

# 多視点映像における視対象に対する興味と選択行動および注視行動の関係性分析

Analysis of relationships among user's interest in visual object, its selection behavior, and gaze behavior while watching multi-viewpoint video

速水慎太郎\*<sup>1</sup>  
Shintarou HAYAMIZU

平山高嗣\*<sup>2</sup>  
Takatsugu HIRAYAMA

Sidney Fels\*<sup>3</sup>

間瀬健二\*<sup>2</sup>  
Kenji MASE

\*<sup>1</sup>名古屋大学工学部

School of Engineering, Nagoya University

\*<sup>2</sup>名古屋大学大学院情報科学研究科

Graduate School of Information Science, Nagoya University

\*<sup>3</sup>The University of British Columbia

Multi-viewpoint video viewers make people see freely visual objects from various viewpoints. However, they need to expend effort on selecting the viewpoints according to their interest. As an assistive approach to reduce the effort, we have proposed the target-centered viewpoint switching that stabilizes a specific object at the center of display field while switching the viewpoints. In order to evaluate the usefulness, we analyzed relationships among user's interest in each object, selection behavior of the object using computer mouse, and gaze behavior. As a result, we found that the approach encouraged the users to direct the gaze to more interesting object at the center.

## 1. はじめに

近年ビデオカメラなどの映像装置の発展に伴い、映像コンテンツはより身近かつ多種多様なものとなっており、娯楽のためだけでなく、知識の増加といったことにも利用できる。ある一つのイベントの撮影を考える。異なる人が撮影を行い、それらの映像を集約する場合、1台のカメラで撮影した映像よりも様々な視点を選ぶことができ、より広く詳細な動画視聴ができる。このような複数のカメラを用いて撮影し、視点選択ができる映像を多視点映像といい、これまでも、多視点映像を使用したスポーツの試合の撮影 [1][2] が行われ、今後さらなる多視点映像の発展と普及が期待されている。

多視点映像がより一般的になり、視聴者自身が視点を選択できるようになると、一つの問題が考えられる。多視点映像では視点が複数存在するため、興味に沿った視点を選ぶことに集中してしまい、その結果、映像そのものを楽しめなくなってしまう可能性がある。この問題を解決するために、視聴にかかる労力を下げる必要がある。その一つとして、特定の被写対象を視聴者が選択することで対象を画面中央に常に表示し、見たい対象を見失うことなく視点を切り替えられる釘付け視聴方式を我々は提案し [3]、それに基づいた多視点映像視聴システム Peg Scope Viewer を開発している [4]。そして、釘付け視聴方式の評価を視聴者の注視行動に基づいて行っている [5]。評価方法は釘付け視聴を行うシステムと行わないシステムを視聴した際の視線を計測し、視聴者が選択した対象と注視点座標との距離を比較するものであった。結果として、釘付け視聴を行う方が注視点座標と対象との距離は短くなり、釘付け視聴方式は、視聴者の視線を物理的に特定の被写対象に集中させることができる。しかしながら、注視行動は視対象への興味の度合いに依存し [6]、また、視聴者は選択した対象に必ずしも興味を持っていないため、必ずしも各被写対象と注視点座標との距離が短くなるわけでない。そこで、本研究では、視聴者の被写対象に対する興味と対象の選択行動および注視行動との関係性を分析する。

## 2. 多視点映像視聴インタフェース

### 2.1 釘付け視聴方式

多視点映像とは複数の視点から撮影された映像であり、視点を選択することで、視聴者は多角的な映像視聴を行うことができる。ここで、複数の固定カメラを用いて撮影を行う場合の模式図を図 1(a) に示し、撮影した映像は図 1(b) とする。図中の円錐に注目していると考えたとき、視点切り替えを行うと、切り替え前と切り替え後で円錐の表示位置が移動するため、円錐に注目し続けることは容易でない。そこで図 1(c) のように円錐が画面中央に表示されるように画面を平行移動およびデジタルズームすることで、視点切り替えが行われたとしても注目対象を安定させて表示できる。

