

報道機関 各位

国立大学法人電気通信大学

光学フィルタを用いてディープフェイク検出器を欺く 新たな攻撃とその対策

【ポイント】

- * モニターをカメラで撮影することで、AI 生成された Deepfake 画像を、あたかも本物であるかのように装う再撮影攻撃が問題となっている。
- * そのような脅威への対策として、撮影と同時にシーンの奥行きを計測することで、平らなモニターと実物を見分ける技術が研究されてきた。
- * 奥行き情報を計測する赤外測距センサーに対して物理的に偽情報を提示することで、従来の対策法を破る新たな攻撃法を提案した。
- * また、提案した攻撃法を防ぐための新しい対策法として、単一のセンサーのみで撮影と奥行き情報の計測を行う方式を提案した。

【概要】

電気通信大学大学院情報理工学研究科情報学専攻博士前期 2 年の礎哲氏、同専攻の菅原健准教授の研究グループは、フロリダ大学と共同で、画像の来歴情報（※1）（Provenance）への新たな攻撃とその対策を提案しました。

社会問題となっている AI 生成した偽画像（Deepfake（※2））へ対策するために、その画像が確かにカメラで撮影されたという来歴情報を付与する技術が普及しつつあります。それに対し、モニターに映した Deepfake 画像をカメラで撮影することで、偽の来歴情報を付ける再撮影攻撃（※3）が脅威として知られています。再撮影攻撃への対策法として、カメラに取り付けた赤外測距センサーを用いた奥行き情報を計測することで、平坦なモニターと実際のシーンを見分ける技術が研究されています。

本研究はまず、そのような従来の対策法を破る新たな攻撃法を提案しました。攻撃のアイデアは、カメラの前に設置した光学フィルタにより、可視光と赤外光の光路を分離することにあります。すると、カメラには平坦なモニターを提示しながら、赤外測距センサーには偽の奥行き情報を提示できるようになり、結果として従来の対策法を回避できてしまいます。この新しい攻撃が可能になるのは、画像の撮影と奥行き情報の計測に、別の色の光（可視光と赤外光）を利用するためです。そこで、撮影と奥行き情報の計測を、どちらも可視光で行うことで、上記の攻撃を防ぐ新たな対策法を提案しました。

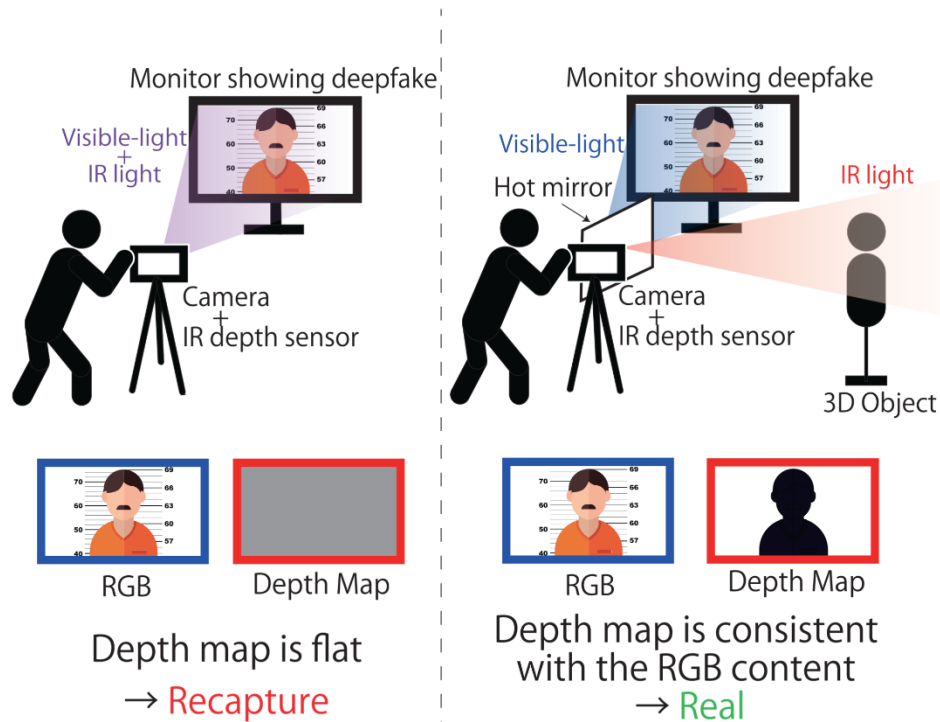


図1：赤外測距センサーを利用した再撮影検出への新たな脅威

【背景】

近年、急速なAI技術の発展により、Deepfake画像の生成が深刻な社会問題となっており、有名人へのなりすましやソーシャルメディア上のフェイクニュースを通じた世論の操作などの事例が報告されています。このDeepfake問題に対応すべく、AIなどによって生成された画像とカメラで撮影された画像を区別するための技術が活発に研究されています。Coalition for Content Provenance and Authenticity (C2PA (※4))はそのための標準規格で、カメラ内で撮影時にデジタル署名を付与することで実際のカメラで撮影したことを証明することができます。

しかし、この方式はモニターなどに写された画像を撮影する攻撃（再撮影攻撃）に対して脆弱です。この方式に準拠するカメラでモニターなどに映したDeepfakeを撮影することによって、画像の内容はDeepfakeであるにもかかわらず、実際のカメラで撮影したというデジタル署名を付与することができてしまい、画像の来歴の保証を損ないます。

