



THE VIRTUAL REALITY SOCIETY OF JAPAN

## 視線情報に基づき主観的現実の自動操作を行うシステムの検証

## Verification of Gaze-Based Automatic Reality Manipulation

伊藤大智<sup>1)</sup>, 高原慧一<sup>1)</sup>, 坂本凜<sup>1)</sup>, 泉原厚史<sup>2)</sup>, 脇坂崇平<sup>3)</sup>, 檜山敦<sup>1,3)</sup>, 稲見昌彦<sup>1,3)</sup>

Daichi ITO, Keiichi TAKAHARA, Rin SAKAMOTO, Atsushi IZUMIHARA, Sohei WAKISAKA, Atsushi HIYAMA, and Masahiko INAMI

1) 東京大学 情報理工学系研究科 (〒 113-8656 東京都文京区本郷 7-3-1, ito@star.rcast.u-tokyo.ac.jp)

2) 東京大学 工学系研究科 (〒 113-8656 東京都文京区本郷 7-3-1)

3) 東京大学 先端科学技術研究センター (〒 153-8904 東京都目黒区駒場 4-6-1)

**概要:** 体験者周辺のリアルタイム映像と録画映像とを、視線情報と映像の画像処理による情報に基づき、時間的・空間的に自動で混ぜ合わせる VR システムを提案する。注視点情報（何を見ている/いないか）を体験シナリオに即座に反映させることで、体験者にとっての主観的現実を、本人に気づかれることなく操作することができる。本発表では変化盲等を考慮した具体的な体験シナリオをいくつか実装し、その検証を心理実験手法により行った。

**キーワード:** 代替現実, アイトラッキング, 変化盲

## 1. 背景

ヒトは現実を認識する際、外部からの五感の入力を自身で解釈した結果を現実として認識する。この自分にとって現実だと思えるもののことを、ここでは主観的現実と呼ぶことにする。この主観的現実とは、ヒトの知覚特性によって生じる錯覚等を用いることで、入力された身の周りの物理世界とは異なる内容にすることが可能である。

過去の研究 [1] では、リアルタイム映像と録画映像とを代わる代わる体験者に提示しつつ、映像を切り替えた瞬間には体験者が切り替えに気づかないようにすることで、録画映像を主観的には眼前の現実であるかのように体験する代替現実 (substitutional reality, SR) システムが提案されている。しかしこのシステムでは、周囲の物体や人が突然消える等の映像切り替えによる時間的・空間的に不連続な変化が体験者の注意を向けている視界の中で起こり、体験者が映像切り替えに気づいてしまう恐れがある。実験者はそれを避けるため、体験者の行動を観察しどこに注意を向けているかを推測しながら、映像切り替えのタイミングを時には自身で判断しなければならなかった。

そこで本研究では、視覚的な連続性を保つことで、体験者に目で見ている世界が現実だと思わせたまま、身の周りの物理世界から視覚情報を変化・乖離させる VR システムの構築を考える。具体的には、ヘッドマウントディスプレイ (HMD) とアイトラッカーを組み合わせ、体験者の視線情報を利用して映像を自動的に切り替える手法を提案する。この 2 つのデバイスを組み合わせた既存システムとして、視線位置情報を VR 空間での物体選択に用いたもの [2] や、周辺視野の解像度を落とすことでレンダリング時間を縮減す

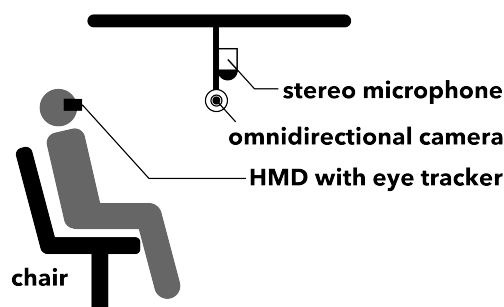


図 1: 提案システムによる体験の様子。体験者はアイトラッカー付きの HMD を装着し、周囲の環境を撮影するための全天周カメラと環境音を録音するためのステレオマイクが体験者の目の前に固定される。

る foveated rendering を実現したもの [3] 等がある。

提案システムでは体験者の視線位置情報を映像に対するリアルタイムの画像処理と組み合わせることで、注視領域では視覚情報が大きく変化しないようにしつつ映像を自動で時間的・空間的に混ぜ合わせる。これにより提示する体験の自動制御が可能になるだけでなく、実験者の判断による映像制御では難しかった体験をも提示できる。