この方式では視聴者が対象に集中して注視することができると考えられ、このような提示方法を釘付け視聴方式と呼び [3]、釘付け視聴を行わない方式をカメラ中心視点切り替え (図 1(b))、釘付け視聴を行う方式をターゲット中心視点切り替えと呼ぶ (図 1(c)) [4][5]。

### 2.2 釘付け視聴インタフェース

釘付け視聴方式を導入した多視点映像インタフェースとして我々は Peg Scope Viewer (以降 PSV と表記) を開発した [4]。図 2 は PSV の GUI のスナップショットである。このインタフェースが備えている機能としてはターゲット中心視点切り替えとカメラ中心視点切り替えの表示方式の切り替え、注目する対象の選択があり、また、それら以外にも映像の視聴操作として再生、停止、カメラ選択、ズーム、再生速度の変更、シークバー操作による再生時刻の変更を備えている。カメラ中心視点切り替えの場合、選択した対象を画面中央へ釘付けしないが、その領域の基準点に赤い四角が図 2 のように表示され、その対象を追跡する。一方、ターゲット中心視点切り替えでは画面の中央に釘付け、赤い四角を表示する。

選択対象としては映像のコンテキストを表現するものを設定する。スポーツの映像を例に考えると選手やボールなどが挙げられ、これらの領域の画面上での座標を基準点とする。また、選択対象の基準点を検出する方法としては手動でアノテーションを行うかコンピュータビジョン技術である物体追跡技術を用いる。

連絡先: 速水慎太郎, 名古屋大学工学部電気電子・情報工学科,  
hayamizu@cmc.ss.is.nagoya-u.ac.jp

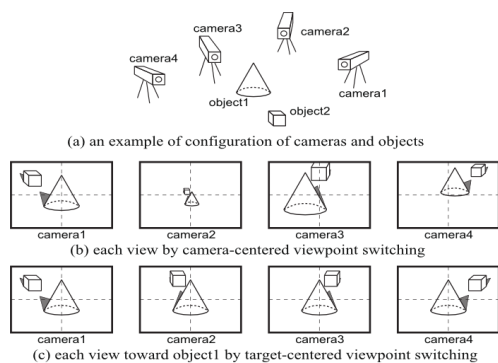


図 1: (a) 多視点映像の撮影と (b) 各視点における画像の例

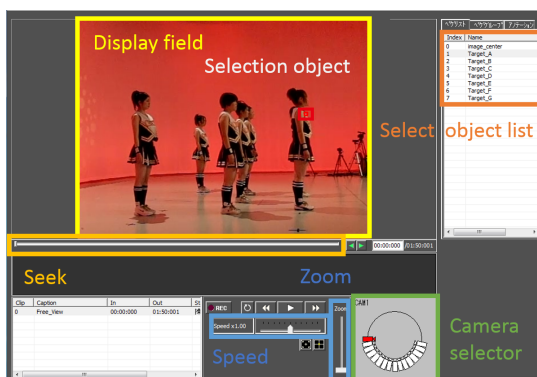


図 2: 多視点映像釘付け視聴システム (Peg Scope Viewer)

### 3. 多視点映像における視対象に対する興味と選択行動および注視対象の関係性の仮説

注視行動は視対象に対する興味と関係があり、静的対象および動的対象それぞれに対する関係が明らかになりつつある [7]. しかしながら、これらの知見は、基本的には単視点映像を能動もしくは受動視聴する状況において得られたものである。多視点映像の能動視聴では、視聴者が自由に視点を切り替えることができ、本研究では被写対象への興味を反映した注視行動が従来知見と同様であるか分析する。

まず、被写対象を十分に知覚するためには、それが網膜上の中心窩に近い領域に投影されるように眼球を動かす必要がある。つまり、選択対象と注視点座標間の距離に興味に影響すると考えられる。そこで、次の仮説を立てる。

- [I] 選択対象と注視点座標との距離は興味が高い対象のほうが興味が高い対象より短くなる。

次に、選択行動は興味と関係があるという知見も得られている。例えば TV 番組の視聴時間と興味との関係 [8] や、番組の視聴履歴を用いた興味推定 [9] が挙げられる。そこで多視点映像においても、対象を選択する行動が興味に依存すると考え、上記の知見における TV 番組の視聴時間は本研究では対象を選択している時間、視聴履歴は対象を選択した回数と対応があるとし、次の 2 つの仮説を立てる。

