



# マルチユーザー VR 遠隔拡大視システムのための 超高速パンチルトカメラの適応的視線制御

田原大輔<sup>1)</sup>, 阿部勇太<sup>1)</sup>, 入山太嗣<sup>1)</sup>, 小室孝<sup>1)</sup>, 島崎航平<sup>2)</sup>, 石井抱<sup>2)</sup>,

Daisuke TAHARA, Yuta ABE, Taishi IRIYAMA, Takashi KOMURO, Kohei SHIMASAKI, and Idaku

ISHII

1) 埼玉大学 理工学研究科 (〒 338-8570 埼玉県さいたま市桜区下大久保 255)

2) 広島大学 先進理工系科学研究科 (〒 739-8527 広島県東広島市鏡山 1-4-1)

**概要:** 本研究では、遠隔地の映像を自由に拡大しながら視ることができる VR 遠隔拡大視システムを、複数ユーザーが利用するための適応的視線制御手法を提案する。本手法では、取得した広角映像を元に顕著性、オプティカルフローを計算し、視線情報を加えた重要度マップから、超高速パンチルトカメラのミラー角を制御することで、各ユーザーに提示される拡大映像のリフレッシュレートを最適化する。

**キーワード:** VR 遠隔拡大視, 超高速パンチルトカメラ, 適応的視線制御

## 1. はじめに

近年 VR デバイスの普及に伴い、遠隔地の映像を HMD を通して視る VR 遠隔視システム [1, 2] が、観光や工事現場など様々な分野で活用され始めている。このシステムは、ビデオカメラなどからリアルタイムで取得した映像を用いることで、従来の監視システムよりも高い没入感を得られるという利点がある。しかし、多くの VR 遠隔視システムは広角映像の表示が主であり、映像を拡大すると解像度が著しく低下するため、対象の詳細な情報を取得することは困難であるという課題があった。

そこでこの課題を解決するため、VR 遠隔視システムを用いつつ、遠隔地の視たい場所の望遠映像を任意のタイミングで取得できる VR 遠隔拡大視システム [3] が登場した。VR 遠隔拡大視システムでは、超高速パンチルトカメラ [4, 5] を用いることで、高速な視線制御により高解像度の望遠映像を効率的に取得できる。このようなシステムは、建設現場での合同点検や、遠隔での技術指導など、複数の専門家や作業者が同時に一つの対象を異なる視点から詳細に確認する場面で、その真価を発揮する。しかし、このシステムを複数のユーザーが同時に利用する場合、一台の超高速カメラが持つ有限の fps を全ユーザーで分け合うことになるため、各ユーザーに割り当てられる望遠映像のフレームレートが低下し、ユーザー体験が損なわれるという新たな問題が生じる。

本研究ではこの複数人利用における課題を解決するため、取得した広角映像から、静的注目度、動的注目度、そしてユーザーの視線方向を基に、ユーザーが興味を持つ可能性の高い領域を判断し、その領域のリフレッシュレートを適応的に制御する手法を提案する。これにより、システムのリソースを重要度の高い情報へ動的に集中させ、複数のユーザー

が同時にシステムを利用する場合でも、各々が見たい対象の詳細な情報を高い品質で取得することができる。

## 2. VR 遠隔拡大視システム

本章では、本研究の基盤となる VR 遠隔拡大視システムの概要と、特に複数人での利用時に生じる課題、そしてその課題を解決する適応的視線制御の必要性について述べる。

### 2.1 システム概要

VR 遠隔拡大視システムとは、従来の VR 遠隔視システムの課題であった、映像を拡大すると解像度が著しく低下し、対象の詳細な情報を得ることが難しいという問題を解決したシステムである。このシステムで用いる超高速パンチルトカメラの外観を図 1 に示す。また、このシステムを実際にユーザーが用いている様子を図 2, 図 3 に示す。本システムでは、超高速パンチルトカメラを用いて広角映像と高解像度の望遠映像を同時に取得する。HMD の動きとカメラのミラーを連動させることで、ユーザーは広角映像を見つつ、望遠映像を重畳表示して詳細な情報を取得する。

