

デュアルカメラスマートフォンを用いた動画像中の不可視QRコードの頑健な認識

覚井 優希、荒木 航太、韓 燦教、福嶋 政期、苗村 健

あらまし 画面とカメラの通信により、画面上の情報を瞬時に取得することができる。不可視のスクリーンカメラ通信は、映像にデータを目立たないように埋め込むものであり、視聴体験を妨げないため、大きな可能性を秘めている。知覚できない色振動法は、各画素に対して同じ輝度の2つの色を交互に表示することでこれを実現する。しかし、元のビデオコンテンツのフレーム間差により、デコード性能が低下する可能性がある。そこで、我々は、変調値の異なる2つの画像を同時に撮影するデュアルカメラスマートフォンを用いた新しいデコーディングアプローチを提案する。この方法により、時間的に近い画像間の色の違いを計算することができ、元の内容の時間的変化によるアーチファクトを低減することができる。我々の実験結果は、従来手法と比較して、復号化率が向上し、認識時間が短縮されることを示している。

キーワード: 不可視の色振動、ディスプレイ-カメラ間通信、可視光通信

1. まえがき

可視光通信(VLC)は、特に近接した機器間の通信において、無線周波数(RF)通信を補完する技術として機能する[1]。RF通信と比較して、VLCは特徴的な特性を有している。まず、可視光の周波数が高いため、VLCは高い指向性を示し、その結果、受信範囲が制限され、干渉が最小限に抑えられる。例えば、可視光線は壁などの障害物を透過することができない。第二に、発光ダイオードやディスプレイを含む既存の照明器具をVLCトランスミッターとして利用できる。したがって、VLCは、これらの特性により、追加のハードウェアを必要としないため、低コストで実装することができる。

VLCのサブセットであるスクリーン・カメラ通信は、ディスプレイやスマートフォンの普及に伴い、脚光を浴びている。画面上の情報を瞬時に検索することができる。QR(Quick RespoSnse)コード[2]のような可視2次元(2D)バーコードは一般的に使用されているが、人間が認識することが困難な記号パターンで構成されており、

コンテンツ表示領域の一部を占めることでビデオ視聴体験を混乱させる可能性がある。

あるいは、見えないスクリーンとカメラのコミュニケーションに焦点を当てた研究も行われており、ここでは、データは人間の目に見えない方法でディスプレイの中に埋め込まれている。このアプローチは、ビデオ視聴体験を維持しつつ、同時にサイドチャンネルにデータを送信する。阿部ら[3]が提案した知覚できない色振動法では、輝度は同じだが色度が画素ごとに異なる2つの変調色を高速に交互に行うことで、知覚できない色振動を実現している。この方法は、急激な色度変化に対する人間の目の鈍感さを利用し、標準的なディスプレイやスマートフォンを使って、見えない情報をリアルタイムで伝送することができる。しかし、激しい動きを含む動画から情報を抽出することは困難である。デコーディングは、元のビデオコンテンツの時間的変化によって損なわれる。なぜなら、埋め込まれた情報は、時間的に離れたフレーム間の差を計算することによって取得されるからである。

この課題を解決するために、我々は、図1に示すように、デュアルカメラスマートフォンを用いて、変調値の異なる2つの画像を同時に撮影する復号化手法を提案する。本手法は、時間的に隣接するフレーム間の色差を捉えることで、元の映像コンテンツの時間的変化によるアーチファクトを軽減し、ロバストな復号化を実現する。我々は以前、この概念を導入し、いくつかの応用を実証した[4]。

Received June 6, 2024; Revised September 25, 2024; Accepted November 12, 2024

† The University of Tokyo *Corresponding Author: Changyo Han (Tokyo, Japan)

This paper includes videos. Note that the videos are not viewable from this PDF file. The videos are available as separate files on the website that hosts this PDF file.

【機械翻訳コンテンツの著作権について】

当サイトに掲載されている原著論文の著作権は映像情報メディア学会に帰属します。原著論文および翻訳論文の無断使用は禁止します。
引用の際には、必ず原著論文の書誌情報をご記載ください。詳細は映像情報メディア学会著作権規定をご覧ください。

Robust Recognition of Imperceptible QR Codes in Videos with a Dual-Camera Smartphone

Yuki Kakui [†], Kota Araki [†], Changyo Han ^{†*},
Shogo Fukushima [†], Takeshi Naemura (member) [†]

Abstract Screen-camera communication enables instantaneous retrieval of on-screen information. Imperceptible screen-camera communication, embedding data in videos unnoticeably, holds great promise as it does not interfere with the viewing experience. The imperceptible color vibration method achieves this by alternately displaying two colors with the same luminance for each pixel. However, decoding performance may deteriorate due to interframe differences in the original video content. To address this, we propose a novel decoding approach using a dual-camera smartphone to capture two images with different modulation values simultaneously. This method allows for computing color differences between temporally close images, reducing artifacts from temporal changes in the original content. Our experimental results demonstrate an improved decoding rate and decreased recognition time compared to the previous method.

Key words: Imperceptible color vibration, screen-camera communication, visible-light communication

1. Introduction

Visible light communication (VLC) serves as a complementary technique to radio frequency (RF) communication, particularly in communication between devices in close proximity [1]. Compared to RF communication, VLC possesses distinctive characteristics. Firstly, due to the high frequency of visible light, VLC exhibits highly directional properties, resulting in a limited reception range and minimal interference. For instance, visible light cannot penetrate obstacles such as walls. Secondly, existing lighting equipment, including light-emitting diodes and displays, can be utilized as VLC transmitters. Hence, VLC can be implemented at a low cost because these properties eliminate the need for additional hardware.

Screen-camera communication, a subset of VLC, has gained prominence with the widespread use of displays and smartphones. It enables the instantaneous retrieval of on-screen information. Although visible two-dimensional (2D) barcodes such as quick response (QR) codes [2] are commonly used, they consist of symbolic patterns that are difficult for humans to recognize and

can disrupt the video-viewing experience by occupying portions of the content display area.

Alternatively, research has focused on invisible screen-camera communication, where data are embedded within the display in a manner invisible to the human eye. This approach maintains the video-viewing experience while simultaneously transmitting data to a side channel. In the imperceptible color vibration method proposed by Abe et al. [3], imperceptibility is achieved by rapidly alternating between two modulated colors of identical luminance but differing chromaticity for each pixel. This method exploits the insensitivity of the human eye to rapid chromaticity changes and allows real-time transmission of invisible information using standard displays and smartphones. However, extracting information from videos containing intense movements poses a challenge. Decoding is impaired by temporal changes in the original video content because the embedded information is retrieved by calculating the difference between temporally distant frames.

To address this challenge, we propose a decoding method that simultaneously captures two images with different modulation values using a dual-camera smartphone, as shown in Fig. 1. Our approach mitigates artifacts caused by temporal changes in the original video content by capturing the color difference between temporally adjacent frames, resulting in robust decoding. We previously introduced this concept and demonstrated its several applications [4]. Building on this,

Received June 6, 2024; Revised September 25, 2024; Accepted November 12, 2024

[†] The University of Tokyo *Corresponding Author: Changyo Han (Tokyo, Japan)

This paper includes videos. Note that the videos are not viewable from this PDF file. The videos are available as separate files on the website that hosts this PDF file.

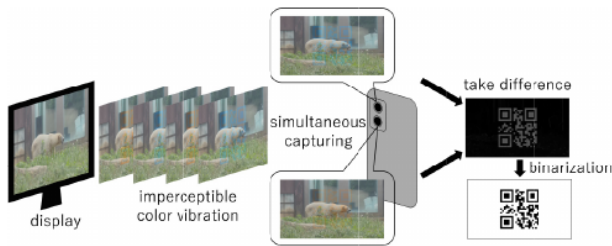


Fig. 1 変調値の異なる2つの画像を同時にキャプチャする。

また、捕捉率と露光時間に関する追加実験を行い、知覚できないQRコードを検索するデュアルカメラのセットアップをより深く分析し、より広い意味合いと最近の開発に関する拡張的な議論を提供する。本研究では、デコーディング性能を向上させるための最適なキャプチャパラメータを経験的に決定し、主にデコーディング処理中に得られた色差に基づく結果について議論する。さらに、提案手法の復号率と認識時間を、阿部ら[3]が提案した先行手法と比較した。

本研究の主な貢献は以下のようにまとめられる：

- デュアルカメラスマートフォンを用いた同時撮影技術を提案し、ロバストなデコーディングを実現することで、元の映像コンテンツのフレーム間差によるアーチファクトを低減する。
- 実験により、適切なキャプチャパラメータを決定し、復号率と認識時間を測定することで、提案システムを評価した。

2. 関連研究

2.1 スクリーン・カメラ通信

スクリーンとカメラの通信の分野では、実質的な情報を埋め込むことができ、エラーロバストな抽出を可能にするコードを設計することに研究が集中している。一般的に使用される2Dコードには、QRコード[2]とデータマトリックス[5]がある。さらに、COBRA[6]やRainbar[7]などの大容量カラーバーコードも提案されている。RescQR[8]は、専用フレーム境界とビタビベースのデータ復元方式を用いることで、QRコードを確実に復元するスクリーンカメラ通信システムを提案した。しかし、これらのコードは、人間の目には直接解釈できない記号的なパターンで構成されており、表示されたコンテンツの視覚的な体験を妨げる可能性がある。さらに、これらの要素はコンテンツ表示領域を空間的または時間的に占有し、動画視聴体験をさらに妨げる。

コード設計アプローチの一つは、通常画像に似た人間が解釈可能な外観を持つコードをレンダリングすることを目的としている。QRコードと画像を組み合わせることでコードを生成した研究もある[9, 10]。生成されたコードは、既存のQRコードリーダーに読みやすいという利点があるが、QRコードのファインダーパターンやその他の特徴を保持しているため、外観が損なわれている。さらに、コード設計のためのいくつかの研究では、画像上に0/1を表すモジュールを重ねたり[11-13]、カラードット行列として表現される画像内のドットの色を変更したり[14]する。これらの方法は、コードを邪魔にならないようにするが、それでも変調された画像を劣化させる。

2.2 見えないスクリーンとカメラのコミュニケーション

一方、もう一つのアプローチは、スクリーンとカメラの通信のために、視覚的に感知できないコードを埋め込むことを提案する。見えないスクリーンとカメラのコミュニケーションとは、人間の目には知覚できない方法でコードを埋め込むことである。この点に関して、Visual MIMO [15]は、画像内の低周波成分の値を変更することで情報を埋め込む。HiLight[16]とUber-in-Light[17]は、それぞれ α チャンネルと赤と青のチャンネルを利用し、周波数シフトキーイングを用いて情報を送信する。これらの方法では、画素値の変化を補正する機構が十分に備えられていないため、知覚できない状態を維持するために変調を十分に小さく保つ必要がある。そのため、環境中のノイズの影響を受けやすく、デコード精度の低下につながる懸念される。