- [II] 対象を選択する時間は興味が高い対象のほうが興味が高い対象よりも長くなる。
- [III] 対象を選択する回数は興味が高い対象のほうが興味が高い対象よりも多くなる。

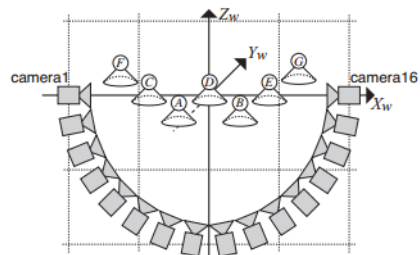


図 3: カメラ配置とチアリーディングの初期フォーメーション

本研究では、ターゲット中心視点切り替えの有用性を評価することも目的としている。従来研究において、カメラ中心視点切り替えよりもターゲット中心視点切り替えの方が、視聴者の注視点座標と選択対象との距離が短くなることが明らかになっている [5] が、選択対象に対する興味度合いを考慮した分析は行われていない。そこで、本研究では、上記の 3 つの仮説で着目した視聴行動と選択対象への興味の高低との関係が、ターゲット中心視点切り替えとカメラ中心視点切り替えでどのように変化するかを分析する。

## 4. 興味と選択行動および注視行動との関係性分析

### 4.1 実験設定と分析方法

仮説の検証を行うために被験者実験を行った。実験参加者は 20 代から 30 代の 12 名であった。実験は、PSV を使用してチアリーディングの映像を、カメラ中心視点切り替えおよびターゲット中心視点切り替えを用いて、それぞれを 330 秒間視聴した。チアリーディングの映像は図 3 のように、16 台のカメラ \*1 で撮影され、映像内では 7 人の女性がチアリーディングの演技を行った。この 7 人の女性が本実験の選択対象であり、胸元を基準点とした。また、映像の長さは 110 秒であった \*2。

映像を視聴する際、実験参加者はモニター \*3 から約 1000mm 離れて座った。注視点座標を計測するため、モニターと実験参加者の間に視線計測装置 \*4 を設置した。

また、選択対象への興味の点数付けを行うため、映像視聴を行った後に、実験参加者にチアリーディングの演技を行っていた女性に対する興味に関する質問をした。質問の内容は For each target, please rate how INTERESTED you are in? (i.e. 1:no, ... 7:extremely) であった。

この質問の後に、実験参加者は 1 から 7 の 7 段階で興味の点数付けを行い、1 がもっとも興味度が低く、7 がもっとも興味度が高いものとした。この興味度を実験参加者毎に平均値を閾値とし、各実験参加者の興味度が高い対象と低い対象に分割した。

### 4.2 仮説 I の検証

仮説 I を検証するために、選択対象と注視点座標とのユークリッド距離を計算した。選択対象の基準点のディスプレイ平面上での座標  $\mathbf{x}_p = (x_p, y_p) \in \mathbb{Z}^2$  と注視点座標  $\mathbf{x}_g = (x_g, y_g) \in \mathbb{Z}^2$  との距離は  $\|\mathbf{d}\| = \sqrt{(x_p - x_g)^2 + (y_p - y_g)^2}$  として算出し

\*1 casio EX-F1,H.264/AVC 準拠,512×384pixels,300fps

\*2 実験参加者は 3 回映像を見返すことが可能である

\*3 三菱 RDF191S, 1024×768pixels

\*4 SR Research EyeLink 1000, 計測精度 0.5°, サンプルングレート 1000Hz

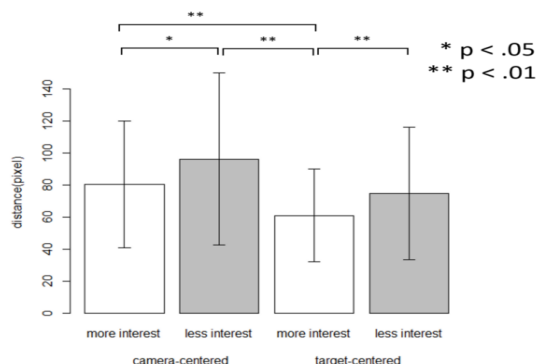


図 4: 選択対象と注視点座標との距離

た。各視聴方式毎に興味度が高い対象と興味度が低い対象それぞれについて距離を算出し、興味度の高低で距離に差が生じるかを統計的有意差検定を行うことで、仮説 I の検証を行った。また、距離は対象を選択していた各時区間で平均したものをを用い、有意水準は 5%とした。