### 2.2 複数ユーザー利用時における課題

超高速パンチルトカメラが提供可能な総リフレッシュレート  $R_{max}$  は有限であり、複数ユーザーが同時に望遠映像を要求すると各ユーザーのフレームレートが低下し、モーションブレイクや入力遅延による没入感の劣化を招く。さらに、ユーザーごとに関心対象が空間的にも時間的にも異なるため、リフレッシュレートを画一的に等分配すると、静止領域を長時間観察しているユーザーへ過剰な帯域を割り、動的かつ高重要度の対象を見ているユーザーへ十分な帯域を供給できないという問題が生じる。

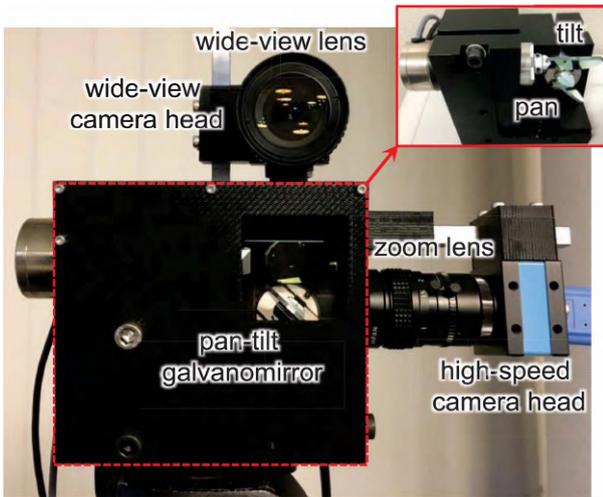


図 1: 超高速パンチルトカメラの外観



図 2: ユーザーがシステムを使用している様子

### 2.3 適応的視線制御手法の必要性

上述の課題を解決するためには、広角映像から得られる静的注目度と動的注目度、およびユーザ視線に基づくユーザの意図を統合的に評価し、重要度の高い領域へリフレッシュレートを優先的に割り当てる適応的視線制御が不可欠である。本研究では、時刻  $t$  における統合重要度マップ  $M_{imp}^t$  を生成し、各ユーザの関心領域に対する平均重要度を指標として総リソースを動的に配分する手法を提案する。これにより、動きの多い領域や複数ユーザが注視する領域には高フレームレートを維持し、静的で注目度の低い領域では更新頻度を抑制することで、限られたリソースを有効活用しながらユーザ体験の均質化を図る。

## 3. 適応的視線制御

限られた撮像リソースを画一的に配分すると、動きの無い映像を見ている場合でも過剰に映像を更新してしまったり、逆に動きのある映像を見ている場合でも更新が遅くなってしまったりする。そこで本研究では、限られた撮像リソースを効果的に配分するための手法である適応的視線制御を提案する。本手法では、映像の時空間的な価値を定量化する「統合重要度マップ」の生成と、そのマップに基づく「動的なリソース配分」によって、複数ユーザー環境下において

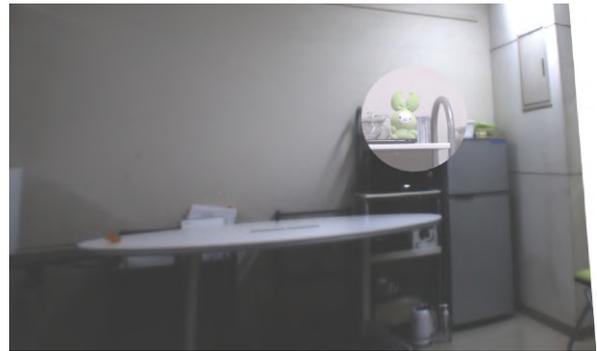


図 3: 広角映像上に望遠映像を重畳表示した様子

も、静的な対象への不要なリソース割り当てを削減し、動きのある重要な対象へリソースを集中させることが可能となる。

### 3.1 統合重要度マップの生成

統合重要度マップは、広角映像内のあらゆる領域の「注目すべき度合い」を数値で表現したものである。図 4 に、本研究で提案する統合重要度マップ生成手法の全体概要を示す。広角映像と視線情報を入力とし、静的注目度マップ ( $M_S^t$ )、動的注目度マップ ( $M_D^t$ ) と、ユーザーの意図を反映したヒートマップ ( $M_U^t$ ) をそれぞれ生成する。生成したマップは正規化され、重み付け ( $w_S, w_D, w_U$ ) されたのちに統合され、最終的な統合重要度マップ ( $M_{imp}^t$ ) が生成される。従来の重要度推定は、静的特徴や動的変化など単一指標への依存が原因で、場面やユーザ意図と乖離するという課題があった。本研究では、この問題を解決するために、「静的注目度」「動的注目度」「ユーザーの意図」の三つの情報源を、互いに補完する形で統合し、環境変化とユーザーの意図の双方を反映した重要度マップを生成する設計方針を採用した。これにより、状況に応じた柔軟な視線制御が可能となる。

