

360度動画に対する視聴覚的な物体削除

島村 僚^{1,a)} 馮 起¹ 小山 裕己² 中塚 貴之¹ 深山 覚² 濱崎 雅弘² 後藤 真孝² 森島 繁生^{3,b)}

概要：

本研究では、360度動画内における対象の映像と音を選択的に削除し補完する手法を提案する。従来の動画補完は、削除対象の映像のみを動画から削除するため、何も無い場所から音が聞こえてくるといったように映像と音に矛盾が生じていた。本研究ではこの問題を解決するために、削除対象を360度映像上で追跡し、削除対象の方角に対応する音を特定する。これにより、360度動画から削除対象の映像と音を同時に削除し、補完した360度動画の生成を可能とした。本手法の有用性を示すために主観評価実験を行った。主観評価実験を通して、提案手法は実験参加者にとって自然で満足度の高い360度動画を生成し、体験の質を高めることを確認した。

1. はじめに

近年、360度動画を対象とした動画共有サービスや撮影デバイスが普及したことで、誰もが360度動画を発信できるようになった。しかし360度動画を撮影する際、撮影領域が広いために撮影者の意図していない物体が写り込んでしまう場合がある。そのため、360度動画から撮影者が不必要だと感じた物体のみを削除する編集方法があると有用である。動画から特定の物体のみを削除する手法としてVideo Inpainting [1]がある。Video Inpainting [1]は動画から削除対象の映像のみを削除し、映像を補完するため、削除対象の音はそのまま残ってしまう。結果として、何も無い場所から音が聞こえるといった不自然な動画が生成される。本研究では、360度動画から削除対象の映像に加え、360度動画に含まれる立体音響データから削除対象が発している音も同時に削除する手法を提案する。本手法の有用性を示すために主観評価実験を実施し、提案手法が実験参加者の動画視聴体験を改善することを示した。

2. 提案手法

提案手法の概要図を図1に示す。まずSiammask [2]を用い、動画の全フレームにおいて削除対象を追跡しマスクを推定する。推定されたマスクに対してDeep Video Inpainting [1]を適用することで、削除対象の映像を削除する。次にDirAC [3]を用いて音源の方角を推定する。Siammask [2]

の追跡結果とDirAC [3]で推定された角周波数の音源方向を比較することで、削除対象から発せられた音源の周波数帯を特定し削除する。

3. 実験

3.1 設定

360度動画から削除対象を矛盾なく削除することで動画視聴体験を高められるという仮説を検証するために17人(年齢:23.1±1.47)の実験参加者による主観評価実験を実施した。主観評価実験では、360度動画から特定の物体の(1)視覚的、(2)聴覚的、(3)視聴覚的に物体を削除する3通りの編集方法を適用した動画について、編集結果の自然さ、満足度を5段階リッカート尺度を用いた質問に回答してもらうことで調査した。また、検定にはウィルコクソンの符号付順位検定を行った。実験で用いた動画は、性質の異なる3つのシーンを撮影して使用した(表1)。

表1 実験使用シーン(太文字:特徴的部分)

シーン	音源	削除対象	動き	周波数変化	環境音
1	話者/ボーダー	ボーダー	あり	ほぼなし	鳥の鳴き声
2	ピアノ/マラカス/バイオリン	ピアノ	なし	あり	なし
3	話者/話者	話者	あり	ほぼなし	鳥の鳴き声

3.2 結果と考察

主観評価実験の自然さの結果を図2、満足度の結果を図3に示す。自然さの結果は図2より、視聴覚的に物体を削除した時が他の削除方法に比べて一番高い点数であることがわかる。特に、視聴覚的削除と視覚的削除の結果には有意差が顕著に現れた(シーン1: $p = 0.0021$, シーン2: $p = 0.0022$, シーン3: $p < 0.001$)。一方、視聴覚的削除と

¹ 早稲田大学

² 産業技術総合研究所

³ 早稲田大学理工学術院総合研究所

a) s-ryo@akane.waseda.jp

b) shigeo@waseda.jp

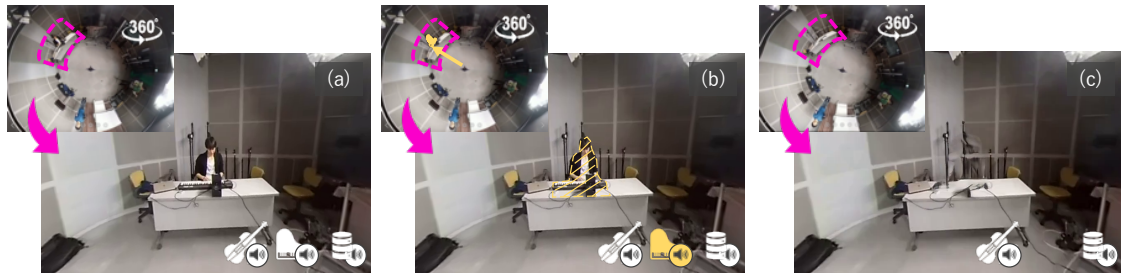


図 1 360 度動画に対する視聴覚的な物体削除の概要図。(a) 360 度映像の初期フレームで削除する対象を選択。(b) 削除対象を視覚情報を用いてトラッキングし、削除対象の方向と推定方向が一致する音の周波数帯を特定。(c) 削除対象を視聴覚的に削除

