

18

音だけから情景を推定

～音から画像認識結果を予測するクロスメディア情景分析～



どんな研究

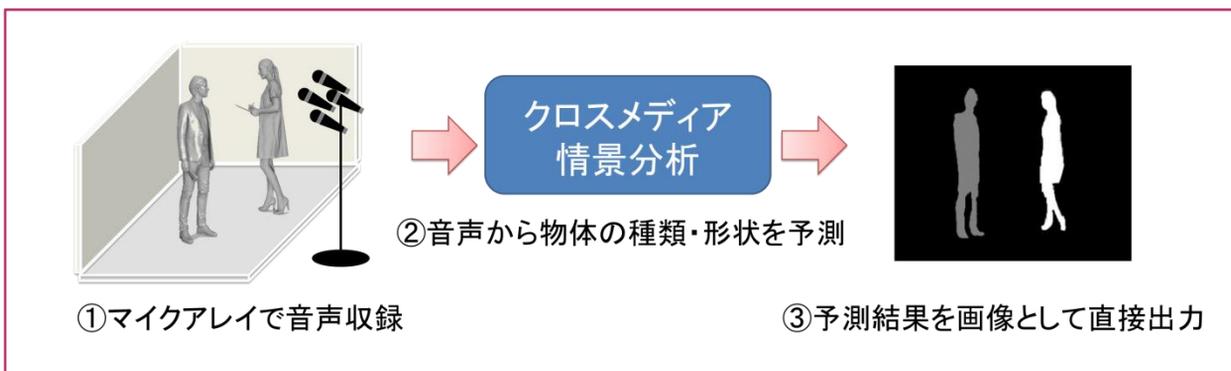
部屋などの屋内空間に配置したマイクロホンアレイにより収録した音だけを分析することによって、まるで画像認識したかのように「どこにどんな物体があるのか」を推定する研究です。生活や公共空間での見守りや防犯等への応用に向けての研究開発を進めています。

どこが凄い

『音から画像認識結果を予測する』という試み自体、これまで取り組み例のない新しい技術課題です。本研究では、入力が音響特徴、出力が画像認識結果となるような、クロスメディアな深層学習モデルを設計し、技術的な実現可能性を初めて示しました。

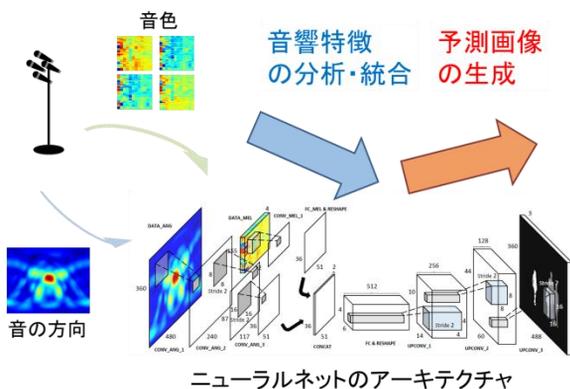
めざす未来

現在の見守り・防犯技術はカメラの利用を前提としており、プライバシー性の高い家庭や公共空間には適用しにくい場合があります。本技術により、カメラの設置が好ましくない空間の様子もわかりやすく確認できるようになるため、適用領域の拡大が期待できます。

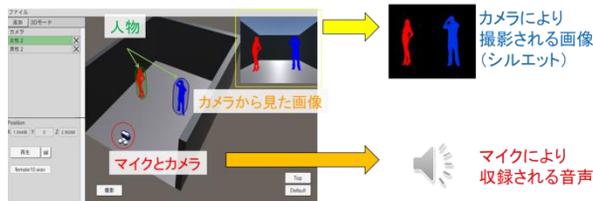


音響特徴から、空間にある物体の種類や位置を予測する深層学習モデルを提案

異なる音響特徴を分析・統合し、出力画像を生成する処理を一つのニューラルネットにより実現

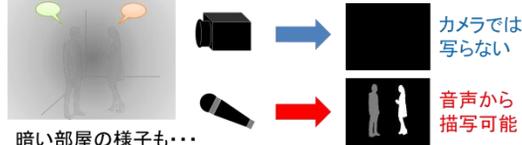


多様な状況をシミュレーションにより再現することで学習を効率化し、実現性を確認



カメラでは写らない・写したくない空間の様子も、音だけから情景を描写可能

見にくい暗がりや写したくない空間の様子も音から描写



関連文献

[1] オストレクミレラ, 入江豪, 亀岡弘和, 木村昭悟, 平松薫, 柏野邦夫, “Seeing through Sounds: 音響情報からの視覚的情景理解に向けて,” 画像の認識・理解シンポジウム, 2017.

担当者

入江 豪 (Go Irie) メディア情報研究部 メディア認識研究グループ