- 静的注目度：入力フレームを多尺度の平滑画像と比較し、各画素の色差・輝度差・エッジ強度などを統合した局所コントラスト統合し、スコア化することで視覚的に際立つ領域を検出する。
- 動的注目度：映像内での時間的な変化がもたらす重要度。移動する人や車両など、多くの応用場面において動きがある対象は最優先で視るべき情報であり、本システムでは、この動きをリアルタイムに検出する。
- ユーザーの意図：システムが評価する客観的な重要度とは別に、ユーザー自身が何を見たいかという主観的な関心度。これはユーザーの視線情報から直接的に取得され [6, 7]、システムがユーザーの要求に応答するための重要な要素となる。また、残像減衰率を考慮することで、以前見ていた映像の重要度も反映される。

これらの異なる性質を持つ重要度は、それぞれ正規化された後、システムの目的や状況に応じて設定された重み係数を用いて一つのマップへと統合される。

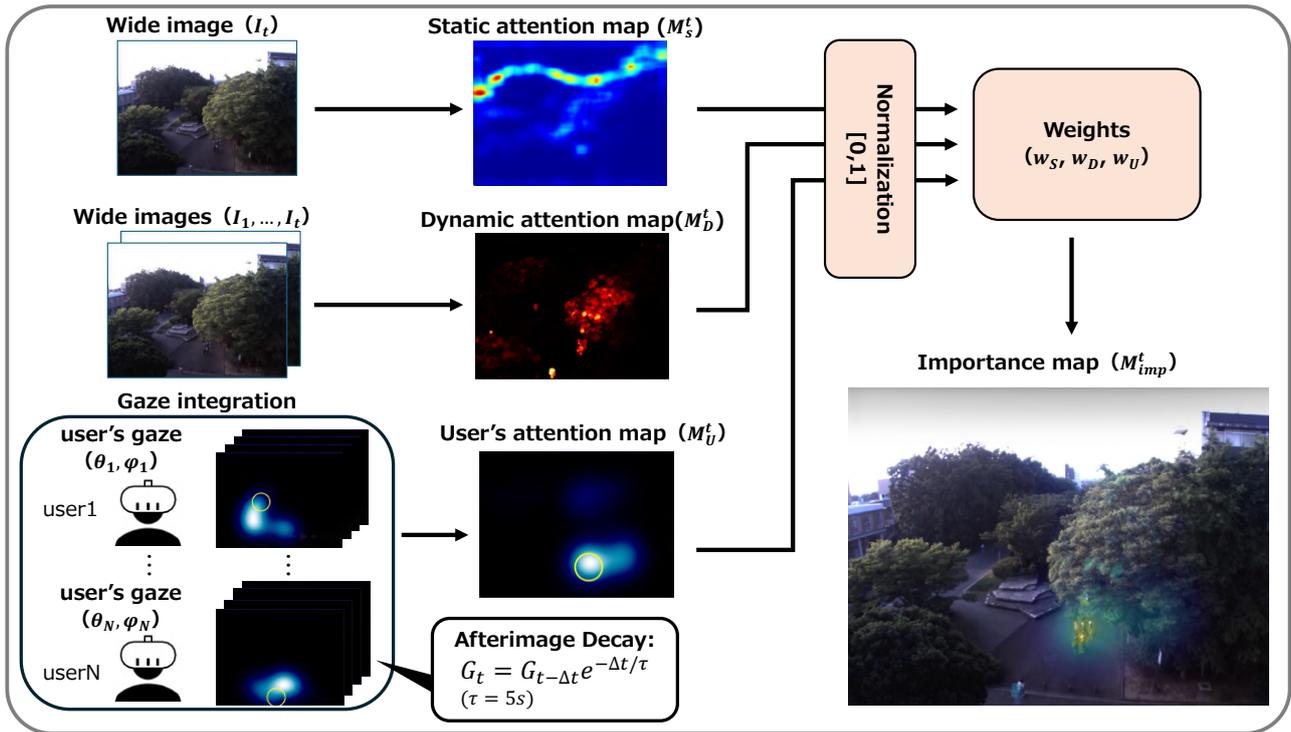


図 4: 統合重要度マップ生成手法の概要

### 3.2 動的なリソース配分

時刻  $t$  における各ユーザー  $u$  の関心領域  $ROI_u(t)$  を視線ベクトルから射影して取得し、その内部画素に対する統合重要度  $M_{imp}^t(x, y)$  の平均値