(1) 輝度変調

注目すべきは、人間の目の時間分解能を利用した研究で、輝度の急激な変化の明滅を知覚するのではなく、平均的な明るさを知覚する研究があることである。境界周波数は臨界フリッカー周波数(CF)と呼ばれ、約40~50Hzである[18]。これらの特性により、知覚できない状態を維持したまま、より大きな変調が可能となる。InFrame [19, 20]は、正のデータフレームと負のデータフレームを交互に高速(120FPS)に表示することで、見えないスクリーン・カメラ通信を実現している。さらに、TextureCode [21]は、フリッカーが知覚されにくい複雑なテクスチャを持つ領域にのみ選択的に情報を埋め込む。さらに、ChromaCode [22]は、埋め込みに一様な知覚色空間の明度成分を採用し、テクスチャの複雑さを考慮して明度変化量を調整する。

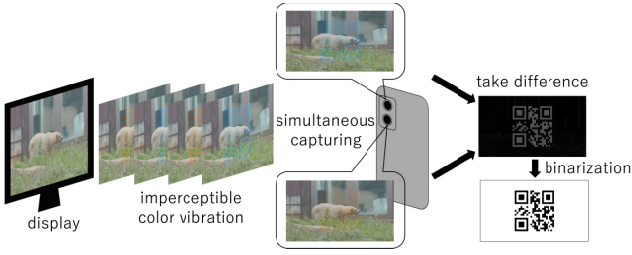


Fig. 1 Simultaneous capture of two images with different modulation values.

we include additional experiments on capture rates and exposure times, offering deeper analysis of the dual-camera setup for retrieving imperceptible QR codes, and providing extended discussions on broader implications and recent developments. This study empirically determines the optimal capture parameters to enhance the decoding performance and primarily discusses the results based on the chromatic differences obtained during the decoding process. Furthermore, we compared the decoding rate and recognition time of the proposed method with those of the previous method proposed by Abe et al. [3].

The main contributions of this study are summarized as follows:

- We propose a simultaneous capturing technique employing a dual-camera smartphone to achieve robust decoding, thereby reducing artifacts caused by inter-frame differences in the original video content.
- Through experiments, we determined appropriate capture parameters and evaluated the proposed system by measuring the decoding rate and recognition time.

2. Related Work

2.1 Screen–Camera Communication

In the field of screen–camera communication, research has focused on designing codes capable of embedding substantial information and enabling error-robust extraction. Typically used 2D codes include QR codes [2] and data matrices [5]. Moreover, high-capacity color barcodes have also been proposed, such as COBRA [6] and Rainbar [7]. RescQR [8] proposed a screen–camera communication system that reliably recover QR codes by using a dedicated frame border and a Viterbi-based data recovery scheme. However, these codes comprise symbolic patterns that are not directly interpretable by human eyes and can interfere with the visual experience of the displayed content. Furthermore, they occupy the content display area, either spatially or temporally, further disrupting the video-

viewing experience.

One of the code design approaches aims to render codes with a human-interpretable appearance similar to normal images. Certain studies have generated codes by combining QR codes and images [9, 10]. Although the generated codes possess the advantage of being readable by existing QR code readers, they retain the finder patterns and other features of QR codes, compromising their appearance. Moreover, some studies for code design involve superimposing a module representing 0/1 on the image [11–13] or altering the color of the dots within an image represented as a color dot matrix [14]. Although these methods render the code less obtrusive, they still degrade the modulated image.

2.2 Invisible Screen–Camera Communication

In contrast, the other approach suggests embedding visually imperceptible codes for screen–camera communication. Invisible screen–camera communication refers to the embedding of codes in a manner imperceptible to the human eye. In this regard, Visual MIMO [15] embeds information by changing the values of low-frequency components within an image. HiLight [16] and Uber-in-Light [17] utilize alpha channel, and red and blue channels, respectively, to transmit information using frequency-shift keying. In these methods, the modulation must remain sufficiently small to preserve imperceptibility because the mechanisms for compensating for changes in pixel values are not well-equipped. Therefore, there is concern that they are susceptible to noise in the environment, leading to a decrease in decoding accuracy.

(1) Luminance Modulation

Notably, certain studies exploit the temporal resolution of the human eye, which does not perceive flickers of rapid luminance changes but rather average brightness. The boundary frequency is called the critical flicker frequency (CFF) and is approximately 40–50 Hz [18]. These characteristics allow greater modulation while maintaining imperceptibility. InFrame [19, 20] achieves invisible screen–camera communication by alternately displaying an original video frame with a positive data frame and a negative data frame at high speed (120 FPS). Moreover, TextureCode [21] selectively embeds information only in regions with complex textures where flickers are less likely to be perceived. Furthermore, ChromaCode [22] employs the lightness component of the uniform perceptual color space for embedding and adjusts the amount of lightness change to ac-

これらの方法は、40～50Hz以上の振動を発生させるために、送信側の表示が120Hzのような輝度のCFFを超えるリフレッシュレートを必要とするため、比較的高いハードウェア要件を必要とする。

(2) 色度変調さらに、いくつかの研究では、約25Hz [23]であり、輝度よりも低い色度のCFFを活用することで、ハードウェア要件を削減しながら、不可視のスクリーン-カメラ通信を実現することが可能となっている。VRCodes[24]は、グレーセルのCIE色度図上で対称な位置に色のペアを生成し、それらを高速で交互に表示することで、不可視色変調を実現した。本手法は、元の色がグレーであるセルに対してのみ情報を埋め込むことを試みており、任意の画像に対する埋め込みを実現していないため、利用が制限される。さらに、山本ら[25]は、モバイルインタラクションのためのカメラのポーズを推定するために、不可視情報としてランダムドットマーカを組み込んでいる。トランスミッター側に60Hzの正変調画像と負変調画像のペアを提示し、カメラのキャプチャレートをレシーバー側に45fpsに設定することで、レシーバー側に3フレーム周期で色の変化を取得する。問題は、オリジナルコンテンツがビデオで、ローリングシャッターカメラが使用されている場合、デコード率が低くなることである。阿部ら[3]は、同一輝度の2つの変調色を60Hzのリフレッシュレートで交互に表示し、QRコードを主に静止画に埋め込む、知覚できない色振動法を提案した。受信側では、スマートフォンのカメラで1/24秒ごとに画像を撮影し、現在のフレームと2番目の前のフレームとの間の色度成分の差の絶対値を計算した。その後、ローリングシャッター効果に対処するため、絶対電流差と事前差を加えた。次に、追加された画像にぼかしフィルタを適用し、二値化した。この方法では、元の映像コンテンツが激しい動きをしている場合、時間的に離れたフレーム間の差分を計算することで埋め込まれた情報を取り出すため、時間的な変化によりデコーディングが損なわれる。(3) デコーディング方法

これらの方法では、画素値を時間的に変調してデータを埋め込むため、デコーディングは通常、受信側で正変調画像と負変調画像の差を計算する。Kleinら[26]は、キャプチャされたフレームのフレーム間差分から適切な垂直分割領域を抽出し、復号化のために表示フレームを再構成することで、キャプチャ率要件を低減する方法を提案した。

UnseenCode[27]は、カラー画像のクロスコンポーネント相関を利用した画像ベースの抽出手法を導入している。この方法は、XYZ色空間のX成分の値の足し算と引き算によって情報を埋め込む。受信側では、1枚のキャプチャ画像から得られた変調されたX成分値と変調されていないZ成分値を用いてデコードが行われる。この方法は、1枚の画像から復号を行うことができるため、フレーム間差に頑健である。しかし、埋め込まれた情報を認識するために、単純で既知のパターンとの相関に依存する。したがって、複雑なパターンや未知のパターンのデコーディングは困難であると考えられる。

3. 方法

我々は、ビデオに埋め込まれた情報を頑健に認識することを可能にする、知覚できない色振動のためのデコーディングアプローチを提案する。

3.1 エンコード処理

提案する復号化方式を説明する前に、知覚できない色振動映像の生成方法について説明する。図2(a)に示すように、QR符号を埋め込む各画素に対して、 $YCbCr$ 色空間の C_b C_r 平面上に、元の色に対して対称に位置する2つの変調色を生成する。次に、この操作を各画素に適用することで、2つの相補的なフレームが得られる。エンコードは30fpsの動画のフレームごとに処理され、2つの画像ペアが生成され、それらを連結して、図2(b)に示すように30Hzの色度フリッカーを示す60fpsの動画が作成される。

3.2 デコーディングプロセス

デコーディング方法は4つのステップからなる。1) 色補正、2) 同時撮影、3) 視野角補正、4) デコーディング。この方法はSwiftを使って実装され、カメラの制御はコードで許可される機能に限定されているため、低レベルのハードウェアへのアクセスは我々の実装には含まれていない。広角カメラとテレフォトカメラを搭載したiPhone 13 Proを使用した。

まず、AppleのAVFoundationフレームワークで提供されている機能を使って、2台のカメラ間で色補正を行った。

その後、異なる露光時間で2つの画像を同時に撮影した。ローリングシャッター効果を考慮し、

count for texture complexity. These methods require relatively high hardware requirements, as the display on the sender side needs to have a refresh rate exceeding the CFF of luminance, such as 120Hz, to generate above 40–50 Hz vibration.

(2) Chromaticity Modulation

In addition, several studies have leveraged the CFF of chromaticity, which is approximately 25 Hz [23] and is lower than that of luminance, making it possible to achieve invisible screen–camera communication with reduced hardware requirements. VRCodes [24] achieved invisible color modulation by generating a pair of colors in symmetrical positions on the CIE chromaticity diagram for a gray cell and by alternately displaying them at high speed. This method attempts to embed information only for cells whose original color is gray, and does not realize embedding for arbitrary images, which limits its use. Moreover, Yamamoto et al. [25] incorporated random dot markers as invisible information to estimate a camera’s pose for mobile interactions. By alternatively presenting positively and negatively modulated image pairs at 60 Hz on the transmitter side and setting the capture rate of the camera to 45 fps on the receiver side, color changes are acquired at a period of three frames on the receiver side. The problem is that the decoding rate is low when the original content is video and a rolling shutter camera is used. Abe et al. [3] proposed an imperceptible color vibration method, wherein two modulated colors with identical luminance were alternately displayed at a refresh rate of 60 Hz, embedding QR codes mainly into still images. On the receiver side, the smartphone camera captured images every $1/24$ s, and the absolute difference in the chromaticity components between the current frame and the second preceding frame was computed. Subsequently, the absolute current difference and prior difference were added to address the rolling shutter effect. Then, a blur filter was applied to the added image, followed by binarization. In this method, decoding is impaired by temporal changes if the original video content has intense movements because the embedded information is retrieved by calculating the difference between temporally distant frames.

