮影する一人称視点映像を基に記憶可能性を推定することを考える. 一人称視点映像を用いる理由として，まず近年のウェアラブルカメラの普及が挙げられる. また，一人称視点映像中から注視行動が検出可能であることが挙げられる. 人は目の前に興味を惹かれるものがあった場合，注視を行う. 興味を惹かれるものは記憶に残りやすいため，記憶可能性は注視行動に現れると考えられる. よって，一人称視点映像から注視行動を検出することで記憶可能性が推定できる.

本稿では，一人称視点映像を基にオブジェクトに対する個人の記憶可能性を推定するためのモデルを構築する. 集められた一人称視点映像それぞれに対してこのモデルを用いることでオブジェクトに対する撮影者個人の記憶可能性を推定し，個人の記憶可能性を統合することでオブジェクトの記憶可能性を推定することを目指す.

2. 関連研究

本章では記憶や一人称視点映像に関する関連研究を挙げ, 本研究の位置づけを明確にする.

記憶に関する関連研究として，Isola ら [3] による memorability の研究が挙げられる. Isola らは画像の覚えやすさ（memorability ）を記録した画像データベースを構築し，記憶のしやすさに影響を与える画像特徴を分析している. その結果を基に，画像記述子を元に予測機を学習することで memorability の予測を行なうことが可能である事を示した.

一方本研究では，画像の記憶しやすさではなく，実環境中のオブジェクトに対する記憶しやすさを扱う. これは, 単なる画像特徴のみならず, 頭部運動に基づくオブジェクトを見た回数等の個人の状況を表すパラメータにも影響を受けていると考えられる.

人の行動を観察するために一人称視点映像を用いる技術に注目が集まっている. Berry ら [4] は, 記憶に障害を抱えた人にウェアラブルカメラで一人称視点映像を記録させ, 介護者と共に閲覧することによって出来事を覚え，想起できる度合いに良い効果があったと述べている. また, 山田ら [5] は，視覚的顕著性マップモデルを用いて一人称視点映像に対して視覚的注意を推定する手法を提案している. 本研究においてもこのような一人称視点映像の特性を個人の状況を記録するデータとして用いる.

3. 記憶可能性モデル

実環境中のオブジェクトの記憶可能性は，興味を惹かれて長時間眺めたり，気になって振り向いたりすると高くなると考えられる. また，見る対象となるオブジェクト自身の持つ特徴も影響する. その他にも，記憶可能性はオブジェクトを見ているときの人の状態にも影響される.

以上より，オブジェクト O_i の記憶可能性 $P_{remember}$ は，フレーム数 T の一人称視点映像 $\mathbf{I} = (I_1, \cdots, I_T)$，撮影者の状態 $\mathbf{S} = (S_1, \cdots, S_T)$ を用いて以下のように表されると仮定する.

$$\underbrace{P_{remember}(O_i, \mathbf{I}, \mathbf{S})}_{\text{記憶可能性}} = \underbrace{P_{personality}}_{\text{個人差}} \cdot \underbrace{P_{memorability}(O_i)}_{O_i \text{ の記憶されやすさ}} \cdot \underbrace{P_{seen}(O_i, \mathbf{I})}_{I \text{ からわかる情報}} \cdot \underbrace{P_{status}(\mathbf{S})}_{\text{撮影者の状態}} \quad (1)$$

ここで，$P_{personality}$ は個人差を表す項である. また，$P_{memorability}(O_i)$ は対象オブジェクト O_i の持つ記憶されやすさを表す項である. これは Isola らによって提案されてい

連絡先: 大泉 建人，京都大学大学院情報学研究科知能情報学専攻，oizumi@ii.ist.i.kyoto-u.ac.jp

る静的シーンでの記憶可能性である memorability を表す [3]. $P_{seen}(O_i, \mathbf{I})$ は一人称視点映像 $\mathbf{I}$ に関する項である．この項に関連する要素として，$\mathbf{I}$ における O_i の出現回数，停留時間，出現位置や $\mathbf{I}$ のオプティカルフロー等が挙げられる．P_{seen} の各要素について，第 3.1 節に述べる．$P_{status}(\mathbf{S})$ は時間によって変化する，撮影者の状態を表す項である．$\mathbf{S}$ には撮影者の興味や集中等，$\mathbf{I}$ からは推測し難い要素が含まれるが，現状では人の状態を測定することは状態の複雑さから困難であるため，本稿では常に P_{status} は一定であると仮定している．

