

追体験メディアにおける注意誘導の呈示手法

黄 楊陽 杉山 治 大塚 琢馬 奥乃 博

京都大学大学院 情報学研究科 知能情報学専攻

1. はじめに

映像・音声に関する収録・再生技術の発展に伴って、個人で追体験メディアコンテンツを作成する環境が整備されている。作成にあたって、必要な編集・鑑賞・共有ツールも数多く提供されており、追体験メディアの繁栄に貢献している。さらに、360°全方位レンズの公開によって新種類の全方位映像コンテンツを収録することが可能になる。従来の映像コンテンツの多くは第三者視点から客観的に撮影されるものであって、「映像の鑑賞=シーンの理解」というイメージが強い。それに対し、人物の視点から収録した全方位映像が「視点の回転」ファクターを加味する。ゆえに、臨場感の高い「映像の鑑賞=シーンの参加」感覚が与えられる。[1] しかしながら、全方位映像に相応する編集・再生技術が欠けている。本研究は以下の観点から、全方位映像を編集・再生する為のシステムを設計し、さらに、当システムにおいて視聴者の注意誘導手法を提案する。

視点回転のために特化した再生システム: 360°全方位レンズで収録された全方位映像が既にウェブ上多く共有されている。[2] それらの公開映像の再生プレイヤーがマウスのドラッグ操作で視点の方向を回転させ、映像内容を選択して表現するが、「シーンの参加」に寄与することが達成されてない。それに対して、本研究はジャイロ付きヘッドマウントディスプレイ (HMD) を持ち上げて、視聴者の顔向きをセンシングし、視点方向を自然に取得する。

臨場感向上のための立体音響技術の適用: 映像コンテンツと同時に音響コンテンツを入手する。視聴者の顔向きに基づいて映像内容の選択を行うゆえに、音響コンテンツが同じセンシングした顔向きに相応する表現の需要が高まる。臨場感の高い立体音響再生技術には、WFS (Wave Field Synthesis) や、sweet spot と呼ばれる特定位置だけに高臨場感を再現するステレオ再生技術などがある。いずれも、設置場所に制約があるので、HMD と組み合わせる使用するには、ヘッドホンの方が好都合である。また、本研究では、全方位映像と立体音響再生を組み合わせ、全方位マルチメディアと呼ぶ。

作成側の「意図」を表現するの再生モデル: 全方位マルチメディアでは、全ての映像を全部目に収めることが不可能である。作成側が「意図」を持って全方位マルチメディアを共有することがあり得る。例えば、背後から声をかけてきた人に視聴者の視点を向けてほしい場面が考えられる。本研究では、音響コンテンツの再生手法に工夫して、作成側の「意図」を表現し、視聴者の注意を誘導する手法を提案する。図1で示したように、視聴者の注意を視点の方向として定義する。

2. 従来研究と本研究の位置づけ

全方位映像に関して、近藤らが主観視点の全方位映像と三次元音場再現を組み合わせることを目指して、First-Person Perspective Omnidirectional camera (FIPPO) を含んだコンテンツ収録、メディア処理、呈示システムを提案・評価した [1]。FIPPO で優れた高解像度の全方位映像を収録して、四つの平面スクリーンで視聴者を囲む形で映像を再生する。

それに対して、本研究で提案する体験システムの収録・再生のための機器は一般視聴者が簡単に入手できる安価なものとなり、誰でも発信・鑑賞できるメディアが特徴である。全方位映像を撮影するために、Home360°社が販売している GoPano Micro を利用する。Apple 社のス

マートフォン iPhone4/4s/5 の専用ケースに固定して、水平方向 360°・垂直 80° のパノラマ動画撮影を可能である。展開される動画の解像度が 1920*448 となる。解像度が低下するわりに、町にて気軽に片手で撮影できること、そして動画展開・共有するソフトウェアがすでに公開されていることが便利で有る。

注意誘導に関して、萩原らが画像中の指定領域の視覚的顕著性を高めて、人の注視を自然に誘導する手法に関する議論した [3]。課題として挙げられていたのは、指定領域での画素処理が画像の見えに違和感を発生する問題がある。それに対して、音響コンテンツを用いた注意誘導が視覚に違和感を与えないメリットがある。

