日本―スペイン、ポーランド 国際共同研究「レジリエント、安全、セキュアな社会のための ICT」 2021 年度　年次報告書	
研究課題名（和文）	ソーシャルメディアプラットフォームにおけるフェイクニュース検出(DISSIMILAR)
研究課題名（英文）	Detection of Fake News on Social Media Platforms (DISSIMILAR)
日本側研究代表者氏名	栗林　稔
所属・役職	岡山大学　大学院自然科学研究科・准教授
研究期間	2021 年 4 月 1 日～2024 年 3 月 31 日

１．日本側の研究実施体制

氏名	所属機関・部局・役職	役割
栗林　稔	岡山大学・大学院自然科学研究科・准教授	研究全体の統括

２．日本側研究チームの研究目標及び計画概要

フェイク画像やビデオ、音声などを正しく見分ける技術として、フェイクコンテンツに含まれる不自然な信号を検出してフェイクを見破る技術を実現するシステムの提案と評価実験を行う。フェイクコンテンツとして操作された情報とオリジナルの情報を深層学習技術により判別できるシステムの設計を目指す。また情報の漏洩元を追跡する電子指紋情報を扱った情報埋め込み手法および抽出法の検証と、暗号プロトコルも含めたシステムの設計を目指す。

3. 日本側研究チームの実施概要

本プロジェクトでは、フェイクコンテンツを検出するためにハイブリッド型のアーキテクチャを構想している。一つは加工・編集などの処理により生じる歪みや深層学習技術に基づ

く画像生成器に由来する不自然な信号成分を解析するフォレンジクス技術であり、受身的な対策である．もう一つは、積極的にコンテンツを守るために原本性保証の情報や改ざん検知信号を忍ばせる対策である。

受身的な対策では、フェイク画像やビデオ、音声などを正しく見分ける技術として、フェイクコンテンツに含まれる不自然な信号を検出してフェイクを見破るフォレンジクス技術を扱う。フェイクコンテンツを解析するためには、フェイクコンテンツが作成される過程について知る必要がある。そこで本年度は、まずその生成方法に関する関連技術を調べるところから手掛け、ディープフェイクと呼ばれるフェイク動画は、架空の顔を生成、人物の顔を他の顔と入れ替え、顔の特徴を変更、人物の表情を他の人の表情と入れ替えなど４つのカテゴリとそれ以外に分類した。またディープフェイクを識別する技術動向についても調べ、現状の課題点と将来的に可能性のある技術的な方向性とその展望についてサーベイ論文(IEEE Access)としてまとめた。

フェイクコンテンツ中の顔領域において人工的に加工・編集された形跡を解析するアプローチは従来から広く研究されているが、顔検出や顔の各パーツの検出を阻害する敵対的ノイズ付加への耐性が課題として指摘されていた。本研究では、同じような特徴を持つ局所領域に分割する手法を用いて領域に分けるアプローチを試みた。また、分割した領域ごとにフェイクであるか否かを識別し、それらの結果をまとめて総合的な判断を行う流れで最終的に判断する手法を考案し、実装を進めている。識別器には畳み込みニューラルに基づく画像二値分類器をファインチューニングしたモデルを採用している。

積極的な対策としては、サイバー空間におけるトレーサビリティを確保するための電子指紋技術を扱っている。本年度は、電子指紋技術と暗号技術とを組み合わせた暗号プロトコルの実装について、クラウドサーバ上で運用することを想定して実装を進めた。クラウドサーバ構築用のパッケージを拡張させて、テキストなどを含む文書ドキュメントファイルに電子指紋情報を埋め込む処理を組み込むところまで実現することに成功している。現時点では対象を扱いやすいテキストに限定している。今後この処理を動画像のビデオストリームに拡張できるようにする予定で研究を進めている。

フェイク動画への電子透かし技術の適用においては、研究成果を国際会議 APSIPA ASC 2021 で発表した。動画中の人物の唇の動きを特徴成分として取り出し、秘密鍵を用いて変換した信号成分を音声フレームのデータ中に埋め込む手法を提案した。シミュレーションにより、通常の編集内において生じる歪みに対して耐性を有することを確認している。今後は、映像と音声との同期まで考慮した手法へと拡張する予定である。