3.1 一人称視点映像から得られる情報

$P_{seen}(O_i, \mathbf{I})$ は一人称視点映像から得られる情報に関連する項である．本稿では以下の要素が関係すると考えるが，ここに示す以外にも関連する要素が存在する可能性はある．

出現回数，停留時間

よく見たオブジェクトほど記憶に残りやすいと考えられるため，一人称視点映像におけるオブジェクトの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くなると考えられる．

画像中の出現位置

人の視野は大きく中心視と周辺視に分けられる．中心視とは視線方向の中心に位置する部分であり，周辺視と比較して高解像度の映像を受容することができる [6]．中心視で捉えたものは周辺視でとらえたものに比べて詳細に観察できていると考えられるため，覚えられやすいと推測できる．よって一人称視点映像の中心位置が注視点と重なると仮定するとき，撮影された画像の中心に出現したオブジェクトの記憶可能性が高いと考えられる．

オプティカルフロー

オプティカルフローは 1 フレームの間に対象点が撮影画像内でどの程度移動しているかを表す．一人称視点映像ではカメラは頭部の動きに合わせて移動するため，撮影対象が運動を行わない場合，オプティカルフローには頭部運動が現れる．オプティカルフローの絶対値が小さいときは頭部が静止していると判断でき，頭部静止中は一点を注視していると考えられる．従って，オプティカルフローの絶対値が小さいときに一人称視点映像内に現れたオブジェクトは，記憶可能性が高いことが考えられる．

また，撮影者が一定の速さで一定の方向に正面を向いて歩いているとき，オプティカルフローは放射状に広がる．このオプティカルフローの拡大中心を FOE（Focus of Expansion）という．FOE は進行方向の無限遠点となる．FOE は視覚的注意を強く引くことが確かめられているため，記憶可能性を推定する際に有用な要素となると考えられる．

3.2 提案モデルの検証

このモデルを実験を通して検証することを考える．被験者が未知であるシーンに対しては $P_{memorability}$ は一定であるとみなすことができる．また，$P_{remember}$，$P_{personality}$ はアンケートを用いることで分かる．従って，3. より，P_{seen} について以下の式が成り立つ．

$$P_{seen} = \frac{\overbrace{P_{remember}}^{\text{アンケートから分かる}}}{\underbrace{P_{personality}}_{\text{アンケートから分かる}} \cdot \underbrace{P_{memorability}}_{const} \cdot \underbrace{P_{status}}_{const}} = f_{(\text{出現回数, 停留時間, 画像中の出現位置, オプティカルフロー})}$$

すなわち，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 P_{seen} との関係が得られ，モデルの妥当性を検証できる．

4. 実験

5 人の被験者に，一人称視点映像を撮影するためのウェアラブルカメラ (Looxcie LX2，図 1) 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．指定したルートを図 2 に示す．ルートは大学

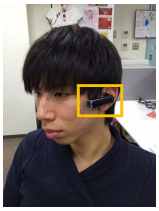

図 1: ウェアラブルカメラ

図 2: 実験ルート

構内の被験者が普段訪れない場所を指定した．アンケートは，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オブジェクト（以下，視認可能オブジェクト群と呼ぶ）の画像 47 枚とルート上からは視認することのできないオブジェクト（以下，視認不可能オブジェクト群と呼ぶ）の画像 47 枚を用意した．アンケートに用いたオブジェクトの画像の一例を図 3(a) に示す．

(a) (b) (c)

図 3: (a) アンケートに用いたオブジェクトの画像
(b)，(c) 一人称視点映像の画像．(b) では対象オブジェクトが中心部分に，(c) では周辺部分に出現

