

XR HMD の外向きカメラを介した周辺文字情報の漏洩リスクの評価

倉崎 翔大^{1,a)} 清水 滯¹ 秋山 晃誠¹ 藤田 真由¹ 大塚 航世¹ 金岡 晃^{1,b)}

概要：VR/AR/MR (XR) HMD の利用拡大に伴い、XR HMD に搭載されたカメラやセンサを介したプライバシー侵害や情報漏洩のリスクが増大している。センサ情報の漏洩によるリスクの議論は進んでいる一方で、XR HMD の外向き RGB カメラを介した周辺情報の漏洩リスクについては十分に議論されていない。本研究では、周辺情報の中でも文字情報に着目し、その漏洩リスクを明らかにすることを目的とした。実験では、Meta Quest 3 を用いて複数の距離・角度から文字を印刷した紙を撮影し、得られた画像を OCR 処理した。その結果の分析を通して文字情報漏洩の可能性を評価すると共に、撮影条件の違いや歪み補正の有無が文字情報漏洩リスクに与える影響についても考察した。

キーワード：XR, VR, AR, プライバシー侵害, 情報漏洩

Evaluation of Ambient Text Leakage Risk via Outward-Facing Cameras on XR HMDs

SHODAI KURASAKI^{1,a)} REI SHIMIZU¹ KOUSEI AKIYAMA¹ MAYU FUJITA¹ KOUSEI OTSUKA¹
AKIRA KANAOKA^{1,b)}

Abstract: As VR/AR/MR (XR) HMDs become widespread, the risks of privacy invasion and information leakage via their integrated cameras and sensors are increasing. While risks associated with sensor data have been investigated, the leakage risk of environmental information through outward facing RGB cameras remains underexplored. This study focuses on textual information in the user's surroundings and aims to clarify its leakage risk. In our experiment, we used a Meta Quest 3 to capture images of printed text at multiple distances and viewing angles, and then processed the images with OCR. By analyzing the OCR outputs, we assess the potential for text leakage and discuss how capture conditions and lens distortion correction affect the leakage risk.

Keywords: XR, VR, AR, Privacy-Invasion, Information-Leakage

1. はじめに

Virtual Reality (VR) /Augmented Reality (AR)/ Mixed Reality (MR) を中心とする Cross Reality (XR) 技術は、ヘッドマウントディスプレイ (Head-Mounted-Display, HMD) を通じた利用が拡大している。しかし、XR HMD

は RGB カメラを含む様々なセンサを搭載しており、利用拡大に伴いそれらを介したプライバシー侵害や情報漏洩の脅威やリスクも増大している。これらの脅威やリスクについての議論や調査は進んでいる一方で、XR HMD の外向き RGB カメラを介した周辺情報の漏洩リスクについては、議論および調査が不十分である。

PC やスマートフォンを始めとする従来デバイスに搭載されている RGB カメラによるプライバシー侵害や情報漏洩についての研究は存在するが、XR HMD の RGB カメラには以下の3つの特性があることから、XR 環境には従来

¹ 東邦大学

Toho University

^{a)} 7525001k@st.toho-u.ac.jp

^{b)} akira.kanaoka@is.sci.toho-u.ac.jp

のコンピューティング環境と異なる脅威やリスクが存在する可能性が考えられる。まず1つ目の特性は、デバイスがユーザの頭部に装着される点である。これによりXR HMD搭載のRGBカメラは、位置や角度が従来のデバイスと比較して変わりやすく、かつより広範な周辺環境がカメラの画角に収まる可能性のあるエリアとなる。更に、人間は注目している情報がある方に頭を向ける傾向にあるため[6]、重要な情報がXR HMDのカメラに映る可能性が高い。加えて、XR HMDユーザの知覚や行動を攻撃者が操ってしまう攻撃も存在し[7], [8], [9]、これにより利用者が攻撃者の意図する方向へ誘導され、結果として攻撃者の求める場所の画像を取得される可能性がある。2つ目は、XR HMDには複数のカメラが搭載されており、とりわけ外向きのRGBカメラは両眼構成で複数設置されることが多い点である。これにより、それぞれのカメラで取得した画像や映像を補完しあえる可能性がある。例えば、1つのカメラからは光が反射するなどして見えなかった場所が、もう1つのカメラからは見える場合である。3つ目は、外向きの両眼用RGBカメラが周囲全体をカバーするため、広角あるいは超広角レンズが採用されることが多い点である。広角レンズを通して画像や映像を取得すると、画角の端側は歪んでうつってしまい、中心部分と端部分では情報の読み取りやすさに差がある可能性が考えられる。また、歪み補正前の画像と補正後の画像を比較したとき、情報の読み取り精度に差が出る可能性が考えられる。

