

複数人の視点映像を利用した図書館内の体験共有支援

和田 聖矢^{1,a)} 松村 耕平^{1,b)} 角 康之^{1,c)}

概要: 本研究では、図書館での体験共有を促すために、ウェアラブルカメラにより撮影された、複数人の一人称視点映像を利用した図書館内の体験共有支援システムを提案する。本システムは、画像の特徴点を用い、一人称視点映像に写り込んだ本棚の判定を行う。写り込んだ本棚別に、一人称視点映像を合成することにより、本棚ごとの複数人体験映像を生成する。6名に、図書館での一人称視点映像を撮影してもらい、その映像を用いシステムの動作実験を行った。その結果、複数人の身体が写り込んだ映像、本棚の前で本を読んでいる様子が写り込んだ映像、本棚を見渡すような映像が生成された。

キーワード: 体験共有, 一人称映像, ライフログ, 可視化

Assist of Sharing the Experiences in Library using Multiple Person's Vision

Abstract: In this paper, I propose the system that assists of sharing the experiences in library using multiple person's view image taken by wearable camera. The system judges bookshelf by using feature points of image. Moreover, the system visualizes action that like choose book and read book by synthesize multiple person's view images. I develop the system, and I check the system behavior.

Keywords: Sharing the Experiences, Visualization, First Person View, LifeLog

1. はじめに

近年ウェアラブルカメラの性能が向上し、高画質な映像を大量に保存することが可能となった。Google Glass^{*1} といったデバイスも注目されており、将来、日常的に一人称視点映像の撮影を行う人が増えると予想される。日常的にウェアラブルカメラでの撮影を行うことで、意識的に撮影できないような突発的な事象や、撮影しようとも思わなかった些細な事象が映像として記録される。こうした映像は、体験の共有に利用することができる。

本研究では、図書館利用者の一人称視点映像に着目した。

図書館利用者は、本を探して歩きまわる、本を手取る、本を読むといった体験をする。こうした体験の多くは、他人と共有されることがない。図書館利用者の体験を記録

し、共有することができれば、以下の様な気づきが得られるのではないかと考える。

- 複数人に利用されている本棚はどこか。
- あまり利用されていない本棚はどこか。
- どんな本が手に取られているのか。

そこで、本研究では、図書館利用者の一人称視点映像を利用し、体験の共有を支援するシステムを提案する。本システムは、各本棚が一人称視点映像に写り込んだ頻度を表すヒートマップと、複数人の行動を合成したコラージュ映像の作成を行う。ウェアラブルカメラで一人称視点映像を撮影しながら、図書館内を移動すると、映像に本棚が映り込む。本システムは一人称視点映像に写り込んだ本棚を判定し、どの本棚に長時間立ち寄ったか、どの本を手にとったか、という情報を可視化する。一人称視点映像に写り込んだ本棚を判定するために、Speeded Up Robust Features(SURF)[1]を用いた、本棚判定プログラムを作成する。また、Photosynthを利用し、同じ本棚が写り込んでいると判定された一人称視点映像を合成することにより、複数人の行動が合成されたコラージュ映像を作成する。

¹ 公立はこだて未来大学
Future University Hakodate

a) s-wada@sumilab.org

b) matsumur@acm.org

c) sumi@acm.org

*1 Google Glass. <http://www.google.com/glass/>.

2. 関連研究

2.1 体験映像の活用

近藤らは、個人視点映像を体験の振り返りに活用する方法を検討している [2]。近藤らは、映像メディアの問題点として、一覧性が悪く、映像すべてを閲覧するには長時間を要する、という点を挙げている。この問題に対して、画像処理により大量の視点映像から重要なシーンを検出し、GPS の位置情報を元に、検出したシーンの位置関係を付加するといったアプローチをとっている。また、視点映像は視野が狭く、動きが激しいため、閲覧するのが不快であるという問題を挙げている。この問題に対しては、複数の視点映像を重ね合わせ、パノラマ映像とすることにより解決を試みている。他にも、体験映像を活用する際に、映像から重要シーンを切り抜くアプローチとして、様々な研究が行われている [3][4][5]。