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，停留時間の情報を取り出し，アンケート結果との関係を調査した．

4.1 アンケートによる記憶可能性調査

視認可能オブジェクト群の画像に対して，一人称視点映像に 1 回以上現れたオブジェクトの内記憶していると回答した割合 (true positive rate)，視認不可能オブジェクト群の画像に対して記憶していると回答した割合（false positive rate）を表 1 に示す．この表から，false positive rate の平均値と比較して true positive rate の平均値が明らかに高くなっていることがわかる．また，被験者 C，D については true positive rate がチャンスレベルである 50.0%を下回っている．被験者 E については false positive rate が他の被験者と比較して高い，すなわち実際には見ていないにも関わらず見たと回答した割合が

表 1: アンケート結果

	true positive rate	false positive rate
被験者 A	24/42(57.1%)	5/47(10.6%)
被験者 B	31/44(70.5%)	5/47(10.6%)
被験者 C	**10/28(35.7%)**	2/47(4.3%)
被験者 D	**18/37(48.6%)**	5/47(10.6%)
被験者 E	23/39(59.0%)	**14/47(29.8%)**
平均	54.2%	13.2%

高くなっている．以上より，被験者 C，D，E についてはアンケート結果の信頼性が低いと判断できる．この結果，回答の信頼性の高い被験者 A，B のデータを解析に用いる．

4.2 一人称視点映像の解析

$P_{memorability}$ が一定であると仮定すると，アンケート結果は P_{seen} の項のみに依存していると考えられる．P_{seen} を推定するために一人称視点映像からどのような情報を取り出し，特徴量として用いることが適当であるかを実験によって得られた結果より検討する．

(a) 出現回数・停留時間

図 4，5 にアンケート結果と出現回数，停留時間との関係を示す．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．点のプロットは，全 47 個の視認可能オブジェクト群について行った．

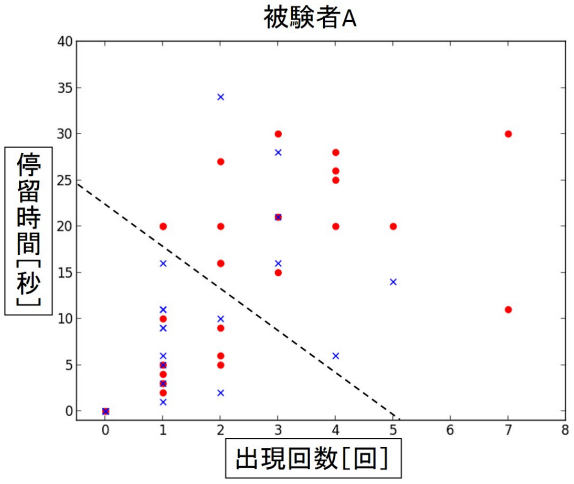

図 4: 被験者 A．x 軸は出現回数，y 軸は停留時間であり，o はアンケートで記憶していると回答したもの，x は記憶していないと回答したもの．

図 4，5 から，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．これはすなわち，何度も見たもの，長く見たものが覚えられやすいことを表しており，我々の仮説と一致する結果である．

(b) オブジェクトの出現位置

被験者 A のアンケートについて，一人称視点映像全体に対する出現回数，停留時間を基に行った推定と比較して，中心部分への出現回数，停留時間を基に行った推定の方がより高い精度で推定を行うことができた．しかし，他の被験者からはそのような結果は得られなかった．この理由として，中心視の範囲と今回用いた一人称視点映像の中心部分が一致しなかったことが考えられる．実際，頭部方向と視線方向は必ずしも一致しないため，注視点は一人称視点映像の中心位置とは一致しない．より正確に注視対象のオブジェクトを判断するためには，注視点推定を用いたアプローチをとることが考えられる．