(3) Decoding Methods

In these methods, which embed data by temporally modulating pixel values, decoding typically involves calculating the difference between positively and negatively modulated images on the receiver side. Klein et al. [26] proposed a method to decrease the capture

rate requirement by extracting appropriate vertically divided regions from the interframe differences of captured frames and reconstructing the displayed frame for decoding purposes. UnseenCode [27] introduces an image-based extraction method that employs the cross-component correlation of color images. This method embeds information by adding and subtracting values of the X component in the XYZ color space. On the receiver side, decoding is performed using the modulated X component values and the unmodulated Z component values obtained from a single captured image. This method is robust to interframe differences since decoding can be performed from a single image. However, it relies on correlation with simple and known patterns to recognize the embedded information. Therefore, decoding of complex or unknown patterns is considered challenging.

3. Method

We propose a decoding approach for imperceptible color vibrations that enables the robust recognition of information embedded in videos.

3.1 Encoding Process

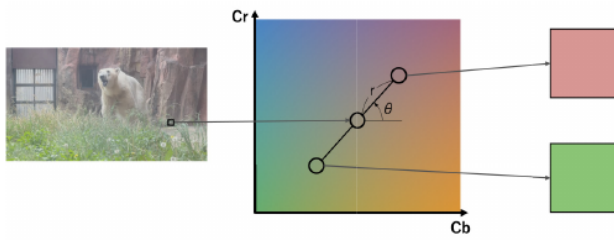
Prior to explaining the proposed decoding method, we describe the generation method of the imperceptible color-vibration videos. As shown in Fig. 2 (a), for each pixel where the QR code is to be embedded, two modulated colors are generated which are symmetrically located on the C_bC_r plane of the YC_bC_r color space to the original color. Then, two complementary frames are obtained by applying this operation to each pixel. The encoding is processed at every frame of 30-fps videos to generate two image pairs and they are concatenated to make 60-fps videos which presents a chromatic flicker of 30 Hz as shown in Fig. 2 (b).

3.2 Decoding Process

The decoding method comprises of four steps. 1) color correction, 2) simultaneous image capture, 3) angle-of-view correction, and 4) decoding. We implemented this method using Swift, and the camera controls are confined to the capabilities allowed by the code, access to lower-level hardware was not part of our implementation. We used an iPhone 13 Pro featuring wide-angle and telephoto cameras.

First, color correction was performed between the two cameras using the functions provided within Apple’s AVFoundation framework.

Subsequently, two images were captured simultaneously at different exposure times. Considering the

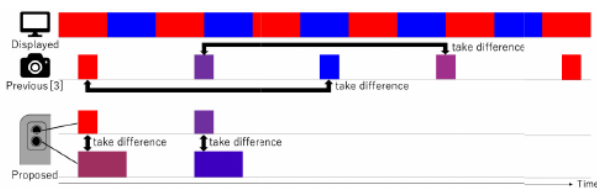


(a) 送信側の各画素の色調変調について

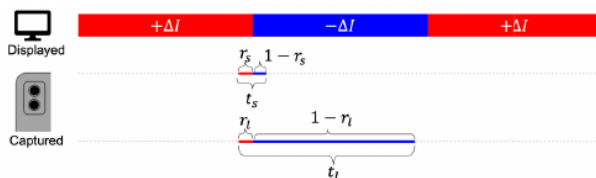


(b) エンコードはフレーム毎に処理され、カラー振動動画を作成する。

Fig. 2 エンコーディング方法の説明



(a) 従来システムと提案システムのタイムシーケンス



(b) 同時キャプチャによって得られる色度差は、変調された画像のキャプチャ時間の比率によって決定される。

Fig. 3 キャプチャ方法。提案手法では、デュアルカメラスマートフォンを用いて、露光時間の異なる2つの画像を同時にキャプチャする。

同時撮影を2回行い、阿部ら[3]の方法と同様に、復号化ステップで4枚の画像を用いて埋め込みQRコードを抽出した。2台のカメラのキャプチャパラメータについては、セクション4で説明する。

画像は、図3(a)に描かれているように、それぞれ露光時間の異なるデュアルカメラスマートフォン上で、2台のカメラを用いて同時に撮影される。露光時間を変化させることで、色変調の量が異なる2つの画像が得られる。

この文脈で、スマートフォンが獲得する色差を計算した。ビデオコンテンツの場合、色の違いを正確にモデル化することは難しく、隣接する2つのフレームはほぼ同じである。

したがって、表示された内容が静止画の色振動であるという仮定のもとで、色差が計算される。知覚できない色振動によって後置変調された2つの画像は、正変調画像と負変調画像と呼ばれる。これらの画像は、画面上に交互に表示される。

色振動の振幅を ΔI とする。二重カメラの露出が正変調された画像が表示されている間に開始された場合、色差は以下のように表される：

$$\begin{aligned} & |\{\Delta I \cdot r_s + (-\Delta I) \cdot (1 - r_s)\} - \{\Delta I \cdot r_l + (-\Delta I) \cdot (1 - r_l)\}| \\ & = 2\Delta I \cdot |r_s - r_l|, \end{aligned} \quad (1)$$

ここで、 r_s , r_l は、図3(b)に示すように、それぞれ、露光時間が短いカメラと長いカメラが、正変調された画像を撮影するのにかかる時間の比率を示す。

現実には、キャプチャされた画像の明るさや色は、絶対的な露光時間によって異なる場合がある。露光時間を変化させることによる入射光量の違いは、見たときにコントラストやシャープネスにわずかな変化を引き起こす可能性があるため、それに応じてISO感度を調整することで最小限に抑えることを試みた。ここでの計算は、セクション4で説明した実験中にISOの設定を変更することで、ダイナミックレンジを適切に調整したという事実に基づいて簡略化した。したがって、提案手法を用いて得られる色差は、各カメラの変調画像のキャプチャ時間の比率に依存し、これはキャプチャプロセスの露光時間とオンセットによって決定される。

次に、AppleのVisionフレームワークのアフィン変換を適用することで、2つのカメラ間で画角補正を行った。計算された変換行列は、2つのカメラのスケールが事前に調整されているため、2つのカメラ間の変換を表している。

最後に、同時に撮影された2つの画像間の輝度を除いた成分の差の絶対値を求めることで、デコーディングを行った。 $YCbCr$ 色空間を前述のように我々の実装で採用し、 C_b 成分と C_r 成分の差の絶対値を計算し、合計した。この復号化処理を同時に撮影したペアに適用し、2種類の画像を得た。2つの差分画像を追加した後、 15×15 のぼかしフィルタを採用し、画角補正時に発生するエッジなどのノイズを軽減した。

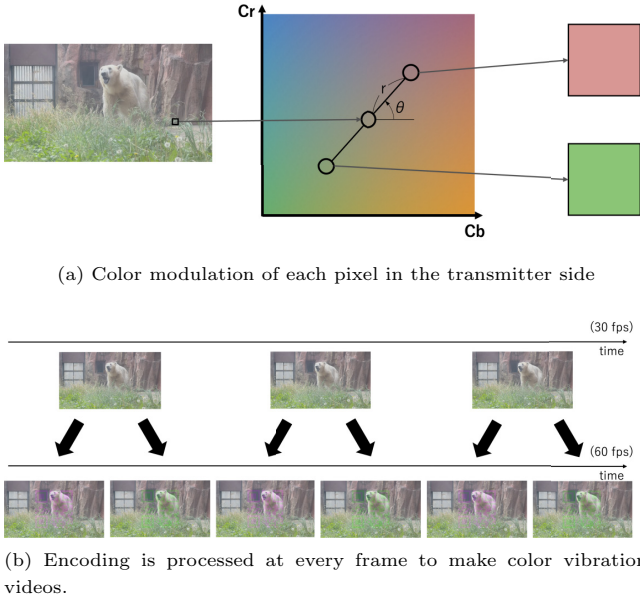


Fig. 2 Description of the encoding method.

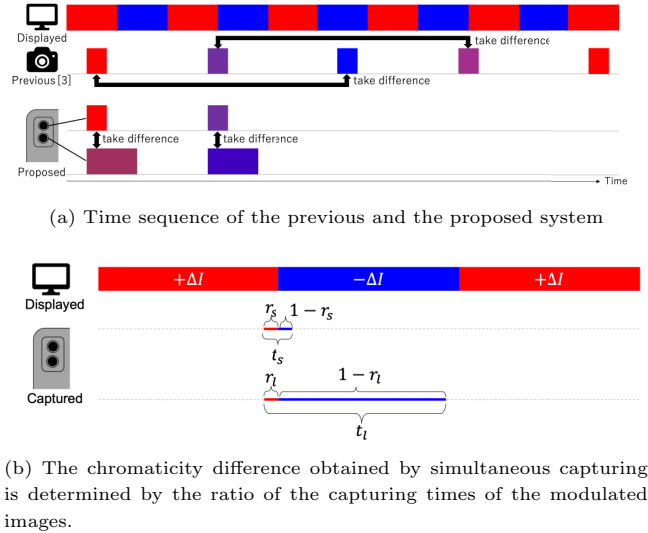


Fig. 3 Capture method. The proposed method employs a dual-camera smartphone to simultaneously capture two images with different exposure times.

rolling shutter effect, simultaneous capture was conducted twice, and the embedded QR code was extracted using four images during the decoding step, similar to the method of Abe et al. [3]. The capture parameters for the two cameras are described in Section 4.

The images are captured simultaneously using two cameras on a dual-camera smartphone, each with a different exposure time, as depicted in Fig. 3 (a). By changing the exposure time, two images with different amounts of color modulation are obtained.

In this context, we calculated the color difference acquired by the smartphone. In the case of video contents, it is difficult to model the color difference accurately, and two adjacent frames are almost identical.

Thus, the color difference is calculated under the assumption that the displayed content is the color vibration of a still image. The two images, post-modulated by imperceptible color vibrations, are referred to as the positive-modulated image and negative-modulated image. These images are displayed alternately on the screen.

Let ΔI represent the amplitude of the color vibration. If the exposure of the dual camera begins while a positive-modulated image is displayed, the color difference is represented as follows:

$$\begin{aligned} & \{ \Delta I \cdot r_s + (-\Delta I) \cdot (1 - r_s) \} - \{ \Delta I \cdot r_l + (-\Delta I) \cdot (1 - r_l) \} \\ &= 2\Delta I \cdot |r_s - r_l|, \end{aligned} \quad (1)$$

where r_s, r_l denote the ratio of the time taken by cameras with short and long exposure times to capture a positive-modulated image, respectively, as shown in Fig. 3(b).