本研究でも同じように、個人視点映像の活用を検討する。活用方法として映像を複数人で共有することにより、体験の共有を支援する。一覧性が悪いという映像メディアの問題点に対して、本研究は、映像に写り込んだ本棚を判定し、シーンの切り分けを行う。また、複数人の映像を合成することにより、複数人の行動が写り込んだコラージュ映像を作成する。

2.2 映像を利用した体験共有

映像を体験共有に利用するシステムとして、田中らが開発した PhotoWalker がある [6]。PhotoWalker は、写真の表示切り替えにアニメーションを使い、写真間を移動する感覚を体験できるシステムである。例えば、個人が旅行中に撮影した写真を、このシステムを用い共有することにより、旅行者の行動を追体験することができる。

本研究では複数人の体験を共有するアプローチとして、複数人の体験が混ざり合った映像を生成する。

2.3 代替現実システム

鈴木らは、過去に撮影された映像を、HMD などの没入感の高いデバイスを通じて提示する、代替現実システム (SR システム) を提案している [7]。本研究で提案するシステムが生成するコラージュ映像を SR システムで提示することにより、一度に複数人の視点に乗り移る、といった体験が得られるのではないかと考える。

3. 複数人の一人称視点映像を利用した図書館内の体験共有支援システム

本研究では複数人の一人称視点映像を利用した、図書館内の体験共有支援システムを提案する。図 1 は提案するシステムの大まかな流れを表している。

はじめに、図書館利用者の複数人の一人称視点映像を記



図 1 システムの流れ

録する。図書館での体験を一人称視点映像として記録すると、映像には利用した本棚や手にとった本が映り込む。

次に、一人称視点映像に写り込んだ本棚が図書館内のどの本棚であるかを、予め用意した各本棚の映像と比較することにより判定する。比較する際には SURF を利用する。

次に、本棚ごとに判定された一人称視点映像の数を可視化したヒートマップを作成する。さらに、各本棚での複数人の体験映像を合成することにより、本に手をのぼす等、複数人の本棚前での行動を可視化したコラージュ映像を作成する。

3.1 一人称視点映像の記録

ウェアラブルカメラを利用し、図書館利用者の一人称視点映像を撮影する。一人称視点映像には以下のシーンが映り込むことが考えられる。

- 歩きまわる
- 本棚を見渡す。
- 本に手を伸ばす。
- 本を手にする。
- 本をめくる。
- 本を戻す。

これらのシーンを映像に捉えるためには、一秒間に一枚以

上の間隔で映像を撮影することが望ましい。撮影間隔を狭めると、その分細かく行動を捉えることができる。しかし、後述する処理を行う際、処理時間が膨大になってしまう。その為、本システムでは、一秒間に一枚の間隔で撮影された一人称視点映像を利用する。

3.2 一人称映像に写り込んだ本棚の判定



図 2 本棚のリファレンス映像の例

本棚の利用頻度の可視化、各本棚での行動を可視化したコラージュ映像の作成を行うために、一人称視点映像に写り込んだ本棚を判定する。本棚を判定するために、ビジュアルマーカーを現実世界に設置する、という方法が考えられる。ビジュアルマーカーとは、AR 技術において、現実世界の物体や位置を検出するために使用するもので、単純図形や模様から成る [8]。ビジュアルマーカーを用いた AR 技術として、有名なものに、ARToolKit*2 がある。ARToolKit の技術を用い、異なる模様のビジュアルマーカーを各本棚に設置することで、それを目印とし、視点映像に写り込んだ本棚を区別することができると思われる。しかし、ARToolKit を用いビジュアルマーカーを認識するためには、ビジュアルマーカーの全体が映像に映り込む必要がある。本棚の広範囲を認識するためには、ひとつの本棚に対し複数のビジュアルマーカーを配置する必要がある。