本研究ではこれらの特徴を踏まえ、XR HMDの外向きカメラを介した周辺情報漏洩リスクの議論と評価を行うこととした。周辺情報の含まれる画像や映像からは、文字情報や位置情報、所有物情報、物の配置情報など、様々な情報が漏洩する可能性が考えられるが、文字情報はより多くのプライバシー関連情報や機密情報を含むと考え、本稿では文字情報に焦点を当て、3つのResearch Question (RQ)を設けた。

RQ1: XR HMDの前面RGBカメラを介して取得した画像情報から、文字情報はどの程度の精度で取得されてしまうのか

RQ2: 文字認識精度は、撮影条件でどのように変わるのか

RQ3: 歪み補正の有無により、文字認識精度は変わるのか

これらのRQに答えるためには、実際のXR HMDを介して撮影した画像に基づき、文字認識精度を実験的に測定することが有効である。本研究では、そのために代表的なXR HMDであるMeta Quest 3を用い、HMDに搭載された2つのRGBカメラで文字列の印刷された紙を撮影し、そこで得られた画像に対して文字認識処理をかけた。文字認識処理の結果を基にCharacter Error Rate (CER)を算出し、これを文字認識精度の評価に用いた。また、RQ2に回答するため、撮影は複数の距離および角度から行った。

加えて、RQ3のため、取得した画像に対して歪み補正を文字認識処理の前に行った場合と行わなかった場合の結果を比較した。

その結果、XR HMDの前面RGBカメラを介して取得した画像から高精度で文字情報が取得されることが明らかとなり、XR HMDの外向きRGBカメラを介した周辺文字情報の漏洩リスクは無視できないものであることが示された。また、撮影角度が大きくなると文字認識の精度が下がる傾向にある可能性、歪み補正により文字認識精度が必ずしも良くなるとは限らないことが示唆され、歪み補正後の画像等だけでなく歪み補正前の画像等の漏洩を防ぐことも重要であることが明らかとなった。

2. 関連研究

本研究は、XR HMDの外向きカメラを介した周辺文字情報の漏洩リスクを対象とする。XRではHMDが頭部に装着され、視線や頭部運動と連動して外向きカメラが環境を捉えるため、従来の端末カメラと異なる特徴的なリスクが生じうる。特に、利用者自身の注意の向きや、知覚操作を介した頭部方向の誘導が撮像に影響し得る点が重要である。

2.1 XR プライバシ

XRのセキュリティ・プライバシー研究は、早期から第三者(bystander)や周辺環境に対する配慮を強調してきた。DenningらはARグラスの現場観察を通じ、周辺者の録画懸念とそれを緩和する設計要件を示した[1]。RoesnerらはARシステムのセキュリティ/プライバシー課題を体系化し[2]、Lebeckらは出力経路の保護[3]やマルチユーザARの基盤[4]を提示した。Ruthらはマルチユーザ環境での安全なコンテンツ共有を実装・評価し[5]、周辺者・共同利用・共有出力というXR特有の文脈における保護要件を具体化している。さらに、知覚操作(PMA)や行動誘導に関する近年の研究は、視線・頭部方向が攻撃的に操作され得ることを示し、結果として外向きカメラが機微情報のある方向へ向けられる可能性を高める[7], [8], [11]。