また提案システムでは、変化盲 (change blindness, CB) や不注意盲 (inattention blindness, IB) といったヒトが視覚的な変化や不連続性を見落としてしまう現象 [4] に着目し、これらを映像切り替え時に引き起こすことを狙う。CB は、サッカード中の抑制やまばたき等の一時的な視覚情報の欠如が起きているときに、その最中に生じた変化を見落としてしまう現象を指す。一方 IB は、ある出来事やタスク

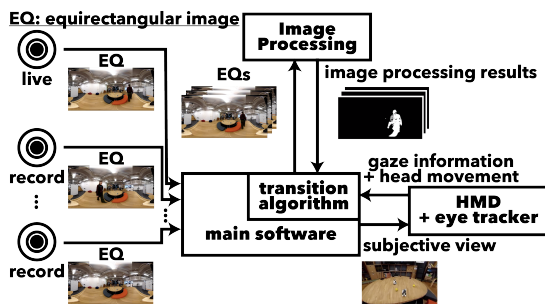


図 2: 映像を時間的・空間的に混ぜ合わせるシステム構成の全体図。

にヒトの注意が向けられているとき、それらとは関係ない領域や文脈で生じた変化を見落としてしまう現象のことである。過去の研究にも CG 空間で IB が生じることを確かめたもの [5] や、映像のレンダリング時間縮減に IB を活用したもの [6]、体験者に気づかれない redirected walking の実現のために CB を活用したもの [7] 等がある。

## 2. 提案システムの詳細

### 2.1 システム構成

提案システムによる体験の様子を図 1 に示す。全天周カメラ (Insta360 Air) を体験者の目の前に固定し、体験者の周囲の環境をリアルタイムに撮影する。この際映像は画像サイズが  $3008 \times 1504$  ピクセルの dual-fisheye 形式で計算機に逐次送信されるが、計算機ではこれを同サイズの equirectangular 形式画像に変換する。それと同時に、ステレオマイク (ZOOM H2n) も体験者の目の前に固定し、体験者の周囲の環境音を 2 チャンネルの信号として録音する。一方全く同一の場所に固定した同一のカメラ・マイクで、予め周囲の様子を録画・録音しておく。そして TouchDesigner, Unity, Python 等で構成したソフトウェアを計算機 (CPU: Intel Core i9-9900K, GPU: NVIDIA GeForce RTX 2080) で動かし、equirectangular 形式のリアルタイム映像と録画映像とを適宜混ぜ合わせ、HMD (HTC Vive Pro) に提示する。この際 equirectangular 形式の画像を一旦球に貼り付けてから、体験者の頭部の向きに応じた部分を切り出して提示するようにする。また HMD 内部にはカメラベースのアイトラッカー (Pupil Labs) を取り付けており、体験者の視線位置情報等を計算機に逐次送信する。なお体験者が HMD を着用した際に、視線位置のキャリブレーションを行う。ここまで述べた映像を混ぜ合わせるシステム構成の全体を図 2 に示す。さらに映像提示と同時に、その映像に同期した環境音を体験者に提示する。先行研究で聴覚情報に着目した SR システム [8] が提案されており、提案システムでも提示映像の内容に矛盾しない立体音響を提示することで、より体験者が提示映像を現実と思いやすくなると期待される。

### アルゴリズム 1 映像切り替え判定アルゴリズム

**Input:**  $p, q$ : 切り替え前後の映像フレーム,  $G$ : 視点周辺の一定範囲に含まれる画素集合,  $v$ : 頭部の回転速度,  $t$ : threshold

**Output:** 切り替えるか否かを表すブール値

```
function ISSAFEFORSWITCHINGIMAGES( $p, q, G, v, t$ )
   $r = \text{CALCULATERISK}(p, q, G, v)$   ▷  $r$ : transition risk
  if  $r \leq t$  then
    return true
  else
    return false
  end if
end function
```