一方で、図書館の本棚には、以下の特徴がある

- 本棚によって収納されている本が異なる。
- 収納されている本の位置は大幅には変わらない。

*2 ARToolKit Home Page. <http://www.hitl.washington.edu/artoolkit/>

● 本棚は単純な直方体の形であることが多い。
これら特徴から、あらかじめ本棚を撮影した映像を、ビジュアルマーカーのように利用できるのではないかと考えた。

そこで、本研究では、本棚のリファレンス映像を作成した。図 2 は本棚のリファレンス映像の例である。各本棚に対してリファレンス映像を作成する。

本棚のリファレンス映像を利用し一人称視点映像に写り込んだ本棚を判定する。判定するために、SURF [1] 特徴点の比較を行う。SURF 特徴点は、回転、スケール変化に対して不変な特徴点であり、ノイズ、照明変化に強固なマッチングに利用することができる。図 3 は、一人称視点映像と本棚リファレンス映像の間で対応する SURF 特徴点を検出している様子である。対応する特徴点が緑の線で結ばれている。一人称視点映像に手が写り込んでいても、複数の対応点が見出されていることが分かる。



図 3 一人称視点映像と本棚リファレンス映像の間で対応する SURF 特徴点の検出

本棚のリファレンス映像と一人称視点映像の SURF 特徴点を比較し、一人称視点映像に写り込んだ本棚を判定する本棚判定プログラムを実装する。本棚判定プログラムは、以下の順で処理を行う。

- (1) 入力された一人称視点映像と各本棚のリファレンス映像を読み込む。
- (2) 一人称視点映像と本棚のリファレンス映像の SURF 特徴点を検出する。
- (3) 一人称視点映像の特徴点と各本棚のリファレンス映像の特徴点を比較し、対応する特徴点の数を計算する。
- (4) 一人称視点映像と、対応する特徴点が多い本棚のリファレンス映像を探し、その本棚のリファレンス映像に写っている本棚が一人称視点映像に写り込んでいる本棚であると判定する。

SURF の検出、対応点の検出には OpenCV*3 を利用した。

各一人称映像に対して、本棚の判定を行い、図 1 の様に本棚のリファレンス映像に一人称視点映像をマッピングしていく。

*3 OpenCV . <http://opencv.org/> .

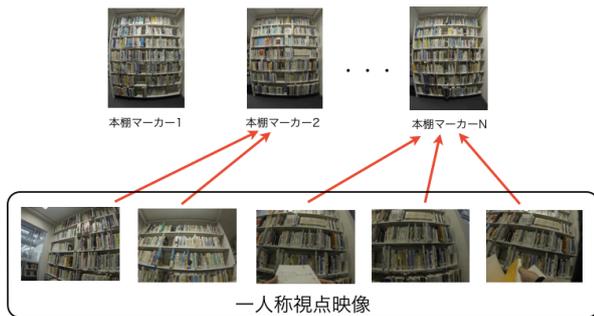


図 4 一人称視点映像の本棚への対応づけ

3.3 ヒートマップの作成

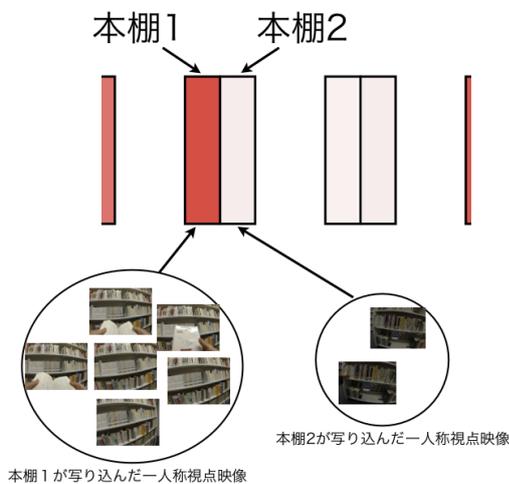


図 5 ヒートマップの作成

各本棚が写り込んでいる一人称視点映像の数を可視化した本棚のヒートマップを作成する。各本棚と、その本棚が写り込んでいると判定された一人称視点映像を紐付ける。紐付けられた一人称視点映像の数により、本棚の色を変化させる。図5の長方形は本棚を表す。図5のように、紐付けられた一人称視点映像の数が多いほど、赤色を濃く表示する。このヒートマップにより、どこの本棚に一人称視点映像が多く紐付けられているか知ることができる。