2.2 従来のカメラプライバシー研究

従来機器のカメラでも、撮像からのセンシティブ情報漏洩は広く議論されてきた。代表例として、スマートフォン等の背面カメラを悪用して屋内の映像から空間情報や機微情報を窃取するPlaceRaider攻撃がある[10]。また、高解像度写真に写り込んだ手指等から指紋パターンを復元し、生体認証の破りにつながるリスクが指摘されている[12]。これらはXR固有の技術に依らず成立する攻撃だが、XRでは撮像の視点がユーザの頭部運動と強く結びつき、かつ常時的・広角的な撮像が起りやすいことから、リスクの

顕在化頻度や被害範囲が拡大し得る点に特徴がある。

3. 脅威モデル

本研究で想定する攻撃者は、複数存在する。

1つ目は、ゼロデイやプラットフォーム脆弱性の悪用者ではなく、Unity Asset Store、Unreal Engine Marketplaceといった拡張リソースのマーケットを経由して無害に見えるアセットやスクリプトを配布する供給者とした。現時点ではそういった攻撃の報告は著者らの知る限りでは無いものの、Google Chrome や VSCode で類似事例が報告されている [13], [14], [15], [16]。開発者や開発組織が当該リソースをプロジェクトに組み込むことで、攻撃者のコードがアプリ権限の範囲で実行される。Unity では、スクリプトを含む Prefab やアセットパッケージの導入だけで多機能オブジェクトが動作し得るという供給形態が、サプライチェーン由来の攻撃面を生む。

このような攻撃者の能力は、公開 API の利用とデータや表示状態の操作に関して自由度が高い。Quest においては、パススルー API が提供されており、その API を通じて外向きカメラの画像が取得可能である。両眼でないことやピクセル数が制限されているなどカメラ本来の性能をフルに反映した画像とはなっていないが、正規の API からの取得が可能である。

2つ目は、非公開低レイヤ API や、デバイス/OS 提供会社のアプリで用いられる内部 API を用いる攻撃者である。こういった低レイヤ API や内部 API の存在はいくつか知られており [17], [18]、これらが悪用されるシナリオは理論上排除できない。非公開低レイヤ API が利用可能な場合、カメラへの直接アクセス等の可能性が上がり、より精度の高い画像が取得できる可能性がある。

3つ目は、root 化された端末の悪用である。悪意のある攻撃者が root 化した端末を、オークションサイト等で配布あるいは販売し、その購入者をターゲットに攻撃をするモデルも考えられる。この場合、上記 2つよりさらに直接的なカメラ制御が可能になり、もっとも精度の高い画像データが取得されうる。

これらの攻撃者は、経路は異なるがそれぞれ外カメラの画像を取得可能である。本研究ではこういった多様な攻撃者モデルが取りうる環境を加味し、直接的にカメラ制御ができもっとも高い画像データが取得できることを前提とした。

4. 実験手法

4.1 撮影環境

Quest 3 で RGB カメラ画像を取得する方法には複数の選択肢がある。ミラーリングを用いた場合は補正済みの片眼映像が得られるが、視野が狭く解像度も限定的である。



図 1 Quest3 の外向け RGB カメラによる画像取得実験：実験環境

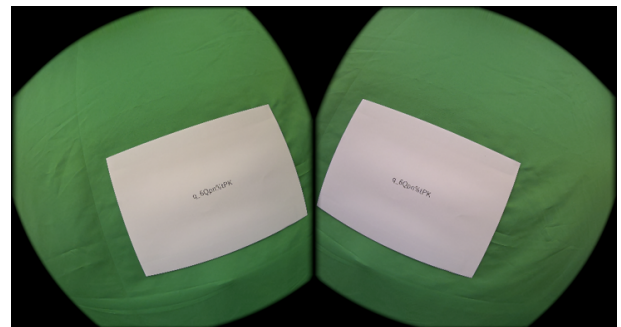


図 2 Quest3 の外向け RGB カメラによる画像取得実験：撮影画像例

パススルー API による取得も可能であるが、同様に補正が施された狭い映像に限られる。さらに、RGB カメラが直接生成した未補正の画像を正規の手段で取得できる API は公開されていない。このため本研究では、3 章で議論した攻撃者モデルも踏まえ、両眼の広角映像を歪み補正なしで取得できる Meta Quest Developer Hub のスクリーンショット機能を採用した。