$$S_u(t) = \frac{1}{|ROI_u(t)|} \sum_{(x,y) \in ROI_u(t)} M_{imp}^t(x, y)$$

を評価スコアと定義する。全ユーザーのスコアを正規化して重み

$$w_u(t) = \frac{S_u(t)}{\sum_v S_v(t)}$$

を求め、超高速パンチルトカメラが提供可能な最大リフレッシュレート  $R_{max}$  から

$$R_u(t) = R_{min} + w_u(t)(R_{max} - NR_{min})$$

を算出して各ユーザーに割り当てる ( $R_{min}$  は最低保証レート、 $N$  はユーザー数)。これにより、静的・動的特徴と視線意図を統合した重要度が高い ROI ほど高フレームレートで撮像され、低い ROI には帯域を抑えた撮像を適応的に適用できる。

## 4. 統合重要度マップの実装

本研究で提案する適応的視線制御を備えた VR 遠隔拡大視システムは、ユーザーが HMD を装着し、その頭部回転情報から視線を推定することを想定している。しかし、本稿では、提案する適応的視線制御アルゴリズムの基礎的な有効性を検証することに焦点を当てるため、ユーザーの視線入力はマウスカーソルで代替して実験を行った。

本実装では、静的注目度を顕著性マップ [8, 9]、動的注目度をオプティカルフロー [10, 11]、ユーザーの意図の反映を、

各ユーザーのマウスカーソルの座標情報をヒートマップとして描画することで、統合重要度マップを実装した。図 5 に、実際に生成したマップとその様子を示す。図 5a は入力となる広角映像である。システムは、この映像からボトムアップの注目度である、顕著性マップとオプティカルフローマップを並行して生成する。図 5b は、顕著性マップを示す。このマップは、映像内で静的だが際立っている領域を検出する。ここでは、木と空の境界部分の重要度が高く算出されていることがわかる。これにより、動きがなくとも視覚的に目立つ対象を捉えることが可能になる。図 5c は、オプティカルフローを可視化したマップを示す。ここでは、広角映像の下部分にいる人の動きが検出されており、その部分の重要度が高く算出されていることがわかる。図 5d は、ユーザーの視線履歴を反映したヒートマップを示す。ここでは、マウスカーソルの履歴をガウシアンで描画したヒートマップが表示されており、現在の注目度が高い領域がわかる。図 5e は、図 5b、図 5c、図 5d のマップを正規化したのち、重み付けして生成された統合重要度マップを示す。静的、動的注目度に加え、ユーザーが意図的に注目している領域を考慮したマップであり、システムはこのマップを用いて適応的に視線制御を行う。

## 5. 結論

本稿では、複数ユーザーが利用する VR 遠隔拡大視システムにおいて、画一的なリソース配分に起因する体験品質の低下という課題に取り組んだ。提案した適応的視線制御手法は、静的注目度、動的注目度、そしてユーザー意図を統合した重要度マップに基づき、限られた撮像リソースを