3.4 複数人の一人称視点映像の合成

本システムでは、Photosynth^{*4} を利用し、複数人の一人称視点映像を合成した映像を作成する。Photosynth とは、写真を合成し、三次元空間を構築するサービスである。同じ場所を様々な位置角度から撮影した複数の映像を元に三次元空間を作成することができる。風景のパノラマ映像や、建物の全景映像を作成する際に利用される。

本研究では本棚ごとに、各本棚が写り込んだ一人称視点映像を Photosynth を利用して合成し、複数人の行動を可

^{*4} Photosynth - Capture your world in 3D.
<http://photosynth.net/>.

視化したコラージュ映像を作成する。一般的に、映像を合成する際、通行人等の動的な物体が写っている映像を使用することは避けられる。しかし、本研究では、本を手にする、本を読むといった行動が写った映像を積極的に合成し、複数人の行動を可視化する。

3.5 コラージュ映像を提示するインターフェース



図 6 本棚ごとに生成されたコラージュ映像の提示

図 6 は、コラージュ映像を提示するためのインターフェースである。ヒートマップとして表示された各本棚をクリックすることにより、クリックした本棚のコラージュ映像が表示される。

4. システムの実装と動作実験

提案システムを実装し、動作実験を行った。図書館は、公立はこだて未来大学情報ライブラリー（以下情報ライブラリー）を利用した。

4.1 一人称視点映像の収集



図 7 ウェアラブルカメラの装着

システムの動作実験を行うために、複数人の一人称視点映像を収集した。合計 6 名、一人あたり、5 分から 7 分の間に毎秒 1 枚の間隔で撮影を行い、合計 2,230 枚の一人称

視点映像を収集した。映像の撮影には GoPro hero 3 を利用した。カメラは、図 7 のように胸部に装着した。撮影した画像の大きさは、横 2560 ピクセル、縦 1920 ピクセルであった。本棚の判定を行う際の処理時間を短縮するために、画像の大きさを横 640 ピクセル、縦 480 ピクセルに縮小した。

4.2 本棚のリファレンス映像の作成

情報ライブラリーの一部の本棚、合計 27 箇所の本棚のリファレンス映像を作成した。各本棚は B1~B27 の名前が付けられていた。本棚のリファレンス映像は一人称視点映像を収集する約一週間前に作成された。撮影には GoPro hero 3 を利用した。本棚の全体が映り込むように撮影を行った。撮影した画像の大きさは、横 2560 ピクセル、縦 1920 ピクセルであった。本棚の判定を行う際の処理時間を短縮するために、画像の大きさを横 1280 ピクセル、縦 960 ピクセルに縮小した。

4.3 システムの動作結果

撮影した一人称視点映像、本棚のリファレンス映像を用い、システムの動作確認を行った。

表 1 各本棚の一人称視点映像

本棚	判定された一人称視点映像数	正解数	正解率
B1	208	167	80.3 %
B2	98	70	71.4 %
B3	104	86	82.7 %
B4	21	3	14.3 %
B5	325	313	96.3 %
B6	182	163	89.6 %
B7	30	11	36.7 %
B8	35	12	34.3 %
B9	42	25	59.5 %
B10	35	1	2.9 %
B11	55	33	60.0 %
B12	32	2	6.3 %
B13	26	23	88.5 %
B14	23	0	0 %
B15	35	2	5.7 %
B16	29	1	3.4 %
B17	102	84	82.4 %
B18	42	24	57.1 %
B19	38	13	34.2 %
B20	72	53	73.6 %
B21	45	23	51.1 %
B22	48	40	83.3 %
B23	79	56	70.9 %
B24	272	226	83.1 %
B25	166	156	94.0 %
B26	63	56	88.9 %
B27	23	0	0 %