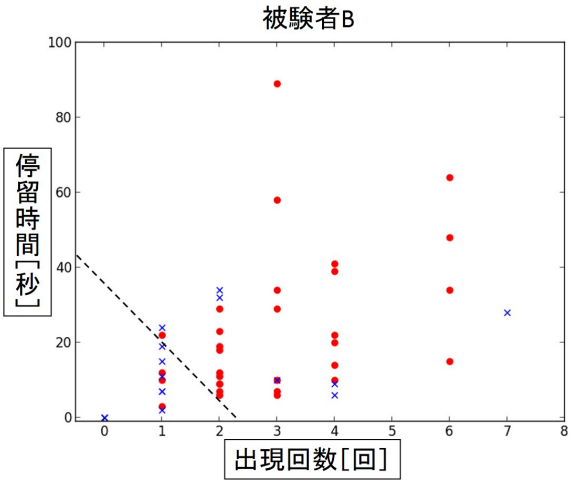

図 5: 被験者 B．x 軸は出現回数，y 軸は停留時間であり，o はアンケートで記憶していると回答したもの，x は記憶していないと回答したもの．

(c) オプティカルフロー

頭部の運動を抽出するための指標として，実験により得られた一人称視点映像からオプティカルフローを計算した．また，x 軸を一人称視点映像開始からのフレーム数，y 軸をオプティカルフローの絶対値の平均としたグラフに，そのフレームに現れていたオブジェクトを記憶していた割合を重ねて描画したものの一部を図 6 に示す．

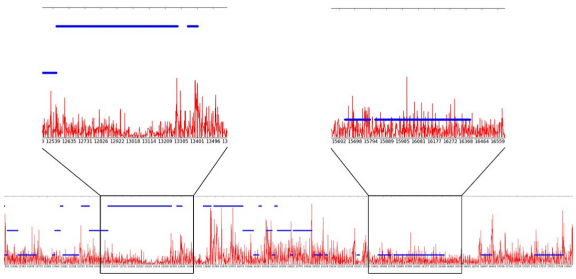

図 6: オプティカルフローの絶対値の時間変動 (赤線) と記憶可能性（青線）．
x 軸は一人称視点映像の開始からのフレーム数

第 3.1 節ではオプティカルフローの絶対値が小さい場合には頭部運動が少なく，すなわちあるオブジェクトに注視しておりそのオブジェクトの記憶可能性が高いと推定されると考察した．しかし，図 6 を見るとオプティカルフローの絶対値が小さい場合であっても記憶可能性が高い部分と低い部分があることがわかる．

そこで，次にオプティカルフローの絶対値の変動から，オブジェクトを注視するために行った頭部運動による変動のみを抽出することを考える．ローパスフィルタをかけた絶対値の時間変化の中で極小値をとる点で興味を持ったオブジェクトの注視を行ったと考え，その時点で一人称視点映像に映っているオブ

ジェクトの記憶可能性が高くなるという仮説を立てた．カットオフ周波数を 1Hz としたローパスフィルタを用いて上記の処理を行い，x 軸を停留時間，y 軸をオブジェクトが一人称視点映像に現れている状態で頭部運動を伴う注視が起こった回数としてアンケート結果をプロットしたグラフを図 7 に示す．ロー