In reality, the brightness or color of the captured images may differ depending on the absolute exposure times. The difference in the amount of light entering due to varying exposure times can cause slight changes in contrast and sharpness when viewed, which we attempted to minimize by adjusting the ISO sensitivity accordingly. The calculation here was simplified based on the fact that the dynamic range was adjusted appropriately by modifying the ISO settings during the experiments described in Section 4. Thus, the color difference obtained using the proposed method depends on the ratio of the capture times of the modulated images for each camera, which is determined by the exposure time and onset of the capture process.

Next, an angle-of-view correction was performed between the two cameras by applying an affine transformation of Apple's Vision framework. The computed transformation matrix represents the translation between the two cameras, as the scaling of the two cameras was adjusted beforehand.

Finally, decoding was performed by obtaining the absolute difference in the components, excluding the luminance, between two simultaneously captured images. $YCbCr$ color space was employed in our implementation, as mentioned previously, and the absolute difference between the C_b and C_r components was calculated and summed. This decoding process was applied to the simultaneously captured pairs to yield two different images. After two difference images were added, a 15×15 blur filter was adopted to mitigate noise such as edges that arise during the angle-of-view correction. The

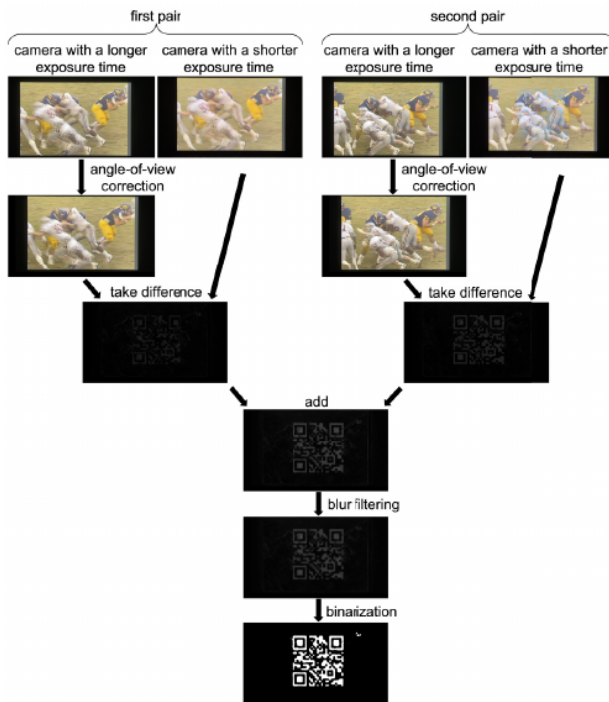


Fig. 4 デコーディング処理のシーケンス。元のビデオコンテンツは「サッカー」[29]である。

グレースケール画像は、大津の方法[28]を用いて2値化し、QRコードを検出した。デコード処理のシーケンスを図4に示す。

4. 実験結果

4. 1 装置と一般パラメータデュアルカメラスマートフォンのキャプチャパラメータを決定するための実験を行い、その後、提案手法の復号性能を阿部らが提案した従来手法と比較した。送信側にMacBook Pro(14inch画面、2021)、受信側にiPhone 13 Proを使用した。MacBook Proの表示プリセットは、Apple XDR Display(P3-1600 nits)のデフォルト設定に設定した。スマートフォンはディスプレイから81cmの距離に設置され、両カメラの画角からディスプレイ全体が見えるようにした。実験はすべて暗室(ADR-F2;ASONE)で行った。

図5(a)に示すQR符号は、不可視情報として埋め込まれている。エラー訂正レベルの変化はデコードレートに影響を与える可能性があるため、一貫性を確保するため、QRコードのエラー訂正レベルは実験中'L'に固定した。 C_b C_r 面における色変調には、輝度成分Yを固定した2つのパラメータ、すなわち図2(a)に示すような振動半径(r)と振動角度(θ)がある。

Table 1 ISO値を変化させた場合の広角カメラとテレフォトカメラの輝度差。

Telephoto ISO	800	400	250	200	100
Difference (wide-angle - telephoto)	-5.9	5.0	8.7	14.5	35.4

振動半径は、 C_b 成分と C_r 成分を-128から127と計算したときの変調色と原色の距離を表す。振動角は C_b 軸と C_b C_r 平面上の変調色と原色を結ぶ線との間の角度である。振動半径 $r = 25$ 、振動角度 $\theta = 135^\circ$ を採用し、MPEG-4コーデックを用いて 57 から 198 の範囲でエンコードされた動画を生成した。エンコード処理で使用したパラメータは付録の通りである。エンコードされた動画は、60Hzのリフレッシュレートでフルスクリーンに表示された。

4. 2 キャプチャパラメータの決定

適切な捕獲率と暴露時間を決定するために実験を行った。足球」ビデオ(図5(b))は、激しい動きを特徴とするため、これらの実験ではオリジナルのビデオコンテンツとして使用された。ビデオは最大輝度設定で表示された。

デフォルトでは、広角カメラとテレフォトカメラの露光時間は、それぞれ1/50秒と1/250秒に設定された。露光時間が短いカメラについては、理論的には露光時間が短いほど良い。これは、露光時間が短いカメラほど、複数のフレームにわたって撮影する可能性が低く、したがって、より大きな変調で撮影することができるためである。しかし、露光時間が極端に短いため、S/N比が低下し、デコードレートが低下する。したがって、一般的な暴露時間範囲内では比較的短い値を使用した。このように、広角カメラの最適な露光時間は後で決定されるが、テレフォトカメラの露光時間は一貫して1/250秒に設定された。

広角カメラのISO感度は、飽和を防ぐため、最低値である50に設定した。2つのカメラ間でほぼ等しい輝度を得るために、様々なISO設定下で輝度差を測定することにより、テレフォトカメラのISO感度を決定した。一様なグレイ画像(RGB値128、128、128)を最大輝度で表示し、グレースケールに変換した後、撮影した画像の輝度を評価した。その結果、表1に示すように、光量と光量の間の輝度差が最も小さいことが明らかになった。

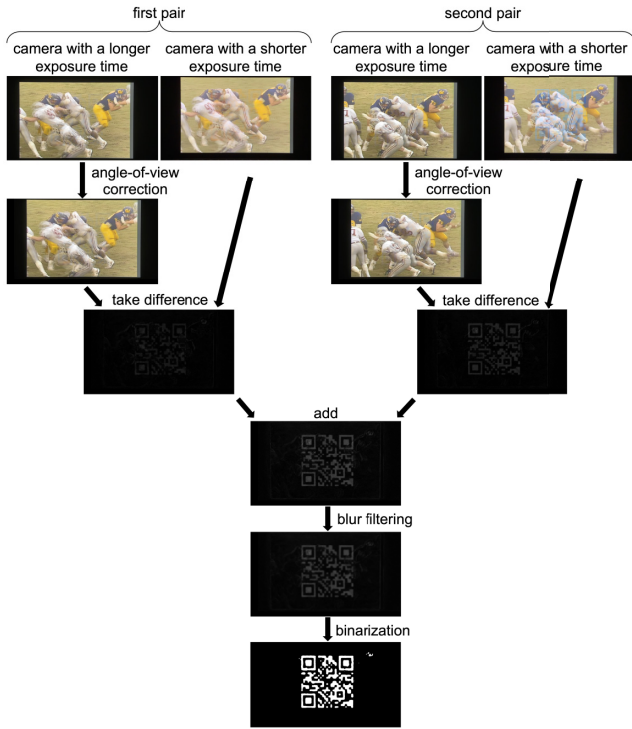


Fig. 4 The sequence of the decoding process. The original video content is ‘football’ [29].

grayscale image was binarized using Otsu’s method [28] to detect the QR codes. The sequence of the decoding process is shown in Fig. 4.

4. Experiments

4.1 Equipment and General Parameters

We first conducted experiments to determine the capturing parameters of a dual-camera smartphone and subsequently compared the decoding performance of the proposed method with that of the previous method proposed by Abe et al. We used a MacBook Pro (14-inch screen, 2021) as the sender and an iPhone 13 Pro as the receiver. The display preset of the MacBook Pro was set to the default setting, which is the Apple XDR Display (P3-1600 nits). The smartphone was placed at a distance of 81 cm from the display, such that the entire display could be seen from the angle of view of both cameras. All experiments were conducted in a dark room (ADR-F2; ASONE).

The QR code shown in Fig. 5 (a) is embedded as invisible information. The error correction level of the QR codes was fixed to ‘L’ throughout the experiments to ensure consistency, as changes in the error correction level might affect the decoding rate. There are two parameters for color modulation in the C_bC_r plane with a fixed luminance component Y : the vibration radius (r) and the vibration angle (θ) as shown in Fig. 2 (a).

Table 1 Brightness difference between the wide-angle camera and the telephoto camera with varying the ISO values.

Telephoto ISO	800	400	250	200	100
Difference (wide-angle – telephoto)	–5.9	5.0	8.7	14.5	35.4

The vibration radius represents the distance between the modulated and original colors when the C_b and C_r components are calculated to be -128 to 127 . The vibration angle is the angle between the C_b axis and the line connecting the modulated and original colors in the C_bC_r plane. We employed a vibration radius of $r = 25$ and vibration angle of $\theta = 135^\circ$, and the encoded videos were generated with a gamut of 57 to 198 using the MPEG-4 codec. The parameters used in the encoding process are described in the Appendix. The encoded videos were displayed on a full screen at a refresh rate of 60 Hz.

4.2 Decision of Capturing Parameters

Experiments were conducted to determine appropriate capture rate and exposure times. A ‘football’ video (Fig. 5 (b)) was used as the original video content for these experiments, as it features intense movements. The video was displayed at the maximum brightness setting.

By default, the exposure times for the wide-angle and telephoto cameras were set to $1/50$ and $1/250$ s, respectively. As for the camera with shorter exposure time, theoretically, the shorter the exposure time, the better. This is because the camera with shorter exposure time is less likely to capture across multiple frames, and therefore can capture images with a larger amount of modulation. However, the extremely short exposure time lowers the signal-to-noise ratio, leading to a lower decoding rate. Therefore, a relatively short value within the general exposure time range was used. Thus, the optimal exposure time for the wide-angle camera will be determined later but the exposure time for the telephoto camera was consistently set at $1/250$ s.