**Input:**  $p(n), q(n)$ : 時刻 (開始からの経過時間)  $n$  での切り替え前後の映像フレーム,  $G(n)$ : 時刻  $n$  での視点周辺の一定範囲に含まれる画素集合,  $v(n)$ : 時刻  $n$  での頭部の回転速度,  $t$ : threshold,  $c$ : 事前に定めた一定の時間幅

```
function MANAGESWITCHING( $p(n), q(n), G(n), v(n), t, c$ )
  ▷ この関数は毎フレーム呼び出される

  Initialize:  $time = 0, s' = 0$ 
   $s =$  現在時刻
  if ISSAFEFORSWITCHINGIMAGES( $p(s), q(s), G(s), v(s), t$ )
  == false then
     $time = 0$ 
  else
     $time = (s - s') + time$ 
  end if
   $s' = s$ 
  if  $time \geq c$  then
     $time = 0$ 
    SWITCHIMAGES()  ▷ 映像を切り替える
  end if
end function
```

### 2.2 映像切り替え判定アルゴリズムの詳細

ソフトウェアで映像を混ぜ合わせる際、あるタイミングで映像を切り替えるか否かの判定はアルゴリズム 1 によって行われる。ここで重要なのは、transition risk と threshold の 2 つの値である。transition risk は、切り替え前後の映像フレームの画像処理による情報と、アイトラッカーで取得した視線位置情報によって計算される、映像切り替えが体験者に気づかれてしまう危険度を表す値である。一方で threshold は、閾値を表す事前に設定した定数である。transition risk が threshold を一定時間下回り続けたときに、映像の切り替えを行う。

transition risk の計算手法 (アルゴリズム 1 内の CALCULATERISK 関数) として、本研究では 2 つの手法を用意した。

1 つ目は各映像の背景差分法による動体の情報を元にした手法である (図 3 も参照されたい)。まず各フレーム画像の解像度を、計算量を減らすために事前に定めた解像度まで落とす。次に各画像のどの部分が動いているのか、混合正規分布 (MoG) に基づき背景画像を適宜更新していく背景差分法を用いて求める。この結果得られた画像の各画素

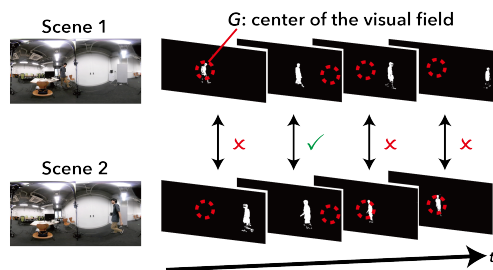


図 3: 各映像の背景差分法に基づく映像切り替え判定アルゴリズム。体験者が見ている領域における動きの量を表す値 (transition risk) を計算し、映像を切り替えるか否かの判定に用いる。

は、ある領域において物体が動いているか否かを表す。ところで映像が不連続に変化し、体験者が注目している動く物体や人等が消えたり、物体や人が何も無い注視領域から突然現れたりすると、体験者が映像切り替えに即座に気づいてしまう。一方で体験者が切り替え前後の映像の両者においてあまり動きのない領域を見ているときに映像を切り替えれば、体験者は変化に気づかないと考えられる。さらに Suzuki ら [1] によると、体験者の頭部回転速度が速いほど、体験者はそのタイミングで生じた映像切り替えに気づきにくい (i.e., head movement suppression)。そこで得られた各画像のある画素  $i$  での輝度値を  $p_i, q_i$ 、視点周辺の一定範囲に含まれる画素集合を  $G$ 、HMD から得た頭部の回転速度を  $v$ 、重みを表す事前に定めた値を  $w_i, w'_i$  として、transition risk  $r$  を

$$r = \sqrt{\left(\sum_{i \in G} w_i p_i\right)^2 + \left(\sum_{j \in G} w'_j q_j\right)^2} - H(v) \quad (1)$$

という形で求める。ここで  $H(v)$  は  $v$  について単調増加する関数であり、これを減らすことで頭部回転速度が速いほど映像切り替えが起こりやすいように誘導している。