Meta Quest 3 は、撮影距離や角度を正確に制御するため、電子的に角度調整が可能な台 (M-6 REMOTE & TRACKING PANTILT) に固定し、さらにその台を三脚に設置した。撮影条件に合わせた位置および角度に設置した。Quest 3 は専用アタッチメントを介して雲台に取り付け、測定条件を再現性高く設定できるようにした。

評価対象となる文字列は紙に印刷し、これを撮影対象とした。距離や角度を変化させる際に安定して設置できるよう、紙はホワイトボードに固定した。ただし紙自体のシワや背景の余計な要素は OCR 処理で誤認識を生じさせる恐れがあるため、文字以外の情報が認識されないよう、背景にはグリーンバックを施した (図 1、2)。これにより、文字列そのものに対する認識精度を適切に評価できるようにした。

4.2 撮影条件、評価対象文字列

撮影条件は、Meta Quest 3 と文字列の印刷された紙の距離を 2 パターン (25cm, 50cm)、文字列の印刷された紙に

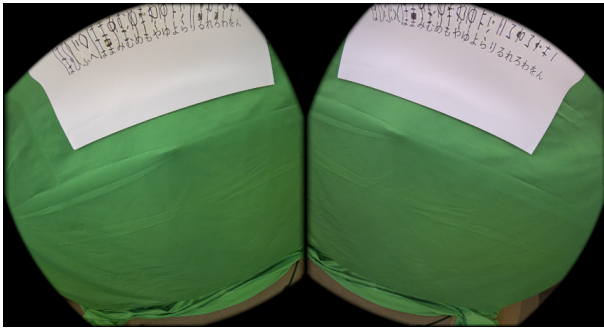


図 3 Quest3 の広角カメラにより周辺部分に強い歪みが出ている撮影画像例

対する Meta Quest 3 の角度を垂直方向に 5 パターン (0° , 15° , 22.5° , 30° , 45°) として設定し、全組み合わせの 10 パターンで撮影を行った。距離の設定については、XR HMD の利用場面において 1~2m 先の対象が現実的な読み取りターゲットとなる一方、攻撃者にとって有利な至近距離での読み取り可能性も検討する必要があると考え、近距離 (25cm) と中距離 (50cm) を選択した。角度の設定については、Quest3 の外向き RGB カメラが超広角レンズを用いており、視野中心から離れるほど歪みが大きくなることを考慮した。歪みは補正されるものの、周辺部では画素が粗くなることで認識精度が低下する可能性があるため、視野中心からの角度を段階的に変化させる条件を設定した。図 3 は、周辺部で強い歪みが発生し、文字自体の読み取りが難しい撮影画像の例である。これらの角度は、Quest3 の視野角に基づき、他の HMD とも比較して特異ではない範囲から選定した。

紙に印刷する文字列には、アラビア数字、英字アルファベットの大文字と小文字、ひらがな、カタカナ、パスワード、住所を準備した。これらは、基本的な文字種を網羅するとともに、実際に利用場面で読み取られる可能性の高い情報を含めることで、文字認識精度を多角的に評価できるようにするために選択した。パスワードの文字列としては、NISC が公開しているハンドブックで安全とされている形式に基づき、英字アルファベットの大文字と小文字、数字、記号を組み合わせたランダムな 10 桁の文字列を 10 個用意した。住所の文字列については、関東 1 都 6 県に加え、北海道、大阪府、京都府を含めた 10 か所の都道府県庁舎の住所とした。なお、紙は A4 サイズで、フォントは游ゴシック、フォントサイズは 30 に設定した。各文字列に割り振った ID、文字列のジャンル、文字列自体を、表 1 に示す。