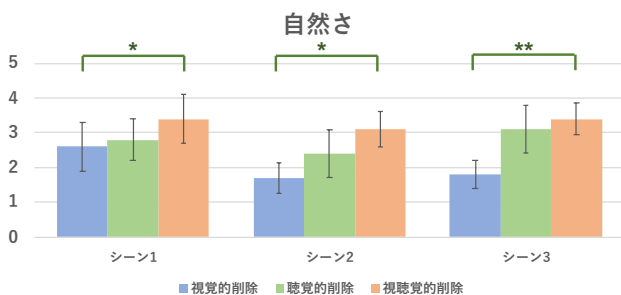


図 2 主観評価実験の自然さのスコア結果 (リッカート尺度 5 段階評価). エラーバー: 標準偏差, *: p 値 0.05 未満, **: p 値 0.01 未満.

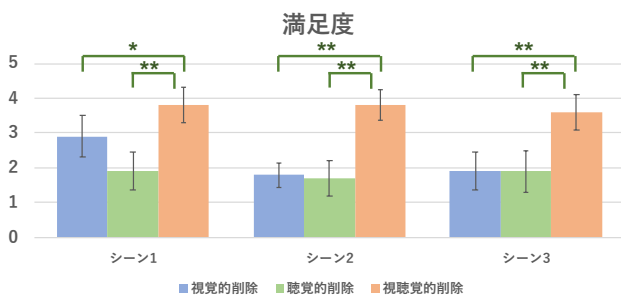


図 3 主観評価実験の満足度のスコア結果 (リッカート尺度 5 段階評価). エラーバー: 標準偏差, *: p 値 0.05 未満, **: p 値 0.01 未満.

聴覚的削除の間には有意差は出なかった。これは、音のみを削除したとしても対象が元から音を発していたのかどうか分からないためそれほど不自然に感じなかったのだと思われる。

満足度の結果は図 3 より、自然さと同様視聴覚的に物体を削除した時が他の削除方法と比べて一番高い点数であることがわかる。満足度に関しては、視聴覚的削除はその他の削除方法のどちらの間にも有意差があった。視覚的削除との間にはシーン 1 では $p = 0.0025$ であった。その他の部分は図 3 にあるように $p < 0.001$ という結果になった。これらの結果から、満足度は自然さと異なり、視覚情報、聴覚情報の片方が残っているだけでも違和感につながり満足度の評価が下がっていることが考えられる。

4. 今後の課題

手法の実装と主観評価実験を通じて、視聴覚的に削除をすることが 360 度動画の自然さ・満足度を向上させることがわかった。一方で、本手法を適用するシーンによっては視聴覚的な削除ができない問題を含むことがわかった。一つ目は、削除対象が他の物体で遮られるとトラッキングが外れてしまい、音の方向推定が失敗し聴覚的な削除ができない点である。二つ目は、本手法は短時間フーリエ変換でいくつかの周波数帯に分割するのでシーンに周波数帯に近い音源があると同一周波数帯に含まれてしまうという点である。上記の問題を解決することが主な今後の課題である。また、今回の主観評価実験では編集済みの動画の視聴しかできないため、今後は任意の対象を選択して削除できるようなインターフェースを作成する予定である。

5. 結論

本研究では、360 度動画から特定の物体を視聴覚的に削除する手法を提案した。本手法は視覚的削除、聴覚的削除の 2 つの処理を行うことによって 360 度動画から特定の対象を視聴覚的に削除することを実現した。実験を通して、視覚的、聴覚的に削除した時が片方のみを削除した時よりも自然で満足度の高い結果を出力できることを示した。

謝辞 本研究は、JST ACCEL NO.JPMJAC1602、及び、JST 未来社会創造事業 NO.JPMJMI19B2 の支援を受けたものである。

参考文献

- [1] D. Kim, S. Woo, J.-Y. Lee, and I. So Kweon. Deep video inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5792–5801, 2019.
- [2] Q. Wang, L. Zhang, L. Bertinetto, W. Hu, and P.H. Torr. Fastonline object tracking and segmentation: A unifying approach. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019.
- [3] V. Pulkki. Spatial sound reproduction with directional audio coding. *Journal of the Audio Engineering Society*. 55(6):503–516, 2007.