2 つ目は各映像間のフレーム画像の輝度値の差を元にした手法である (図 4 も参照されたい)。各フレーム画像の解像度を事前に定めた解像度まで落とし、各画像の輝度値画像を求める。この結果得られた画像においてある画素での輝度値の差は、ある領域における映像差分の大きさを表す。ところで体験者がいる物体や人等を見ているときに、映像が不連続に変化しそれが消えると、先と同様映像切り替えに気づいてしまう。一方で体験者が切り替え前後であまり変化のない領域を見ているときに映像を切り替えれば、体験者は変化に気づかないと考えられる。そこで各輝度値画像のある画素  $i$  での輝度値を  $p_i, q_i$ 、視点周辺の一定範囲に含まれる画素集合を  $G$ 、HMD から得た頭部の回転速度を  $v$ 、重みを表す事前に定めた値を  $w_i$  として、transition risk  $r$  を

$$r = \sum_{i \in G} w_i |p_i - q_i| - H(v) \quad (2)$$

という形で求める。ここで  $H(v)$  は先と同様、 $v$  について単調増加する関数である。

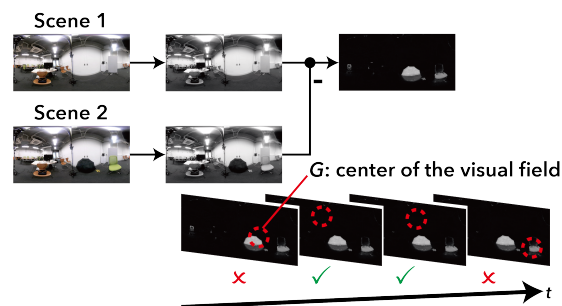


図 4: 各映像間の輝度値の差に基づく映像切り替え判定アルゴリズム。体験者が見ている領域における映像差分の大きさを表す値 (transition risk) を計算し、映像を切り替えるか否かの判定に用いる。

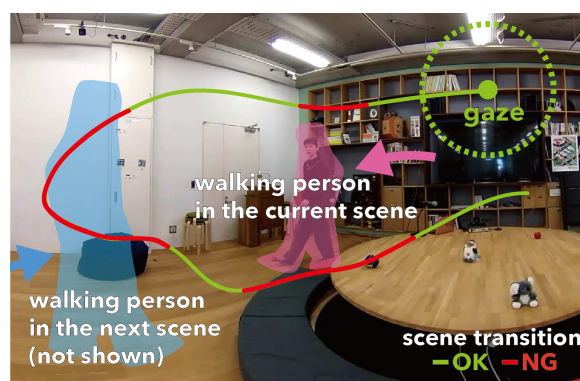


図 5: 実現可能な体験の一例。体験者が見ていない隙に、部屋の中で歩く人を消滅させたり出現させたりする。なおこの図は [9] より引用した。

またシーン切り替え所用時間 (どれくらいの時間をかけてシーンを切り替えるか) を頭部運動に合わせて動的に変更する。一般に徐々に提示映像が変化する場合、瞬時に変化する場合に比べて変化が知覚されにくいことが知られている [4]。本システムにおいてもこれを考慮し、基本的には数百ミリ秒以上かけてシーン切り替えを行う。ただし一定以上の速度で頭部が運動している場合は、瞬時にシーン切り替えを行う。これは以下の 2 つの理由による: (1) 前述した head movement suppression により、頭部運動時には瞬時の切り替えが知覚特性上適している。(2) 時間をかけて切り替えた場合には切り替え開始時と完了時で注視領域が大きくなり、そのために判定アルゴリズムが適切に機能しなくなるリスクが増大する。この手法の有効性については、本発表で心理実験手法を基に議論する予定である。

### 3. 提案システムで可能な体験シナリオ

提案システムを用いることで、事前に用意した体験シナリオや transition risk の計算手法に応じ、様々な体験を作り出すことが可能である。具体例を以下に 2 つ述べる。なおこれらの体験シナリオ自体は、筆者らが以前 [9] で述べたものと同様のものである。

- 体験者が見ていない隙に、部屋の中で歩く人を消滅さ

せたり出現させたりする(図5)。体験者は何かがおかしいという違和感を感じ、何が起きたのか後から理解することは可能だが、変化の瞬間にはその違和感の正体が分からない。これには各映像の背景差分法に基づく映像切り替え判定アルゴリズムを用いる。

- 周囲に転がっているボールの数を体験者に数えてもらう。ボールは体験者の気づかないタイミングで消えたり現れたりするため、体験者はボールを正確に数えることができない。これには各映像間の輝度値の差に基づく映像切り替え判定アルゴリズムを用いる。

どちらの例においても、体験者に見えている物体や人がまさに今存在しているものなのか、過去に存在していたものなのか、区別がつかなくなることに注意されたい。また映像切り替えを自動で行うため、切り替えタイミングを考慮した綿密な準備を事前に行っていない映像、例えば体験が始まった後に映像提示の裏で録画した映像も混ぜ合わせることが可能な点も指摘しておく。