4.3 文字認識精度の計測

最初に、10 パターンの条件で 25 パターンの文字列の印刷された紙を撮影した合計 250 枚の画像を分析のため、左右に分割して 500 枚の画像にし、これを歪み補正なしの画

表 1 撮影実験の評価対象文字列のジャンルとその文字列
ジャンル 正解データの文字列

アラビア数字	0123456789
英小文字	abcdefghijklmnopqrstuvwxyz
英大文字	ABCDEFGHIJKLMNOPQRSTUVWXYZXYZ
ひらがな	あいうえおかきくけこさしすせそたちつてとなにぬねのはひふへほまみむめもやゆよりるれろわをん
カタカナ	アイウエオカキクケコサシスセソタチツテトナニヌネノハヒフヘホマミムメモヤユヨラリルレロワラン
パスワード	4Kjcas/7.8
パスワード	q.6Qpn%tPK
パスワード	3nf&6hk#nZ
パスワード	z*YQ2L—Z)
パスワード	Qm+A(HRz-3
パスワード	et2GwkQ\$27
パスワード	YpNR,4jJFg
パスワード	&f]u9d@Ea2
パスワード	;?A-i WS5p
パスワード	{G.J0"m}e=
住所	〒310-8555 茨城県水戸市笠原町978-6
住所	〒320-8501 栃木県宇都宮市塙田1-1-20
住所	〒371-8570 群馬県前橋市大手町1-1-1
住所	〒330-9301 埼玉県さいたま市浦和区高砂3-15-1
住所	〒260-8667 千葉県千葉市中央区市場町1-1
住所	〒163-8001 東京都新宿区西新宿2-8-1
住所	〒231-8588 神奈川県横浜市中区日本大通1
住所	〒060-8588 北海道札幌市中央区北三条西6丁目
住所	〒540-8570 大阪府大阪市中央区大手町2丁目
住所	〒602-8570 京都府京都市上京区下立売通新町西入藪ノ内町

像群とした。次に、歪み補正の有無による文字認識精度の違いを確認するため、歪み補正なしの画像群に対して歪み補正を行うことで歪み補正ありの画像 500 枚を準備した。歪み補正は、OpenCV のカメラキャリブレーション API を用いて、内角 7×10 のチェスボード画像を様々な角度から撮影した画像から対応点を抽出した後、左右のカメラそれぞれの内部行列と歪み係数を推定することで行った。

表 2 各角度における最良の CER：歪み補正なし

文字列\撮影角度	0°	15°	22.5°	30°	45°
アラビア数字	0.100	0.100	0.100	0.100	0.900
英字小文字	0.038	0.038	0.038	0.346	0.230
英字大文字	0.038	0.153	0.115	0.115	0.346
ひらがな	0.021	0.042	0.106	0.319	0.531
カタカナ	0.063	0.085	0.148	0.361	0.595
パスワード (平均)	0.550	0.180	0.350	0.530	0.790
住所 (平均)	0.556	0.700	0.646	0.615	0.942

表 3 各角度における最良の CER：歪み補正あり

文字列\撮影角度	0°	15°	22.5°	30°	45°
アラビア数字	0.800	0.700	0.800	0.900	0.900
英字小文字	0.423	0.730	0.500	0.576	0.192
英字大文字	0.500	0.269	0.307	0.076	0.576
ひらがな	0.021	0.021	0.063	0.042	0.659
カタカナ	0.063	0.063	0.063	0.085	0.659
パスワード (平均)	0.830	0.850	0.780	0.550	0.860
住所 (平均)	0.558	0.541	0.567	0.526	0.942