The ISO sensitivity of the wide-angle camera was set to 50, which is the lowest value, to prevent saturation. To obtain an approximately equal brightness between the two cameras, we determined the ISO sensitivity of the telephoto camera by measuring the brightness differences under various ISO settings. A uniform gray image (RGB values 128, 128, 128) was displayed at maximum brightness, and the brightness of the captured images was evaluated after converting them to grayscale. The results, as presented in, Table 1, reveals that the smallest brightness discrepancy between the



(a) QRコード(<https://www.nae-lab.org/> バージョン2、誤り訂正レベルL)



(b) football [29]



(c) 大きな降圧バニーシーン [29]



(d) walking [29]

Fig. 5 実験に使用したQRコードとビデオ

テレフォトカメラのISO設定を400に設定した場合、2台のカメラが発生した。この結果に基づき、このISO設定を以降の実験に使用した。

(1) キャプチャ率

セクション3で説明したように、デコード処理ごとに2回同時キャプチャを行った。

このように、捕捉率は、1回の復号処理で同時に捕捉されたペアの間隔を決定した。この区間は、撮影された画像ペアの間に適切な色差を生成する上で重要な役割を果たす。ディスプレイのフレーム遷移に対してキャプチャのタイミングがずれているため、あるペアの色差が不十分な場合は、別のペアに頼って補正する。さらに、実装上の理由から、2組のキャプチャを完了してから復号後にそれらを再キャプチャするまでの時間も、キャプチャ間隔の整数倍を構成し、キャプチャ率に依存する。これにより、タイミングの問題で失敗した場合、後続の復号化処理のキャプチャタイミングを調整することができ、復号化に必要な色差を得る可能性が向上する可能性がある。したがって、復号化率とその変動は捕捉率に影響されると仮定する。

スマートフォンのキャプチャレートを15fpsから30fpsまで変化させ、デコードレートを測定した。各捕捉率は50回の測定で評価した。各測定において、ループ再生で「サッカー」映像を表示し、1つの映像シーケンスに相当する時間、復号処理を行った。さらに、各測定後にデコーディングアプリケーションを再構築し、ディスプレイのフレーム遷移タイミングと2台のカメラのキャプチャタイミングとの間のシフトをリセットするようにカメラ接続を初期化した。デコーディング率は、デコーディング処理の総実行回数のうち、成功したQRコード認識回数として算出した。

図6に示すように、復号化率の平均値には大きな差があり、復号化率のばらつきは捕捉率によって異なることがわかる。これらの結果は、2つのカメラ間のキャプチャタイミングの偏りと、ディスプレイフレーム遷移タイミングに対する2つのカメラのキャプチャタイミングのずれという2つの観点から議論される。

まず、2台のカメラ間の撮影タイミングの不一致を検討する。セクション3では、2台のカメラが同時に画像を撮影することを想定した復号化手法を提案する。しかし、カメラのAPIの制限により、2つのカメラ間の撮影タイミングに若干の不一致があった。2つのカメラの露光時間を異なる設定にする場合、個別に処理する必要があり、完全な同期が達成できない。この不一致が大きいほど、2台のカメラで撮影された画像間のフレーム間差が大きくなり、その結果、デコード成功率が低下する。



(a) QR code (“<https://www.nae-lab.org/>”, version 2, error correction level L)



(b) football [29]



(c) big_buck_bunny_scene [29]



(d) walking [29]

Fig. 5 QR code and Videos used in the experiments

two cameras occurred when the ISO setting of the telephoto camera was set to 400. Based on the results, this ISO setting was used in our subsequent experiments.

(1) Capture Rate

As described in Section 3, we conducted simultaneous captures twice per decoding process. Thus, the

capture rate determined the interval between simultaneously captured pairs for a single decoding process. This interval plays a crucial role in generating an appropriate color difference between captured image pairs. If there is insufficient color difference in one pair due to the timing of capture relative to the frame transition of the display, we rely on another pair to compensate. Furthermore, for implementation reasons, the time from completing the capture of two pairs to recapturing them after decoding also constitutes an integer multiple of the capture interval, which depends on the capture rate. This allows for adjustments in the capture timing of subsequent decoding processes if one fails due to timing issues, potentially improving the likelihood of obtaining the necessary color difference for decoding. Therefore, the decoding rate and its variation are assumed to be influenced by the capture rate.

The decoding rate was measured by varying the smartphone’s capture rate from 15 fps to 30 fps. Each capture rate was evaluated fifty measurements. In each measurement, we displayed the ‘football’ video in loop playback and performed the decoding process for a duration equivalent to one video sequence. In addition, the decoding application was rebuilt after each measurement, and the camera connections were initialized to reset the shift between the frame transition timing of the display and the capture timing of the two cameras. The decoding rate was calculated as the number of successful QR code recognitions out of the total number of decoding processes performed.

The results, shown in Fig. 6, exhibit a substantial difference in the mean value of the decoding rate and variation in the decoding rate depending on the capture rate. These results are discussed from two perspectives: the skew in the capture timing between the two cameras and the shift in the capture timing of the two cameras relative to the display frame transition timing.

First, we consider the discrepancy in capture timing between the two cameras. In Section 3, we propose a decoding method that assumes that two cameras simultaneously capture images. However, due to camera API limitations, there was a slight discrepancy in the capture timing between the two cameras. When setting different exposure times for the two cameras, they must be handled individually and perfect synchronization is not achievable. The larger this discrepancy, the greater will be the interframe difference between the images captured by the two cameras, resulting in a decreased decoding success rate.

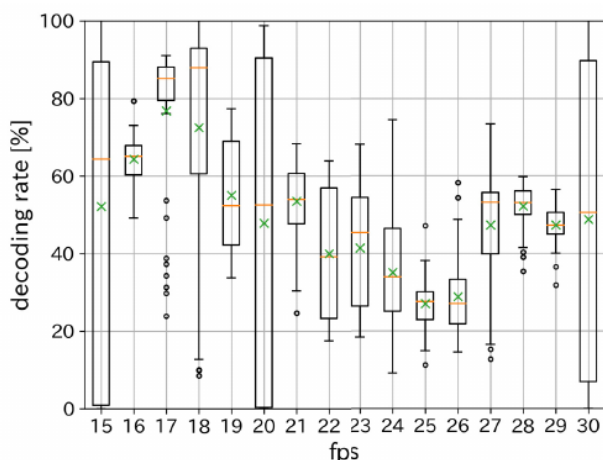


Fig. 6 キャプチャレートを変化させた場合のデコードレートの箱ひげ図

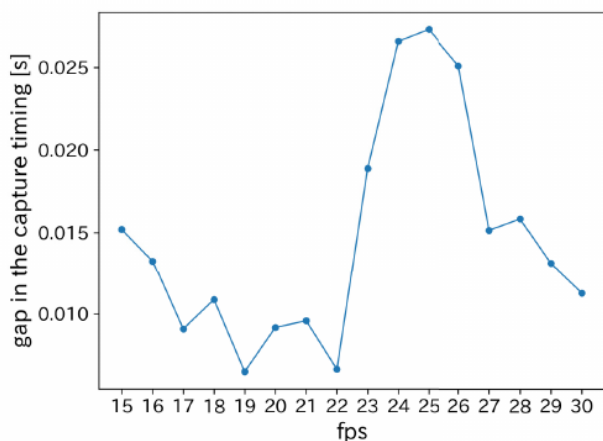
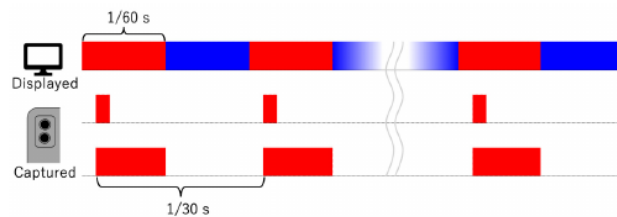


Fig. 7 2台のカメラ間の撮影時間のスキューは、撮影速度が異なる。

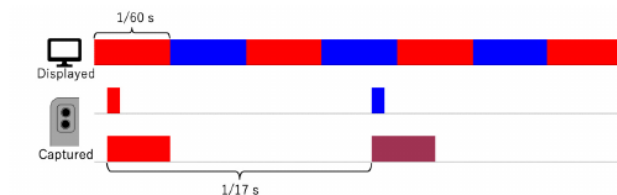
スマートフォンアプリでは、2台のカメラ間の撮影時間の偏りは一定ではなかった。そこで、同じ撮影速度条件下で、2台のカメラ間の画像ペアの撮影タイミングの違いを測定した。測定は、連続する20のデコーディング・プロセスごとに10回行われ、その結果、各条件について合計200回の測定が行われた。図7は、2つのカメラ間のキャプチャ時間の平均差を示している。

その結果、2台のカメラ間の撮影タイミングの差は23~26fpsと特に大きいことがわかった。1フレームの表示時間が1/60 0.0167秒であることを考慮すると、2台のカメラの撮影時間のスキューは平均して1フレーム以上離れている。したがって、フレーム間差は23~26fpsで大きくなり、図6に示すようにデコードレートが低下する。

その後、ディスプレイのフ



(a) 捕捉率が30fpsの場合、十分な色差が得られないことがある。



(b) キャプチャレートが17fpsの場合、表示フレーム遷移に対してキャプチャタイミングが約半分にシフトしている。

図8 キャプチャレートによる違い

フレーム遷移タイミングに対する2つのカメラのキャプチャタイミングのシフトについて議論する。ディスプレイとスマートフォンは同期させることができなかったため、2台のカメラの最初の同時撮影のフレーム遷移に対する撮影タイミングを、各測定におけるカメラ接続の初期化ごとにランダムに決定した。その結果、その後の捕獲タイミングは、最初の同時捕獲時刻と捕獲率によって決定される。

15、20、30fpsの場合、キャプチャ間隔が1/60秒(ディスプレイのフレーム遷移間隔)の整数倍であると考え、2つの画像ペアのキャプチャタイミングは、各デコーディング処理においてディスプレイのフレーム遷移と完全に一致した。このような条件下で、2組のうち1組から十分な色差が得られない場合は、もう1組で十分な色差が得られるとは考えにくい。さらに、2組をキャプチャーする終了からデコード後に2組をキャプチャーするまでの時間は、キャプチャー間隔の整数倍である。これは復号率の偏光を引き起こし、最初の同時撮影で撮影された画像から十分な色差が得られれば復号率が高くなり、そうでなければ復号率が低くなる。

図8(a)は、捕捉率が30fpsの場合、十分な色差が連続して得られない例である。その結果、図6に示すように、15、20、30fpsでデコードレートに大きなばらつきが見られた。