図 8 に作成されたヒートマップと対応する情報ライブラリー内の位置を示した。システムによる本棚の判定が正しく行われているか確認するため、各本棚に紐付けられた一人称視点映像を目視により確認した。

また、表 1 に各本棚に紐付けられた一人称視点映像の数、本棚と正しく紐付けされた一人称視点映像数（正解数）、正しく紐付けされた確率（正解率）を示した。収集した一人称視点映像 2,230 枚中、本棚と正しく紐付けされた映像は 1,643 枚、正解率は 73.7%であった。

4.4 考察

本棚の判定に関して、一人称視点映像に大きく手や本が写り込んでいても、正しく本棚を判定することができた。また、本棚のリファレンス映像は一人称視点映像を収集する約一週間前に作成されたものであったが、多くの一人称視点映像から本棚を正しく判定することができた。多少であれば、本の出入りがあっても本棚の判定を行うことが可能であると考えられる。

一方で、手にとった本が大きく写り込んだ一人称視点映像、ぼやけた一人称視点映像等、本棚を正しく判定することができない映像があった。また、通路を歩いている映像や、前かがみになったため地面のみが写り込んだ映像等、本棚が写り込んでいない映像があった。これらの映像に関しては、本棚が写り込んでいないと判定する必要がある。

作成されたコラージュ映像に関して、以下の特徴が見られた。

複数人の身体が写り込み

図 9 のように複数人の身体が写り込んでいる映像があった。これらの一人称視点映像は撮影された時間、撮影者がそれぞれ異なっていた。複数の独立した一人称視点映像を見るよりも効率よく、複数人の行動を確認することができた。

視点の多様性の可視化

図 10 のように、複数人の一人称視点映像が合成されることにより、空間の広がりが見られた。本棚内で、注目されている箇所、注目されていない箇所を知ることができると考えられる。