検証を行った結果を図 4 に示す。カメラ中心視点切り替えについて、興味度が高い選択対象と注視点座標との距離（中央値:73.3px, 標準偏差:39.6px）は興味度が低い選択対象に対する距離（中央値:84.7px, 標準偏差:53.7px）よりも短くなり、Wilcoxon の順位和検定を行った結果、 $W = 9870$ ,  $p$  値  $< .05$  となり、有意差が確かめられた。また、ターゲット中心視点切り替えでは、興味度が高い選択対象との距離（中央値:56.0px, 標準偏差:28.9px）は、カメラ中心視点切り替えの場合と同様に、興味度が低い選択対象との距離（中央値:67.7px, 標準偏差:41.4px）よりも短くなり、Wilcoxon の順位和検定を行った結果、 $W = 15859$ ,  $p$  値  $< .05$  となり、有意差が確かめられた。

カメラ中心視点切り替え、ターゲット中心視点切り替えともに興味度が高い選択対象と注視点座標との距離は興味度が低い選択対象に比べて短くなり、有意差が確かめられた。これは選択した対象への興味の高低によって注視行動に影響があることを示す。この結果から仮説 I が支持される。

### 4.3 仮説 II の検証

仮説 II を検証するために、ある対象が選択されてから別の対象が選択されるまでの持続時間を計測した。検証方法は仮説 I と同様に興味度の高低で持続時間に有意な差が生じるかを検定した。

検定を行った結果を図 5 に示す。カメラ中心視点切り替えの結果として、興味度が高い対象に対しての持続時間（中央値: $5.59 \times 10^3$  msec, 標準偏差: $3.45 \times 10^4$  msec）は興味度が低い対象に対しての持続時間（中央値: $2.96 \times 10^3$  msec, 標準偏差: $1.26 \times 10^4$  msec）よりも長かった。しかし、Wilcoxon の順位和検定を行った結果、 $W = 13031.5$ ,  $p$  値  $> .05$  となり有意差は確かめられなかった。一方、ターゲット中心視点切り替えの結果として、興味度が高い対象に対しての持続時間（中央値: $7.52 \times 10^3$  msec, 標準偏差: $1.78 \times 10^4$  msec）は、興味度が低い対象に対しての持続時間（中央値: $3.94 \times 10^3$  msec, 標準偏差: $1.78 \times 10^4$  msec）よりも長くなり、Wilcoxon の順位和検定の結果、 $W = 11319$ ,  $p$  値  $< .05$  となり有意差が確かめられた。

カメラ中心視点切り替えでは、興味の高低で有意差は得られなかったが、興味度が高い対象の持続時間のほうが興味度が低い対象の持続時間より長くなった。また、ターゲット中心視

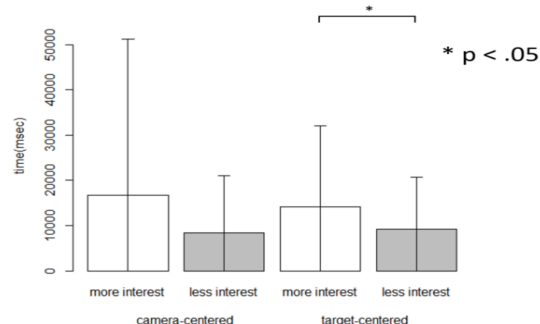


図 5: 対象を連続的に選択していた持続時間

点切り替えでは、興味度が高い対象のほうが持続時間が長く、また有意差も確かめられた。選択対象に対する興味の高低により、対象を選択していた持続時間に影響があることが分かり、この結果から、仮説 II が支持される。