逆に、デコードレートのばらつきを抑えるためには、

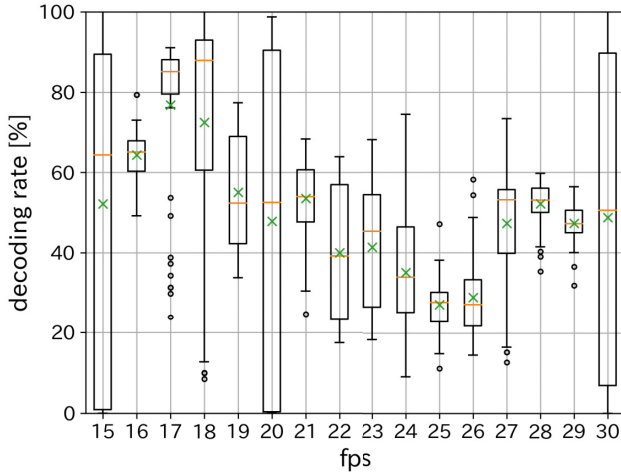


Fig. 6 Box plot of decoding rates at varying capture rates

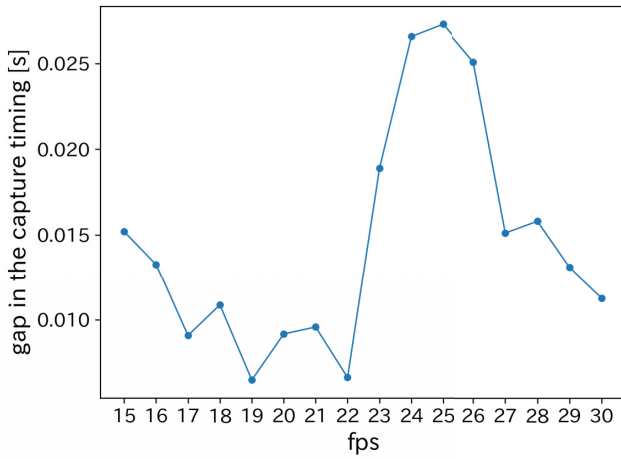
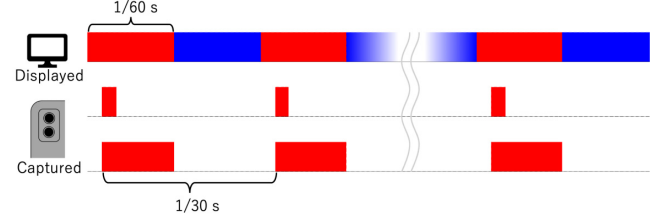


Fig. 7 The skew in capture time between the two cameras varying capture rates

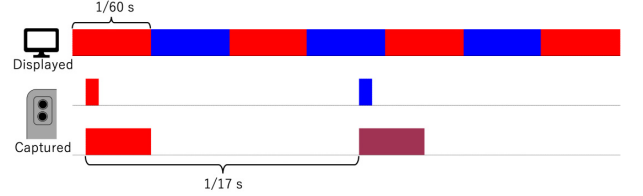
In the implemented smartphone application, the skew in capture time between the two cameras was not constant. Therefore, we measured the difference in the capture timing of the image pairs between the two cameras under the same capture rate conditions. Measurements were performed 10 times for each of the 20 consecutive decoding processes, resulting in a total of 200 measurements for each condition. Fig. 7 shows the average difference in the capture time between the two cameras.

The results indicate that the difference in the capture timing between the two cameras is particularly large at 23 to 26 fps. Considering that the display time of one frame is $1/60 \approx 0.0167$ s, the skew in the capture times of the two cameras is on average more than one frame apart. Therefore, the interframe difference becomes significant at 23 to 26 fps, resulting in lower decoding rates, as shown in Fig. 6.

Subsequently, we discuss the shift in capture timing



(a) When the capture rate is 30 fps, a sufficient color difference may never be obtained.



(b) When the capture rate is 17 fps, the capture timing is shifted by approximately half a frame relative to the display frame transition.

Fig. 8 Differences due to the capture rate

of the two cameras relative to the frame transition timing of the display. Because the display and smartphone could not be synchronized, the capture timing relative to the frame transition of the first simultaneous capture of the two cameras was randomly determined for each initialization of the camera connections at each measurement. Consequently, the subsequent capture timings are determined by the time of the first simultaneous capture and the capture rate.

Considering the cases of 15, 20, and 30 fps, wherein the capture interval was an integer multiple of $1/60$ s (the frame transition interval of the display), the capture timing of the two image pairs aligned completely with the frame transitions of the display for each decoding process. Under these conditions, if a sufficient color difference cannot be obtained from one of the two pairs, it is unlikely that a sufficient color difference can be obtained with the other pair. Additionally, the time from the end of capturing two pairs to the time of recapturing two pairs after decoding is an integer multiple of the capture interval. This causes polarization of the decoding rate, with a high decoding rate if a sufficient color difference is obtained from images captured in the first simultaneous capture; otherwise, a low decoding rate is obtained.

Fig. 8 (a) presents an example wherein a sufficient color difference cannot be obtained consecutively when the capture rate is 30 fps. Consequently, the decoding rate exhibited significant variations at 15, 20, and 30 fps, as shown in Fig. 6.

Conversely, to reduce the variability in the decoding

ディスプレイのフレーム遷移に対してキャプチャタイミングがわずかにずれるようなキャプチャレートが必要である。このキャプチャタイミングシフトの周期を長くするために、ディスプレイのリフレッシュレート(60Hz)とキャプチャレートは大きな共通分周器を共有してはならない。特に、17、19、23、29fpsは、それぞれ60に関しても、互いにも素数であるため、選ばれた。

前述したように、2台のカメラ間の撮影時間の差が大きいため、23fpsでの復号率は低かった。19fpsと29fpsのレートは、それぞれ20fpsと30fpsのレートに近かった。このことは、デコーディング処理に使用した2組の画像のキャプチャタイミングが、ディスプレイのフレーム遷移に対して類似していたことを示唆している。したがって、第1のペアから十分な色差が得られない場合、第2のペアからは補正のための十分な色差が得られない場合がある。

一方、17fpsでは、キャプチャ間隔は $1/17 \approx 0.0588$ 秒であり、 $3.5/60 \approx 0.0583$ に近い。図8(b)に示すように、2つの画像ペアのキャプチャタイミングは、表示フレーム遷移に対して、表示フレームの約半分ずつずれていた。したがって、キャプチャした画像の一方のペアが十分な色差を提供しないとしても、もう一方のペアでデコードするための色差を補完することは比較的容易であると考えられる。図6に示すように、復号率は17fpsで最も高かった。以上の結果から、17fpsの捕捉率を用いることにした。

捕獲率の影響に関する我々の調査は、我々の実験で使用した装置であるiPhone 13 Proに特に最適化されていることに注意すべきである。様々なAndroidスマートフォンのようなデュアルカメラセットアップを持つ他のデバイスでは、これらのデバイスに固有の異なるハードウェアとソフトウェア構成のため、キャプチャレートの再評価が不可欠である。

(2) 露光時間

その後、露光時間が長い広角カメラの最適な露光時間を決定するための実験を行った。

デコーディングレートは7つの異なる露光条件下で測定された: 1/30、1/40、1/45、1/50、1/60、1/80、1/100秒。露光時間が短いテレフォトカメラの露光時間は、前述のように一貫して1/250秒に設定した。広角カメラとテレフォトカメラのISO感度は、前回の実験と同様に、それぞれ50と400に設定した。

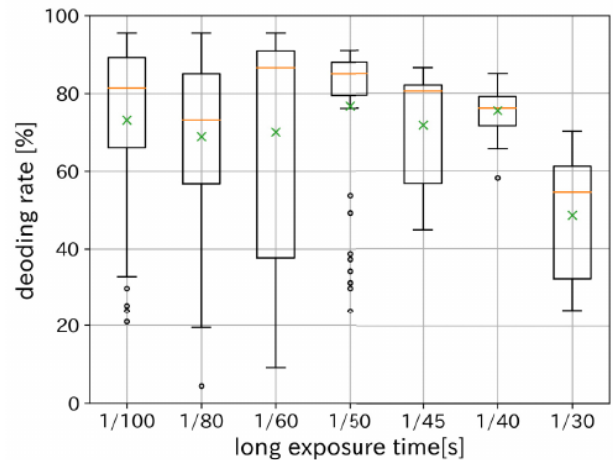


Fig. 9 露光時間を変化させた場合の復号化率の箱ひげ図

正確には、広角カメラの露光時間を変更するには、2台のカメラ間の明るさに合わせてISO感度を調整する必要がある。ISO感度値を変化させながら輝度の変化を視覚的に比較したところ、ISO感度値に有意な差は見られなかったため、固定値を使用した。これまでの実験と同様に、各条件について17fpsの捕捉率で50回の測定を行った。

図9に結果を示す。平均復号化率は、露光時間が1/30秒のときに低くなり、他の条件では有意差は観察されなかった。曝露時間が短いほど変動が大きくなる傾向があった。1/30秒の条件下での復号率の低さを説明する説明として、2つの可能性が考えられる。

まず、十分な色差が得られない場合がある。1/30秒は2つの表示フレームで等価であるため、露光時間が長いカメラでは色変調を捉えることができないと考えられる。このことは、2台のカメラで撮影された画像の色差が小さく、埋め込み符号の抽出が困難であることを意味する。

第二に、輝度飽和が要因になる可能性がある。露光時間が長くなると、カメラはより多くの光を集めるようになり、広角カメラとテレフォトカメラで撮影された画像の輝度差が大きくなる。その結果、埋め込まれたQRコードの色差を適切に取得することができない。

その後、短い露光時間でこのような大きなばらつきが生じる理由について考察する。

まず、2台のカメラの撮影時刻が完全に同期していると仮定し、以下のようにした。

rate, the capture rate should be such that the capture timing shifts slightly with respect to the frame transition of the display. To lengthen the period of this capture timing shift, the refresh rate of the display (60 Hz) and the capture rate should not share large common divisors. In particular, 17, 19, 23, and 29 fps were chosen, as they are prime with respect to 60 and to each other, respectively.

The decoding rate at 23 fps was low because of the large difference in capture time between the two cameras, as mentioned previously. The rates at 19 and 29 fps were close to those at 20 and 30 fps, respectively. This implies that the capture timings of the two pairs of images used for the decoding process were similar relative to the frame transitions of the display. Therefore, if a sufficient color difference cannot be obtained from the first pair, then a sufficient color difference for compensation may not be obtained from the second pair.

In contrast, at 17 fps, the capture interval was $1/17 \approx 0.0588$ s, which is close to $3.5/60 \approx 0.0583$. The capture timings of the two image pairs shifted from each other by approximately half of the displayed frame relative to the display frame transition, as shown in Fig. 8 (b). Therefore, it is assumed that even if one pair of captured images does not provide sufficient color difference, it is relatively easy to complement the color difference for decoding with the other pair. The decoding rate was highest at 17 fps, as shown in Fig. 6. Based on the aforementioned findings, we decided to use a capture rate of 17 fps.

It should be noted that our investigation concerning the impact of capture rate is specifically optimized for the iPhone 13 Pro, which was the device used in our experiments. For other devices with a dual-camera setup, such as various Android smartphones, a re-evaluation of the capture rate is essential, due to the distinct hardware and software configurations inherent to these devices.

(2) Exposure Time

We subsequently conducted an experiment to determine the optimal exposure time for a wide-angle camera with a longer exposure time.