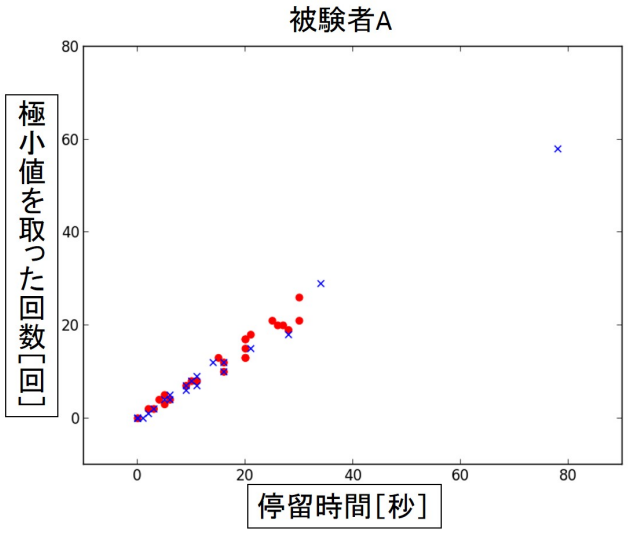

図 7: 停留時間と頭部運動を伴う注視が起こった回数の関係

パスフィルタをかけたグラフの極小値をとる時点と一人称視点映像を見比べたとき，我々が注視を行ったと判断した時点は概ね検出された．しかし，図 7 を見ると，対象オブジェクトの停留時間と注視が起こった回数はどちらも概ね比例している．これは単位時間当たりの極小値をとる回数がほぼ一定であることを示している．この原因として，ローパスフィルタをかけることで除くことができるとしていた，注視のために行う運動以外の情報，特に注視を行うための頭部運動に比べて周期の短い運動の影響が考えられる．この影響により，取り出したい点以外の多くの時点で極小値をとり，結果として停留時間に比例した回数の注視が起こったという解析結果が得られたと推測できる．実際，我々が注視を行っていないと判断するような点，例えば頭部をほとんど運動させずに足を踏み出した時点などが，極小値をとることが確認できた．

4.3 一人称視点映像の特徴とアンケート結果の関係

図 4 及び図 5 より，被験者 A，B に関して，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 4，5 に点線で示す．被験者 A の (b) 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．被験者 B の (a) については正解率 78.7%，適合率 80.0%，再現率 90.3%の精度で分類を行うことができた．

5. 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．本稿ではこの内 P_{seen} 項を一人称視点映像を基に推測することを試みた．P_{seen} 項を推測するため，オブジェクトの一人称視点映像における出現回数，停留時間，出現位置の他，一人称視点映像のオプティカルフローやフレーム間の輝度値の差分について，その扱いを検討した．その結果，対象オブジェクトの一人称視点映像における出現回数，停留時間は記憶可能性の推定に有用であることが示唆された．一方，オブジェクトの出現位置や，一人称視点映像のオプティカルフロー，フレーム間の輝度値の差分については，どのように記憶可能性推定に用いることが適切であるか結論付けることができなかった．出現回数や停留時間について，追実験を行い今回考察された記憶可能性との関係が普遍的なものである検討を行うことを今後の課題とする．また，同時に出現位置をどのように利用するか，オプティカルフローやフレーム間の輝度値の差分からどのように記憶可能性の推定に有用と思われる頭部運動を抽出するかについて，さらに検討を行う．今回考慮していない要因についても，記憶可能性との関係を調査することを今後の課題とする．

参考文献

[1] 松本創, 岩田伸一郎, 古賀利郎: 7180 中心市街地の景観画像における注視行動特性と色彩分布に関する研究 (夜間景観, 都市計画), 学術講演梗概集. F-1, 都市計画, 建築経済・住宅問題, Vol. 2009, pp. 429–430 (2009).

[2] 谷岡誠一, 槇田史郎: 洪水ハザードマップの認知と理解の向上を目指して, 平成 16 年度河川情報シンポジウム講演集, pp. 40–47 (2004).

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

[4] Berry, E., Kapur, N., Williams, L., Hodges, S., Watson, P., Smyth, G., Srinivasan, J., Smith, R., Wilson, B. and Wood, K.: The use of a wearable camera, SenseCam, as a pictorial diary to improve autobiographical memory in a patient with limbic encephalitis: A preliminary report, *Neuropsychological Rehabilitation*, Vol. 17, No. 4-5, pp. 582–601 (2007).

[5] 山田健太郎, 菅野裕介, 岡部孝弘, 佐藤洋一, 杉本晃宏, 開一夫: 一人称視点における視覚的顕著性マップモデルの性能評価, 電子情報通信学会技術研究報告. HIP, ヒューマン情報処理, Vol. 110, No. 422, pp. 81–86 (2011).

[6] Turvey, M. T.: On peripheral and central processes in vision: inferences from an information-processing analysis of masking with patterned stimuli., *Psychological review*, Vol. 80, No. 1, p. 1 (1973).