本研究が提案するシステムの構成が図1で示すように、収録用機材はスマートフォンとそれに装着するレンズで、持ち歩きやすい形式を整えている。全方位マルチメディアを再生する時に、体験者の顔向き情報を HMD (Oculus Rift) についているジャイロセンサから取得し、センシングされた向き情報に基づいて、視覚、聴覚情報の制御を行う。視覚の部分では、カメラで撮影したパノラマ画像を HMD の再生に適合させるために、ユーザの顔向きに基づいて、部分画像の切り出しとディストーションを行う。聴覚の部分では、三次元仮想空間に配置された音源を再現するために、ユーザの顔向きに基づいて、注意誘導操作を行う。

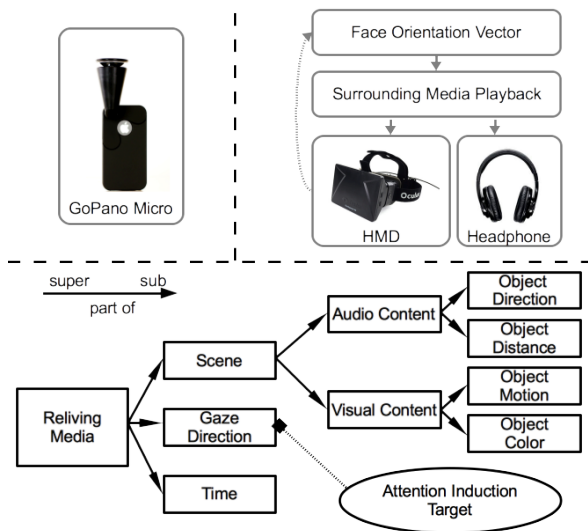


Figure 1: 左上: 収録用 GoPanoMicro の実物写真, 右上: 再生の概要, 下: 全方位マルチメディアオントロジー

3. 追体験メディアのための編集ツールの開発

収録されたローデータが展開したパノラマ画像の時系列データとモノチャンネル音響信号を含んでいる。本研究で提案した全方位マルチメディアを再生するために、収録されたローデータをアニメーションする必要があるため、追体験メディアのための編集ツールを開発した。編集ツールで作成側が付与するメタデータが二種類ある。

立体音響再生するための方向情報と距離情報: 立体音響再生を達成するための前提条件は音源と視聴者の位置関係が確定されることである。方向情報と距離情報を付

与するために、音響信号と画像において音源となる物体の対応関係、音源となる物体の方向情報を常に確認できるように要求される。例えば、人の話し声の音響信号に対して、方向・距離の時系列情報を作成するために、画像で話者を確認できることと、話者の方向を明示することが重要である。

作成側の「意図」を表す注意曲線： 全方位マルチメディアにおいて、360°の画像情報が提供されて、それを一斉に目に収めることが不可能である。作成側が視聴者に見てほしい物体・方向があって、全方位マルチメディアを共有すると想定できる。ゆえに、見てほしい物体・方向の軌跡を表現する時系列情報を作成するツールが必要である。

編集ツールの作成にあたってオープンソースツールキット OpenFrameworks 0.74 を利用した。立体音響再生にあたって、公開ライブラリ OpenAL Soft 1.15.1 を利用した。

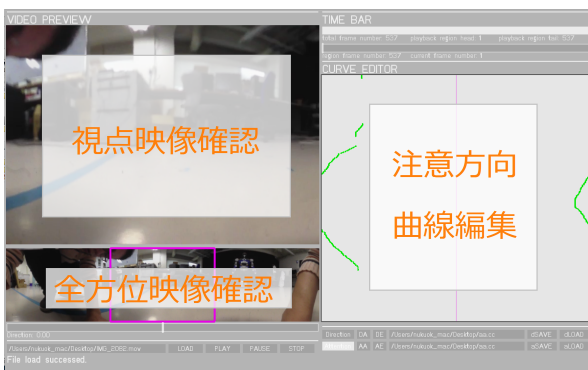


Figure 2: 編集ツールを用いてメタデータ付与する場面、全方位映像・視点映像を確認しながら音響方向および注意曲線を編集するように設計する。

4. 注意誘導する手法の提案

従来研究では、視覚的な外部刺激を注意誘導するためのボトムアップ方式の要因として、顕著性マップモデルが構築される。[4] 顕著性マップにモデル化されるのは色相、輝度、運動方向などの視覚情報である。音響信号で注意誘導するボトムアップ方式の要因として、運動方向であると考慮して、注意誘導する手法を提案する。