The decoding rate was measured under seven different exposure conditions: 1/30, 1/40, 1/45, 1/50, 1/60, 1/80, and 1/100 s. The exposure time for the telephoto camera, which had a shorter exposure time, was consistently set at 1/250 s, as previously mentioned. The ISO sensitivities of the wide-angle camera and the telephoto camera were set to 50 and 400, respectively, as

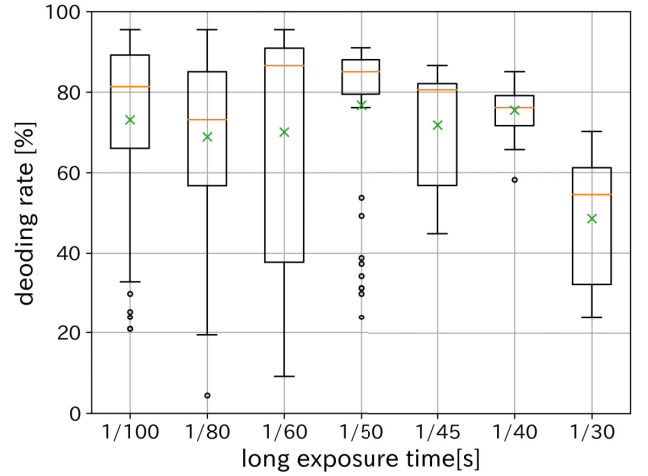


Fig. 9 Box plot of decoding rates at varying exposure times

in the previous experiments. To be precise, changing the exposure time of the wide-angle camera will require adjusting its ISO sensitivity to match the brightness between two cameras. Upon visually comparing the change in brightness while altering the ISO sensitivity value, we found that it did not make a significant difference, hence a fixed value was used. Similarly to the previous experiments, we performed 50 measurements for each condition at a capture rate of 17 fps.

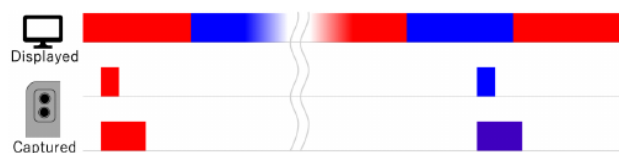
Fig. 9 shows the result. The average decoding rate was lower when the exposure time was 1/30 s, with no significant differences observed for the other conditions. The variability tended to be greater when the exposure time was shorter. Two possible explanations account for the low decoding rate under the 1/30 s condition.

First, it is likely that a sufficient color difference cannot be obtained. As 1/30 s is equivalent for two display frames, it is assumed that a camera with a longer exposure time cannot capture color modulation. This implies that the color difference between the images captured by the two cameras is small, making it difficult to extract the embedded code.

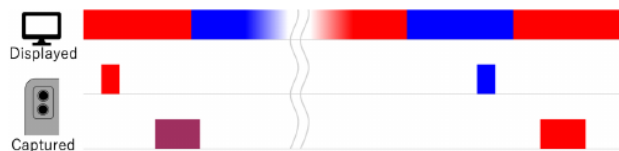
Second, the brightness saturation could be a factor. When the exposure time is longer, the camera collects more light, thereby increasing the brightness difference between the images captured by the wide-angle camera and telephoto camera. This, in turn, hinders proper acquisition of the color difference of the embedded QR code.

Subsequently, we discuss the reasons for this large variation under short exposure times.

First, we assumed that the capture times of the two cameras were perfectly synchronized, as shown in



(a) 2台のカメラの撮影時刻を同期させた場合。



(b) When a skew exists between the capture time of the two cameras.

Fig. 10 2 台のカメラの撮影時間による違い

図10 (a). 長いカメラの露光時間が1/60秒以下であれば、両カメラが純粋なフレームを捉えている、すなわち、正変調画像と負変調画像の混合フレームを捉えていない確率は、フレーム表示時間と同等である。したがって、広角カメラの露光時間が短くなるにつれて、2つの画像の色変調量が等しいことから、復号率は低下する。しかし、実際には、先に述べたように、2台のカメラの撮影タイミングにはスキューが存在し、図10(b)に示すようになる。露光時間が短いほど、より大きな変調量の画像が撮影されることが期待される。

さらに、図7に示すように、キャプチャレート17 fpsで2つのカメラのキャプチャタイミングの差は約0.009sであり、ディスプレイのフレーム表示時間の約半分である。各カメラは反対方向に変調された画像を撮影している可能性が高く、その結果、色差が大きくなり、デコード率が高くなると予想される。その結果、露光時間が短い場合、復号化レートのばらつきが大きくなる。

最適な露光時間を決定するためには、ユーザーエクスペリエンスを向上させるために、安定したデコードレートを維持することが極めて重要である。そこで、平均復号率が比較的高く、すべてのテスト条件の中で最小値が最も高い1/40秒の露光時間を選択した。

4.3 従来手法との比較

(1) デコード率

提案手法の復号率を阿部らの手法と比較した: 図5に示すように、「サッカー」、「大きな降圧バニーシーン」、「歩行」の3つの動画を用いて、提案手法の復号率を阿部らの手法と比較した。デコーディング率は、ループ内で再生された各映像について、

Table 2 のデコード率[%]の比較。
 動画における従来手法と提案手法

	method of Abe et al. [3]		proposed method	
	ISO200	ISO300	maximum rate	maximum brightness
football	6.97	7.45	75.5	75.5
big_buck_bunny_scene	62.6	69.7	96.4	56.5
walking	18.0	23.9	95.5	74.4

映像シーケンスの長さに対応する時間、デコーディング処理を繰り返すことで測定した。各ビデオについて50回の測定が記録された。

阿部らが提案した方法では、ISO感度を200と300に設定し、ディスプレイの明るさを最大に設定したスマートフォンのテレフォトカメラを使用した。逆に、提案手法では、ディスプレイの明るさが最大になるだけでなく、各映像のデコードレートが最大になるようにディスプレイの明るさを設定して測定を行った。このアプローチを採用したのは、ディスプレイの明るさが増すと、埋め込まれたコードの色調変調が大きくなるからである。しかし、露光時間が長いカメラで撮影された画像では、コンテンツによって、過度に明るいディスプレイが飽和を引き起こす可能性がある。表 2 は、3 つのビデオの平均復号率を示している。ディスプレイの明るさを適切に設定することで、すべての動画でデコード率が向上した。

フットボール」映像の場合、ディスプレイの明るさが最大になったときにデコードレートが最大値に達したが、他の2つの映像ではそうではなかった。フットボール」ビデオの背景は草で、緑色が薄くなり、他のビデオは明るい色をしている。その結果、表示輝度が最大となる明るい動画を使用すると、キャプチャされた画像は飽和する。提案手法は、元の映像コンテンツに起因するフレーム間差分を抑制し、ロバストなデコーディングを実現する。

(2) 認識時間

提案手法と阿部らの手法の認識時間を、ISO感度300で比較したところ、表2に示すように、より高い復号率を示した。2. 認識時間は、画像取り込み開始から埋め込みQRコードの正確な読み出しまでの時間と定義した。本実験で使用した映像は、前回の実験で使用したのと同じで、ループで再生したものである。測定は各条件下で50回の反復で行われ、各測定の開始時刻はビデオシーケンス全体を通してランダム化された。

結果を図11にプロットする。阿部らが提案した方法と比較して、提案した方法は、埋め込まれたQRコードを読み取るのにかかる時間が短く、ばらつきも少ない。

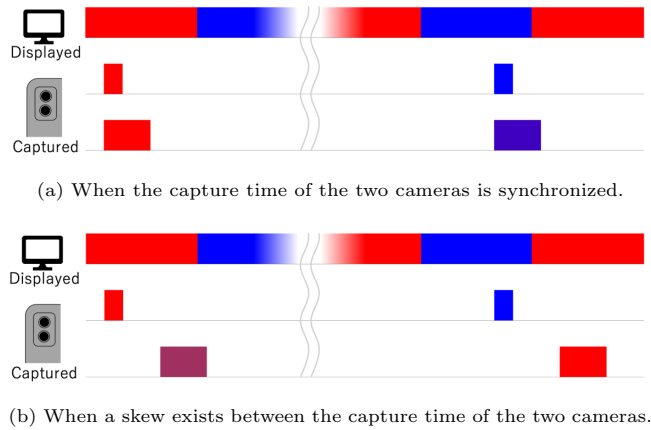


Fig. 10 Differences due to the capture time of the two cameras

Fig. 10 (a). If the exposure time of the longer camera is less than $1/60$ s, equivalent to the frame display time, the probability that both cameras capture a pure frame—i.e., not a mixed frame of positive-modulated and negative-modulated images—increases. Therefore, as the exposure time of the wide-angle camera decreases, the decoding rate decreases, given that the color modulation amounts of the two images are equal. However, in reality, a skew exists between the capture timings of the two cameras, as previously mentioned and shown in Fig. 10 (b). With shorter exposure times, it is expected that images with larger modulation amounts can be captured.

Additionally, the difference in capture timing between the two cameras is about 0.009s at a capture rate of 17 fps, as shown in Fig. 7, which is approximately half the display's frame display time. It is likely that each camera captures images modulated in opposite directions, leading to the expectation that the resulting color difference will be larger and the decoding rate will increase. Consequently, the variation in the decoding rate is substantial when the exposure time is short.

To determine the optimal exposure time, it is crucial to maintain a stable decoding rate for enhanced user experience. Therefore, we opted for an exposure time of $1/40$ s, which had a relatively high average decoding rate and the highest minimum value among all tested conditions.

4.3 Comparison between the Previous Method

(1) Decoding rate

We compared the decoding rate of the proposed method with that of Abe et al.'s method using three videos: 'football', 'big-buck-bunny_scene', and 'walking' videos, as shown in Fig. 5. The decoding rate was measured by repeating the decoding process for a

Table 2 Comparison of decoding rate [%] between the previous and the proposed method in videos

	method of Abe et al. [3]		proposed method	
	ISO200	ISO300	maximum rate	maximum brightness
football	6.97	7.45	75.5	75.5
big_buck_bunny_scene	62.6	69.7	96.4	56.5
walking	18.0	23.9	95.5	74.4

time equivalent to the length of the video sequence for each video played in the loop. Fifty measurements were recorded for each video.

For the method proposed by Abe et al., the telephoto camera of the smartphone was used with ISO sensitivities of 200 and 300, and the brightness of the display was set to maximum. Conversely, for our proposed method, measurements were carried out not only at maximum display brightness, but also with the display brightness set to achieve the maximum decoding rate for each video. This approach was adopted because an increased display brightness results in greater color modulation of the embedded code; however, overly bright displays may cause saturation in images captured by cameras with longer exposure times, depending on the content. Tab. 2 shows the average decoding rates for the three videos. The decoding rate improved for all the videos when setting the display brightness appropriately.

In the case of a 'football' video, the decoding rate reached its maximum value when the display brightness was at maximum, but this was not the case for the other two videos. The background of the 'football' video is grass, which has a faded green color, while the other videos have brighter colors. Consequently, captured images became saturated when using bright videos at maximum display brightness. The proposed method suppresses interframe differences stemming from the original video content, resulting in robust decoding.