時間経過の多様性の可視化

図 11 のように、個別の本を読んでいる様子が写り込んだ一人称視点映像が、複数枚合成されることにより、個別の本を読み続けているという時間情報が可視化された。

5. 課題と今後の展望

5.1 本棚の判定精度向上

作成したシステムの動作実験を行った結果、本棚の判定に対し、以下の問題があった。

(1) 図 12 の様に、本棚が写り込んでいない映像であって

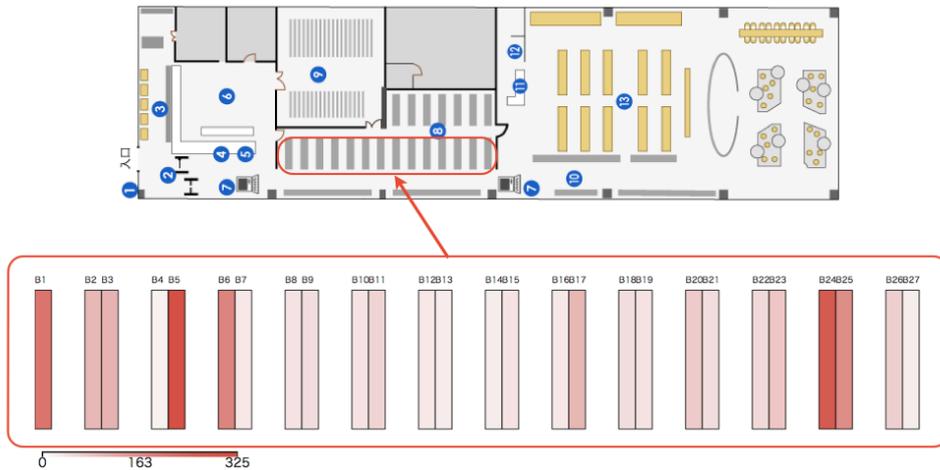


図 8 作成されたヒートマップと対応する情報ライブラリー内の位置



図 9 複数人の身体の映り込み



図 12 本棚が写り込んでいない映像



図 13 左:ぼやけた映像, 右:手にとった本が大きく写り込んだ映像

も、本棚のリファレンス映像と対応する特徴点が多少検出される為、いずれかの本棚が写っていると判断されてしまう。

(2) 図 13 の様に、映像がぼやけている場合や手にとった本が大きく写り込んだ場合、本棚のリファレンス映像と同じ本棚が写り込んでいても、対応する特徴点を多く検出できない。

一人称視点映像に写り込んだ本棚の特定精度を高めるためには、これらの問題に対応する必要がある。

(1) に関して、本棚が写り込まない映像に対する処理として、一人称視点映像と本棚のリファレンス映像の間で、対応する特徴点が一定数以下の場合、本棚が写っていないと判定する、といった処理が考えられる。

(2) に関して、ぼやけた映像や本が大きく写り込んだ一人称視点映像に対しては、その映像の前後に撮影された一人称視点映像の情報を用い、前後の一人称視点映像に写り込んだ本棚が同じであった場合、間に撮影された映像にも同じ本棚が写っていると判断する、といった処理を行うことにより、本棚を正しく判定することができると考える。

これらの処理を実装し、本棚特定プログラムの精度を向上させることを考えている。

5.2 コラージュ映像を図書館内で提示するシステム

今後の展望として、HMD を用い、コラージュ映像を図書館内で、リアルタイムに提示するといったシステムが考えられる。例えば、SR システムの様に [7]、現実で眺めている本棚と同じ本棚が写り込んでいる過去に撮影された映像をリアルタイムで提示することにより、他人の視点に乗り移り、行動を追体験できると考える。こうしたシステムを実現するためには、現実で眺めている一人称視点映像を高速に判定する必要がある。



図 10 本棚を閲覧する様々な視点のコラージュ



図 11 個別の本をしばらく読み続けている様子のコラージュ

参考文献

- [1] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features." Computer Vision-ECCV 2006. Springer Berlin Heidelberg, 2006. 404-417.
- [2] 近藤一晃, 高瀬恵三郎, 小泉敬寛, 中村裕一, 森幹彦, and 喜多一. "個人視点映像を用いた気づき体験の回想と整理支援: フィールド調査における問題発見を通じて (一般セッション, PRMU のフロンティア・グランドチャレンジ)." 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解 110.330 2010. 13-18.
- [3] Aizawa, Kiyoharu, Kenichiro Ishijima, and Makoto Shina. "Summarizing wearable video." Image Processing, 2001.
- [4] Kawamura, Tatsuyuki, Yasuyuki Kono, and Masatsugu Kidode. "Wearable interfaces for a video diary: towards memory retrieval, exchange, and transportation." Wearable Computers, 2002.(ISWC 2002). Proceedings. Sixth International Symposium on. IEEE, 2002. Proceedings. 2001 International Conference on. Vol. 3. IEEE, 2001.
- [5] Nakamura, Yuichi, Jun'ya Ohde, and Yuichi Ohta. "Structuring personal activity records based on attention-analyzing videos from head mounted camera." Pattern Recognition, 2000. Proceedings. 15th International Conference on. Vol. 4. IEEE, 2000.
- [6] 田中浩也, 有川正俊, and 柴崎亮介. "写真画像群の重なりを用いた広域的な擬似3次元空間." インタラクティブシステムとソフトウェア IX 2001. 75-84
- [7] 鈴木啓介, 脇坂崇平, and 藤井直敬. "代替現実:「いま・ここ」を体験する VR システム." インタラクシオン 2011. 2011. 869-870.
- [8] 亀田能成. "技術解説." 映像情報メディア学会誌 66.1 2012. 45-51.