### 4.4 仮説 III の検証

仮説 III を検証するために、ある対象を選択した回数を対象を選択した総数に対する割合として算出した。検証方法は仮説 I, II と同様に興味の高低で選択回数の割合に有意な差が生じるかを検定した。

検定結果を図 6 に示す。カメラ中心視点切り替えについて、興味度が高い対象を選択した回数の割合（平均:0.136, 標準偏差:0.0861）の平均は興味度が低い対象を選択した回数の割合（平均:0.122, 標準偏差:0.0573）の平均よりも多くなった。しかし、 $t$  検定の結果、自由度 66.0,  $t = 0.839$ ,  $p$  値  $> .05$  となり、有意差は確かめられなかった。また、ターゲット中心視点切り替えでも同様に興味度が高い対象を選択した回数の割合（平均:0.133, 標準偏差:0.0455）の平均は興味度が低い選択対象の選択回数の割合（平均:0.123, 標準偏差:0.0455）の平均よりも多くなった。しかし、 $t$  検定の結果、自由度 76.5,  $t = 0.9922$ ,  $p$  値  $> .05$  となり、有意差は確かめられなかった。

選択対象を選んだ回数についてはカメラ中心視点切り替え、ターゲット中心視点切り替えともに有意差が確かめられなかった。このことから仮説 III が支持されるとは言えない。この理由として、対象の選択は図 2 の右上に示される候補リストが用いられたため、このリストと被写対象の対応を把握するのは容易でなく、実際に選択したかった対象とは異なる対象を選んでしまうことがしばしばあった。そのため、選択回数と興味の高低との間に関係が得られなかったのではないかと考えられる。

## 5. ターゲット中心視点切り替えの効果の検証

ターゲット中心視点切り替えは選択した対象を中央に表示する方式であるため、対象に視線を集中しやすくなることや対象を視知覚しやすくなることで、より視聴者の興味度が注視行動や選択行動に表れることが期待できる。上記で行った分析結果を踏まえ、ターゲット中心視点切り替えの効果を検証した。検証方法は各視聴方式間で、上記の分析で興味との関係性が確かめられた選択対象と注視点座標間の距離および対象を選択した持続時間に有意差があるかについて有意差検定した。

### 5.1 選択対象と注視点座標間の距離を用いた検証

仮説 I の検証で用いた選択対象と注視点座標間の距離（図 4）を指標として検証した。結果として、興味度が高い対象との距



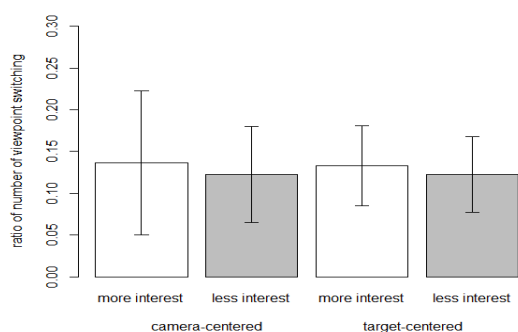


図 6: 全対象の選択回数の合計に対する対象を選んだ回数の割合

離の中央値は、ターゲット中心視点切り替えで 56.0pixel となり、カメラ中心視点切り替えの 73.3pixel より短く、Wilcoxon の順位和検定の結果、 $W = 13830$ ,  $p$  値  $< .05$  となり有意差が確かめられ、また、興味度が低い対象との距離の平均は、ターゲット中心視点切り替えで 67.7pixel となり、カメラ中心視点切り替えでの 84.7pixel より短く、Wilcoxon の順位和検定の結果、 $W = 18306$ ,  $p$  値  $< .05$  となり有意差が確かめられた。一方、興味度が低い対象との距離の中央値は、ターゲット中心視点切り替えで 67.7pixel であり、興味度が高い対象との距離の中央値は、カメラ中心視点切り替えで 73.3pixel となり、Wilcoxon の順位和検定の結果、 $W = 13088$ ,  $p$  値  $> .05$  となり有意差は確かめられなかった。

検証の結果、興味度が高い対象との距離はターゲット中心視点切り替えが最も短くなった。このことから、ターゲット中心視点切り替えは興味度が高い対象に物理的に視線を釘付けるのに効果的であると言える。また、ターゲット中心視点切り替えでの興味度が低い対象と注視点座標間の距離は、カメラ中心視点切り替えの興味度が高い対象との距離とほぼ同じ値となった。この結果から、ターゲット中心視点切り替えは、常に対象に視線を集中させるわけではないと言える。