ヘッドホンの左右チャンネルに出力する波形 Y^L, Y^R が OpenAL ライブラリによって以下の入出力形式で計算される。各音源 i の相対位置を (θ_i, d_i) 、波形を x_i とし、視聴者の顔向き ω 、出力をステレオ波形 Y^L, Y^R 、

$$\begin{cases} Y^L(t) = \sum_i G_i^L(x_i, \theta_{i,t}, d_{i,t}, \omega_t) \\ Y^R(t) = \sum_i G_i^R(x_i, \theta_{i,t}, d_{i,t}, \omega_t) \end{cases} \quad (1)$$

注意誘導手法を加味して、OpenAL ライブラリへの入力を動的に修正する。 β が作成側の「意図」を表す注意してほしい方向を表す。

$$\begin{cases} Y^L(t) = \sum_i G_i^L(x_i, \theta_{i,t} + \Delta\theta_t, d_{i,t} + \Delta d_t, \omega_t) \\ Y^R(t) = \sum_i G_i^R(x_i, \theta_{i,t} + \Delta\theta_t, d_{i,t} + \Delta d_t, \omega_t) \end{cases} \quad (2)$$

$$\Delta\theta_t = A(\theta_{i,t}, \omega_t, \beta_t) \quad (3)$$

$$\Delta d_t = A(d_{i,t}, \omega_t, \beta_t) \quad (4)$$

5. 注意誘導する手法の評価実験

仮定： 提案した注意誘導操作に関して、それを適用された音響信号が適用されていない音響信号に比べて、視聴者の視点を回転させることには優位である。

実験概要： 仮定を検証するために、収録されたローデータから全方位マルチメディアを編集して、被験者に視聴してもらい、左側と右側からそれぞれ注意誘導操作の適用あり・なしの音響信号を再生する。被験者が見たい方向に首を回すように指示する。ただし、視覚的に刺激の影響を排除するために、映像コンテンツについて、収録された映像の半分を棄却し、左右が対称するように加工する。音響信号に関して、両側から立体音響計算するための波形 x_i を同一される。

実験手順：

1. 被験者教示。被験者に対してシステムの動作とHMDに掲示される画像の関係を説明し、HMDとヘッドホンを着用させて、デモ用全方位マルチメディアを視聴してもらう。
2. 実験用の全方位マルチメディアの視聴にあたって、「音声聞こえるまで、視点を回転せずに顔を前方向に向ける。音声が再生し終わったら、左右方向からより気になる一つの方向を選んで、首を回すこと。」
3. 手順2を繰り返して、首の回った方向を記録する。
4. 実験用全方位マルチメディアの再生が終了した後、「被験者左右両眼の視力差、全方位マルチメディアの画質、要望するシーン」などの項目にアンケートを行う。

6. おわりに

本研究では、全方位マルチメディアの共有に需要の高まった編集・再生システムを設計・実装し、さらに、作成側の「意図」を表現するための音響信号による注意誘導手法を提案して検証する。

謝辞： 本研究の一部は科研費基盤 (S) の支援を受けた。

References

- [1] Kondo Kazuaki, Yasuhiro Mukaigawa, and Yasushi Yagi. "Wearable imaging system for capturing omnidirectional movies from a first-person perspective." In Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology, pp. 11-18. ACM, 2009.
- [2] www.gopano.com
- [3] Hagiwara Aiko, Akihiro Sugimoto, and Kazuhiko Kawamoto. "Saliency-based image editing for guiding visual attention." In Proceedings of the 1st international workshop on pervasive eye tracking & mobile eye-based interaction, pp. 43-48. ACM, 2011.
- [4] Itti, Laurent, and Christof Koch. "Computational modeling of visual attention." Nature reviews neuroscience 2, no. 3, pp. 194-203. 2001.