この再撮影攻撃を防ぐために、奥行き情報を利用する再撮影検出方式が提案されました。この方式では、スマートフォンに搭載されている赤外光による深度センサー（LiDAR (※5)）を使用して、カメラでの画像の撮影と同時にその奥行き情報を計測します。撮影された画像とその奥行き情報の整合性を検証することによって再撮影を検出します。例えば、図1の左に示されるDeepfakeの再撮影では、撮影されるモニターは平面で奥行き情報は均一で撮影された画像中の人物の奥行き情報に整合しないため、再撮影として検出されます。この方式は高い精度で再撮影を検出することができ、再撮影攻撃に対する安全な検出方式とされていました。

【手法】

本研究では、赤外光による深度センサーを利用する従来の再撮影検出方式を迂回する、新たな脅威を発見しました。ホットミラー (※6) と呼ばれる特殊なフィルターは、可視光を透過する一方で、ある角度で入射した赤外光を反射する性質を持ちます。このホットミラーをスマートフォンの前に設置することによって、画像ではモニターの画面が撮影されているにも関わらず、奥行き情報の計測は攻撃者の用意した空間で行われるようになり、Deepfakeに対応した奥行き情報が計測されるようになってしまいます（図1の右）。その結果、画像と奥行き情報が整合してしまい、赤外測距センサーを利用した再撮影検出が回避されてしまいます。

この脅威を防ぐために、単一のセンサーのみを利用した安全な再撮影検出方式を構築しました。

本方式では Dual-Pixel センサー（※7）（DP センサー）と呼ばれる、効率的なオートフォーカスのためにスマートフォンやデジタルカメラに広く使用されるイメージセンサーを使用します。このセンサーは各画素が左右に分割されており、画像の撮影に加えて左右の視点の画像（DP 画像）を取得することができます。DP 画像には奥行き情報が含まれており、その奥行き情報を復元して撮影された画像との整合性を検証することにより、再撮影の検出を行います。同じ可視光で画像の撮影と奥行き情報の計測が行われるため、本研究で発見された脅威のような、画像の撮影と奥行き情報の計測の空間を分離する攻撃が不可能で、安全な撮影を行うことができます。

【成果】

ホットミラーと呼ばれる、可視光を透過して赤外光を反射する特殊なフィルターをスマートフォンの前に設置することで、撮影の可視光と、奥行き情報計測の赤外光を分離してそれぞれの空間を別々にできることを実証しました。実験の結果、この原理を利用して最先端の赤外測距センサーを利用した再撮影検出方式が 100%の成功率で回避されてしまうことが実証されました。

この脅威に対応するために、Dual-Pixel センサーを利用した、新たな再撮影検出方式を構築しました。単一のセンサー上で、撮影と奥行き情報の計測が同じ可視光で行われるため、本研究で新たに発見された、撮影と奥行き情報の計測の空間を分離する攻撃の適用が不可能であり、安全な撮影を行うことができます。その方式が 100%の精度で再撮影を検出できることが評価実験において実証されました。

【今後の期待】

本研究で確立した、Dual-Pixel センサーのみを利用した再撮影検出方式は、ソーシャルメディアや報道をはじめとする画像や動画の真正性を確保する基盤となります。Dual-Pixel センサーはスマートフォンやデジタルカメラに広く使用されているため、この方式が普及することによって、Deepfake などによる誤情報の拡散を防ぐことに繋がることが期待されます。

（論文情報）

タイトル：Breaking Infrared Recapture Detection: Optical-Synthesis Attacks and Depth-Aware In-Sensor Countermeasures

著者：Tetsu Ishizue, Sara Rampazzi, Takeshi Sugawara

掲載誌：USENIX WOOT Conference 2026

（外部資金情報）

科学技術振興機構（JST）戦略的創造研究推進事業 CREST「基礎理論とシステム基盤技術の融合による Society 5.0 のための基盤ソフトウェアの創出：AI 駆動型サイバーフィジカルシステムのセキュリティ評価・対策基盤」（JPMJCR23M4）

（用語説明）

※1：来歴情報

画像や動画などのコンテンツが、いつ、だれによって、どのように作成され、どのように編集されたかという、コンテンツの出所に関する情報。

※2：Deepfake

AI によって生成された偽の画像や動画。

※3：再撮影攻撃

モニターをカメラで撮影する行為。AI によって生成された Deepfake を本物のように装うために行われる。

※4 : C2PA

来歴情報を記録・証明するための技術標準。C2PA に準拠するカメラはカメラ内の秘密鍵でデジタル署名を作成することによって、その画像や動画がカメラで撮られたものであることを証明する。

※5 : LiDAR

レーザー光を照射して、その反射した光の情報から距離を計測するセンサー。

※6 : ホットミラー

可視光を透過し、赤外光を反射する特殊なフィルター。

※7 : Dual-Pixel センサー

効率的なオートフォーカスのために開発された、各画素が2つに分割されている特殊なイメージセンサー。

【連絡先】

＜研究内容に関すること＞

電気通信大学 情報理工学研究科 情報学専攻

【職名】 准教授

【氏名】 菅原 健

Tel : 042-443-5896 E-Mail : sugawara@uec.ac.jp

＜報道に関すること＞

電気通信大学総務部総務企画課広報係

Tel : 042-443-5019 Fax : 042-443-5887

E-Mail : kouhou-k@office.uec.ac.jp