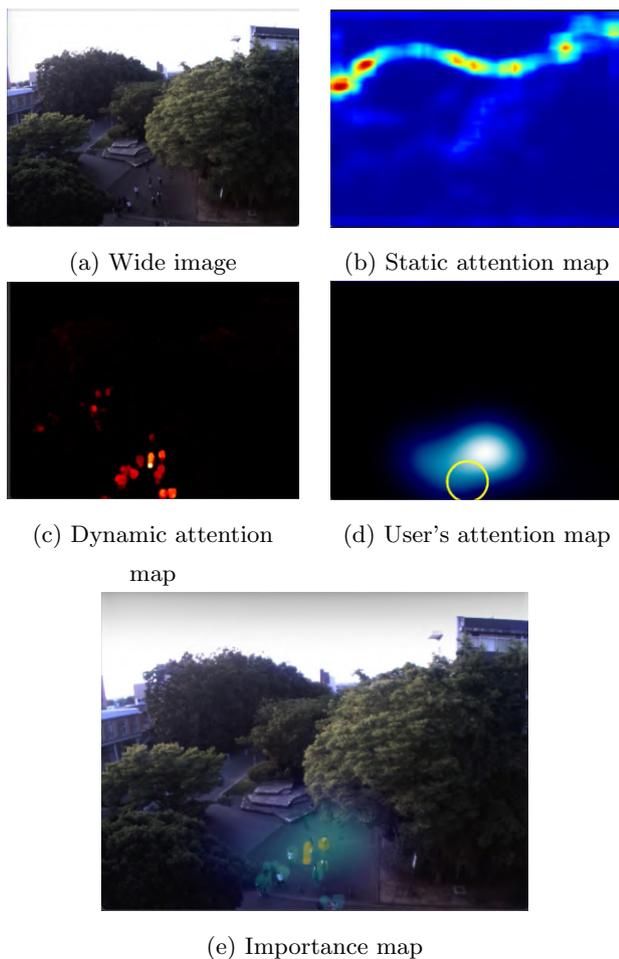


図 5: 統合重要度マップの実装例

動的に配分するものである。生成した統合重要度マップで高重要度領域を適切に抽出できることを確認した。

今後は、第一に、統合重要度マップに基づくリフレッシュレートの動的割当アルゴリズムを実装し、第二に、超高速パンチルトカメラを用いた VR 遠隔拡大視システムへ完全に適用する。さらに、各要素の重みをタスクや環境に応じて自動最適化する手法を検討し、実環境下での長時間評価を通じてユーザー体験の向上度合いを定量的に検証する。これらの取組みにより、限られた撮像リソースを最大限に活用し、高品質な遠隔拡大視体験を提供する実用的なシステムの構築を目指す。

#### 参考文献

- [1] Yi Ren and Henry Fuchs. Faster feedback for remote scene viewing with pan-tilt stereo camera. In *Proceedings of 2016 IEEE Virtual Reality (VR)*, pp. 273–274. IEEE, 2016.
- [2] Tamay Aykut, Stefan Lochbrunner, Mojtaba Karimi, Burak Cizmeci, and Eckehard Steinbach. A stereoscopic vision system with delay compensation for 360 remote reality. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*, pp. 201–209, 2017.

- [3] 田原大輔, 阿部勇太, 入山太嗣, 小室孝, 島崎航平, 石井抱. 超高速ミラー駆動パンチルトカメラを用いた vr 遠隔拡大視システム. *インタラクション 2024 論文集*, pp. 1073–1075, 2024.
- [4] Shaopeng Hu, Kohei Shimasaki, Mingjun Jiang, Taku Senoo, and Idaku Ishii. A simultaneous multi-object zooming system using an ultrafast pan-tilt camera. *IEEE Sensors Journal*, Vol. 21, No. 7, pp. 9436–9448, 2021.
- [5] Shaopeng Hu, Hongyu Dong, Kohei Shimasaki, Mingjun Jiang, Taku Senoo, and Idaku Ishii. Omni-directional panoramic video system with frame-by-frame ultrafast viewpoint control. *IEEE Robotics and Automation Letters*, Vol. 7, No. 2, pp. 4086–4093, 2022.
- [6] Ehsan Amini, Saleh Javadi, and Siamak Khatibi. Saliency map generation based on human level performance. In *2024 IEEE Gaming, Entertainment, and Media Conference (GEM)*, pp. 1–5. IEEE, 2024.
- [7] Shibo Wang, Shusen Yang, Hailiang Li, Xiaodan Zhang, Chen Zhou, Chenren Xu, Feng Qian, Nanbin Wang, and Zongben Xu. Salientvr: Saliency-driven mobile 360-degree video streaming with gaze information. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, pp. 542–555, 2022.
- [8] Akshay Jindal, Krzysztof Wolski, Karol Myszkowski, and Rafał K Mantiuk. Perceptual model for adaptive local shading and refresh rate. *ACM Transactions on Graphics (TOG)*, Vol. 40, No. 6, pp. 1–18, 2021.
- [9] Xiaodi Hou and Liqing Zhang. Saliency detection: A spectral residual approach. In *2007 IEEE Conference on computer vision and pattern recognition*, pp. 1–8. Ieee, 2007.
- [10] Gunnar Farneback. Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis*, pp. 363–370. Springer, 2003.
- [11] Haofei Xu, Jing Zhang, Jianfei Cai, Hamid Rezatofghi, and Dacheng Tao. Gmflow: Learning optical flow via global matching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8121–8130, 2022.