#### 4. 結論と今後の展望

本研究では体験者に気づかれずに、体験者周辺のリアルタイム映像と録画映像とを時間的・空間的に自動で混ぜ合わせるVRシステムを提案した。体験者の視線位置情報と映像の画像処理による情報を組み合わせることで、体験者にとっての主観的現実の変容を引き起こすことが可能である。

今後の展望としてまず、体験者がその瞬間には気づかないような映像変化を起こす手法を、より詳細に検討することが挙げられる。提案システムにおいては映像切り替えにかかる時間の他にも、transition riskの計算に用いる画素集合 $G$ の大きさやthresholdの値等、調整可能なパラメータが複数存在する。これらを各々の体験者にとって最適なものに調整する手法を確立することは今後の課題とする。またtransition riskを計算するアルゴリズムにおいて、視線の移動速度や軌跡等の、アイトラッカーから取得可能な視線位置以外の情報も活用することを考えている。特に視線の移動速度に関しては、サッカード中に提示映像を微小回転させることでredirected walkingを実現する研究[10]のように、CBを引き起こしうるという観点において着目している。さらに視線を誘導する手法を用いることで、より積極的に主観的現実を操作することも考えられる。例えばHataら[11]は提示画像の一部に気づかれぬ程度のブラーをかけ、その部分に無意識に注意を誘導する手法を提案している。このような手段で映像の不連続な変化が生じない領域に視線を誘導し、大きな変化が生じる領域から積極的に注意を逸らすことで、映像変化に気づかれぬタイミングを待つだけに留まらず作り出すことも可能と考えられる。

謝辞 本研究はJST ERATO (JPMJER1701) 及びJSPS 科研費 (15K16014) の助成を受けたものである。

#### 参考文献

- [1] K. Suzuki, S. Wakisaka, and N. Fujii. 2012. Substitutional Reality System: A Novel Experimental

Platform for Experiencing Alternative Reality. *Scientific Reports*, 2, 459.

- [2] V. Tanriverdi and R. J. K. Jacob. 2000. Interacting with eye movements in virtual environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '00)*. Association for Computing Machinery, New York, NY, USA, 265-272.
- [3] A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans. Graph.* 35, 6, 179.
- [4] R. A. Rensink. 2009. Attention : Change Blindness and Inattentive Blindness.
- [5] J. Triesch, D. H. Ballard, M. M. Hayhoe, and B. T. Sullivan. 2003. What you see is what you need. *Journal of Vision*, 3, 1, 9.
- [6] K. Cater, A. Chalmers, and P. Ledda. 2002. Selective quality rendering by exploiting human inattentive blindness: looking but not seeing. In *Proceedings of the ACM symposium on Virtual reality software and technology (VRST '02)*. Association for Computing Machinery, New York, NY, USA, 17-24.
- [7] E. A. Suma, S. Clark, D. Krum, S. Finkelstein, M. Bolas, and Z. Warte. 2011. Leveraging change blindness for redirection in virtual environments. *2011 IEEE Virtual Reality Conference*, Singapore, 159-166.
- [8] K. Fan, H. Izumi, Y. Sugiura, K. Minamizawa, S. Wakisaka, M. Inami, N. Fujii, and S. Tachi. 2013. Reality jockey: lifting the barrier between alternate realities through audio and haptic feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2557-2566.
- [9] D. Ito, S. Wakisaka, A. Izumihara, T. Yamaguchi, A. Hiyama, and M. Inami. 2019. EyeHacker: Gaze-Based Automatic Reality Manipulation. In *ACM SIGGRAPH 2019 Emerging Technologies (SIGGRAPH '19)*.
- [10] Q. Sun, A. Patney, L. Y. Wei, O. Shapira, J. Lu, P. Asente, S. Zhu, M. McGuire, D. Luebke, and A. Kaufman. 2018. Towards virtual reality infinite walking: dynamic saccadic redirection. *ACM Trans. Graph.* 37, 4, 67.
- [11] H. Hata, H. Koike, and Y. Sato. 2016. Visual Guidance with Unnoticed Blur Effect. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '16)*. Association for Computing Machinery, New York, NY, USA, 28-35.