## 5.2 対象を選択していた持続時間を用いた検証

仮説 II の検証で用いた、対象を選択していた持続時間 (図 5) を指標として検証した。結果として、興味度が高い対象を選択していた持続時間の中央値は、ターゲット中心視点切り替えで  $5.59 \times 10^3$  msec, カメラ中心視点切り替えで  $7.52 \times 10^3$  msec であり、Wilcoxon の順位和検定の結果、 $W = 9379.5$ ,  $p$  値  $> .05$  となり有意差は確かめられなかった。また、興味度が低い対象を選択していた持続時間の中央値は、ターゲット中心視点切り替えで  $3.94 \times 10^3$  msec, カメラ中心視点切り替えで  $2.96 \times 10^3$  msec であり、Wilcoxon の順位和検定の結果、 $W = 13103.5$ ,  $p$  値  $> .05$  となり有意差は確かめられなかった。

この結果から、対象を選択していた持続時間は視聴方式に関わらず、ほとんど差がないということが分かり、釘付け視聴方式を使用しても、対象を選択する時間と興味との関係が強まらないということが示唆される。

## 6. おわりに

本研究では釘付け視聴方式の効果を検証するために、視聴者の視対象に対する興味の高低と注視行動および対象の選択行動との関係性について分析した。結果として、注視行動については、選択対象と注視点座標との距離は興味が高いと短くな

り、注視行動は従来知見通り、興味との関係性が確認できた。一方、選択行動については、持続時間は興味が高いと長くなったが、選択回数は興味の高低によらず、ほぼ同じであった。また、各視聴方式間で、選択対象と注視点座標間の距離および対象を選択した持続時間に有意差があるかについて統計的検定を行った結果、持続時間は視聴方式間で差がなかったが、対象と注視点座標との距離は、興味が高い対象に対して、釘付け視聴方式を適用したほうが短くなった。つまり、釘付け視聴方式は興味が高い対象に視線を釘付けるために効果的であると分かった。また、多視点映像視聴時は、無意識的な行動である注視行動の方が意識的な行動である選択行動より興味との関係性が強い可能性がある。

今後の課題として、より詳細な指標を用いた分析や広い範囲で対象が動く動画を視聴した場合の分析が挙げられる。

## 7. 謝辞

本研究の一部は、総務省 SCOPE(111706003) および科研費若手研究 B (23700168) の支援によるものである。

## 参考文献

- [1] Kanade, T., et al. "Virtualized reality: Perspectives on 4D Digitization of Dynamic Events.", pp.32-40, 2007.
- [2] Kameda, Y., et al. "Free viewpoint browsing of live soccer games." IEEE International Conference on Multimedia and Expo, Vol. 1., 2004.
- [3] Tokai, S., et al. "Pegged to point browsing: An approach to browse multi-view video with view-switching, and its applications." ICPR workshop on Sensing Web. 2008.
- [4] Mase, K., et al, "Socially-assisted Multi-view Video Viewer", International Conference on Multimodal Interaction, pp. 319-322, 2011.
- [5] Hirayama, T., et al. "Analysis of Gaze Behavior while using a Multi-Viewpoint Video Viewer", Symposium on Eye Tracking Research and Applications, 2014.
- [6] 平山高嗣, "人間の内部状態を顕在化する視覚的インタラクション", 情報処理学会研究報告 CVIM, 2013.
- [7] Qvarfordt, Q., Zhai, S., "Conversing with the User Based on Eye-Gaze Patterns", Proc. the ACM CHI Human Factors in Computing Systems Conf., pp.221-230, 2005.
- [8] 黒木修隆ら, "テレビ視聴者の選局行動に基づく番組嗜好度の推定. 映像情報メディア学会誌", vol.60, no.3, pp.454-457, 2006.
- [9] 土屋誠司ら, "TV 番組推薦システムの構築とその有用性の検証", 情報処理学会研究報告 HI, pp.95-102, 2006.