歪み補正なしの画像群と歪み補正ありの画像群の計 1000 枚の画像すべてに対して、文字認識精度を高めるための前処理として、文字の印刷された紙部分の抽出、グレースケール化、固定閾値 128 による二値化、ガウシアンフィルタを用いたノイズ除去を行った。その後、文字列が斜めに傾いて映っている場合を考慮し、画像を 1 度ずつ、360 度回転させながら文字認識を行い、その結果から CER を算出した。文字認識には、オープンソースの OCR エンジンである Tesseract と、Tesseract を Python で扱うためのライブラリである pytesseract を用いた。また、各画像に対する CER の結果は、各画像ごとに 360 度で最も良い CER とした。

5. 結果

撮影角度の違いによる文字認識精度への影響を見るため、各角度における最良の CER を正解の文字列ごとに表にまとめた。ただし、パスワードと住所については、それぞれ 10 個の文字列ごとの最良 CER の平均値とした。歪み補正なしの場合を表 2、歪み補正ありの場合を表 3 に示す。これらの表から、撮影角度が大きければ大きいほど CER が悪くなる傾向がみられる。更に、文字列ごとにも CER に差がみられ、特にパスワードと住所については他と比較して悪い値となった。

また各撮影角度において、撮影距離が 25cm の場合と 50cm の場合の最良 CER を文字列ごとに比較し、25cm の方が良かった文字列数、50cm の方が良かった文字列数、同じ CER だった文字列数も表にまとめた。歪み補正なしの賀愛を表 4、歪み補正ありの場合を表 5 に示す。これらの 2 つの表より、0 度から 22.5 度においては撮影距離が遠いほど CER は悪い傾向にあることがわかる。歪み補正ありの

表 4 撮影距離ごとの最良 CER 比較：歪み補正なし

撮影距離\撮影角度	0°	15°	22.5°	30°	45°
25cm < 50cm	11	16	15	10	1
25cm = 50cm	8	2	2	3	14
25cm > 50cm	6	7	8	12	10

表 5 撮影距離ごとの最良 CER 比較：歪み補正あり

撮影距離\撮影角度	0°	15°	22.5°	30°	45°
25cm < 50cm	15	14	15	18	2
25cm = 50cm	7	6	7	3	15
25cm > 50cm	3	5	3	4	8

場合においては 30 度においても同様の傾向がみられたが、歪み補正なしの場合では撮影距離の近いほうが CER が悪い傾向がみられる。撮影角度 45 度では撮影距離が 25cm の場合の CER が悪い傾向がつゆ億出ているが、これは撮影時に評価対象の文字列が画角からはみ出て映っていないためである。また、全体的に歪み補正を行う場合は 50cm よりも 25cm の方が良い CER を出す傾向が強く見られる。

6. 考察

6.1 撮影条件による差

本実験において、撮影角度や撮影距離が大きくなると文字認識精度が下がる傾向がみられたことから、実験環境でない一般的な環境においても、撮影角度や撮影距離は小さいほうが文字認識精度は高いと考えられる。しかし、歪み補正ありの撮影角度 30 度の「英語大文字」「ひらがな」「カタカナ」の画像に対する CER が 0.100 よりも 0 に近く良いことから、撮影角度が大きくてもカメラの画角内にある文字情報は十分に漏洩しうる可能性を示唆する結果であったともいえよう。

6.2 歪み補正の有無による差

本実験において、撮影角度 22.5 度と 30 度において、日本語の含まれる「ひらがな」「カタカナ」「住所」の文字列に対する CER が、歪み補正により改善された。これは、日本語の文字認識は英語の文字認識と比べ、カメラの歪みの影響を大きく受ける可能性を示唆した結果であると言える。

一方で、アラビア数字や英語小文字に対する CER は、歪み補正により悪化した。この結果から、必ずしも歪み補正を行うことが文字認識精度を上げることに繋がるわけではないことが考えられる。更にこの可能性は、歪み補正後の画像や映像が漏洩した場合よりも、歪み補正前の画像や映像が漏洩した場合の方が、文字情報漏洩のリスクが高いことを示唆しているとも考えられる。

7. まとめ

本研究では、XR HMD の外向き RGB カメラを介した周

謝辞 本研究は、JST、CREST、JPMJCR22M4 の支援を受けたものである。