(2) Recognition time

We compared the recognition times for the proposed method and the method of Abe et al. using an ISO sensitivity of 300, which exhibited a higher decoding rate, as shown in Tab. 2. We defined the recognition time as the time taken from the start of image capture until accurate reading of the embedded QR code. The videos used in this experiment were identical to those used in the previous experiment and were played in a loop. Measurements were performed over 50 iterations under each condition, and the start time of each measurement was randomized throughout the video sequence.

The results are plotted in Fig. 11. Compared to the method proposed by Abe et al., the proposed method

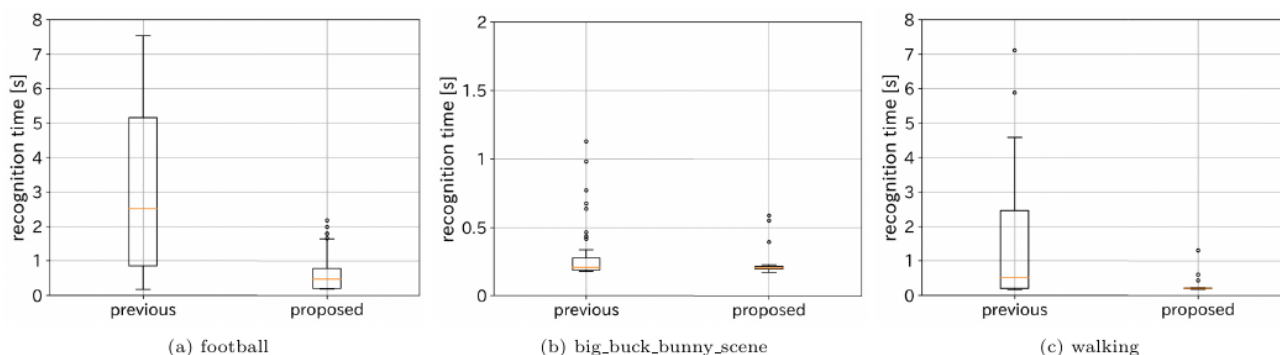


図11 動画における従来手法と提案手法の認識時間の箱ひげ図

阿部らは、大きな動きのあるシーンに埋め込まれたQRコードのデコードに失敗した。したがって、キャプチャ開始時刻にもよるが、QRコードの読み込みに時間がかかる場合がある。しかし、提案手法はフレーム間差の影響を抑制できるため、連続したデコードの失敗が比較的少なかった。今後の課題としては、提案手法を用いた復号化処理時間の短縮化に着目した実装が考えられる。Abeら[3]の方法では、24fps(60Hzの画面では2.5フレームごとに撮影)から40fps(60Hzの画面では1.5フレームごとに撮影)に撮影率を上げることができるが、提案手法で撮影された画像は時間的に近い。その結果、2つの画像間の差は、元のビデオコンテンツの時間的変化の影響を受けにくくなり、その結果、デコード性能が向上する。

5. アプリケーション

提案手法は、映像コンテンツに埋め込まれた不可視情報の抽出を容易にする。システムの汎用性を示すために、3つのアプリケーションを詳細に説明する。

最初のシナリオでは、図12(a)に示すように、アーティストのウェブサイトへのハイパーリンクがミュージックビデオに埋め込まれた。視聴者は、スマートフォンをディスプレイに保持することで、アーティストのウェブサイトと同時にアクセスし、CDやアルバムなどの商品を確認することができる。広告ビデオの対応する製品にハイパーリンクを埋め込むことは、このシステムの基本的かつ有用な応用例である。

さらに、元のビデオコンテンツに基づいて、時間情報を埋め込むことができる。第二のシナリオは、図12(b)に示すように、ユーザーがスマートフォンに表示することで、動画コンテンツの説明にアクセスできるように、時間的に埋め込まれたクローズドキャプションである。

このアプリケーションでは、各QRコードに対して多言語のクローズドキャプションが埋め込まれ、このシステムは複数のユーザーの同時データ検索をサポートするため、各ユーザーが好みの言語を選択できるようにした。キャプションはスマートフォンに表示されるため、視聴体験を妨げない。見えない情報は、ビデオコンテンツに独立して埋め込むことができ、情報提示の柔軟性を提供する。

図12(c)は、観光映像内の特定の場所にスマートフォンを点在させることで、観光スポットに関する情報を取得できるアプリケーションである。この実装では、観光地の名前と画像がスマートフォンに表示される。時間的・空間的に情報を変化させることで、映像内のオブジェクトに対応する情報を埋め込むことが可能である。

6. むすび

本研究では、動画中に埋め込まれた不可視QRコードを頑健に認識するための復号化手法を提案する。本手法では、デュアルカメラスマートフォンを利用し、変調値の異なる画像のペアを同時に撮影する。

この方法は、変調されたQRコードの色差を、時間的に隣接する2つの画像を用いて抽出することができるため、元のビデオコンテンツのフレーム間差に起因するアーチファクトを低減することができる。実験により、ディスプレイのフレーム遷移に対するスマートフォンのキャプチャタイミングのシフトと、スマートフォンの2台のカメラ間のキャプチャタイミングのスキューの影響が明らかになった。実験結果に基づき、これらの影響を抑制し、安定したデコードを確保するためのパラメータを導出した。



Fig. 11 Box plot of recognition time for the previous and the proposed methods in videos

requires less time to read the embedded QR code and exhibits less variation. Abe et al. failed to decode QR codes embedded in scenes with large motions. Therefore, depending on the capture start time, it may take a long time to read the QR code. However, our proposed method experienced relatively fewer consecutive decoding failures since it can suppress the effects of interframe differences. Future work can focus on implementing improvements to reduce the processing time for decoding using our proposed method. While the capture rate in the method of Abe et al. [3] can be increased from 24 fps (capturing every 2.5 frames for a 60 Hz screen) to 40 fps (capturing every 1.5 frames for a 60 Hz screen), the images captured by the proposed method are temporally closer. Consequently, the difference between the two images is less impacted by temporal changes in the original video content, resulting in improved decoding performance.

5. Applications

The proposed method facilitates the extraction of invisible information embedded in video content. Three applications are described in detail to demonstrate the versatility of the system.

In the first scenario, a hyperlink to the artist’s website was embedded in the music video, as shown in Fig. 12 (a). Viewers can instantaneously access an artist’s website to check their products, such as CDs and albums, by holding a smartphone to the display. Embedding a hyperlink into a corresponding product in an advertising video exemplifies the fundamental and useful application of this system.

Additionally, temporal information can be embedded based on the original video content. The second scenario involves temporally embedded closed captions that enable users to access explanations of video con-

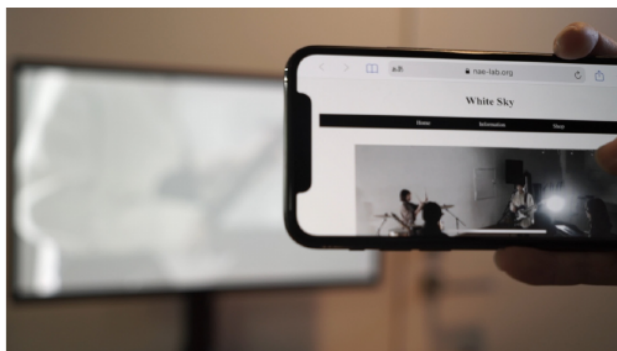
tent by displaying them on their smartphones, as shown in Fig. 12 (b). In this application, multiple-language closed captions were embedded for each QR code, allowing each user to select their preferred language since this system supports simultaneous data retrieval for multiple users. Captions do not obstruct the viewing experience because they are displayed on a smartphone. Invisible information can be independently embedded in the video content, providing flexibility in information presentation.

Fig. 12 (c) shows an application that enables users to acquire information about sightseeing spots by pointing their smartphone at specific spots within a sightseeing video. In this implementation, the name and image of a sightseeing location are displayed on a smartphone. By temporally and spatially varying the information, it is possible to embed information corresponding to the objects within a video.

6. Conclusion

In this study, we propose a decoding method for robustly recognizing embedded invisible QR codes in videos. Our method utilizes a dual-camera smartphone to simultaneously capture a pair of images with different modulation values.

This method can reduce the artifacts caused by interframe differences in the original video content because the color difference of the modulated QR codes can be extracted using two temporally adjacent images. The experiments revealed the effects of the capture timing shift in the smartphone relative to the frame transition of the display and the skew in the capture timing between the two cameras of the smartphone. Based on the experimental results, parameters were derived to suppress these effects and ensure stable decoding. Furthermore, we presented several applications that embed



(a) 音楽ビデオでアーティストのウェブサイト返信する



(b) スマートフォン画面上の即時クローズドキャプション(左:英語、右:日本語)



(c) 観光映像に空間的に埋め込まれたランドマーク情報。

Fig. 12 動画に埋め込まれた情報を抽出するアプリケーション例

さらに、不可視情報を動画に埋め込むアプリケーションをいくつか紹介し、提案システムの実現可能性を実証した。

さらに、様々な照明条件や、様々なディスプレイやスマートフォンの組み合わせで、このシステムの適応性を探ることで、実世界のシナリオにおける汎用性がさらに実証されると考えている。また、変調された映像の範囲を広げることで、視聴体験を向上させることができる。

謝辞

このプロジェクトは中山未来研究所の支援を受けている。

付録

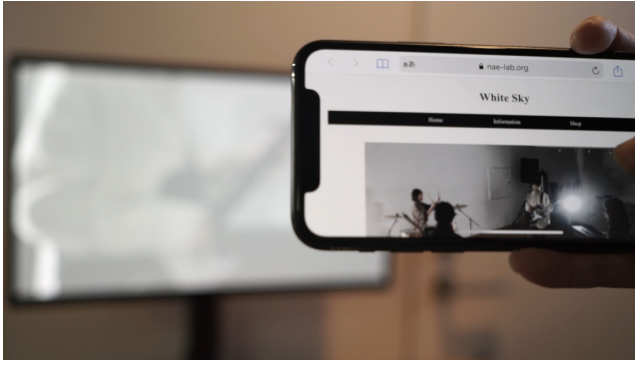
A. 符号化処理で使用するパラメータ 4 章で述べたように、振動半径(r)と角度(θ)は色調変調に使用するパラメータである。特に、振動半径が大きいとデコードが容易になるが、このパラメータが過度に大きくなると、変調が人間の目に知覚できるようになる。安定したデコードのためには変調量が十分に大きくなければならないため、キャプチャパラメータを決定する前に、これら2つのパラメータを決定する実験が行われた。

果実の静止画(図1(a))を原コンテンツとし、QRコード(図1(b))を不可視情報として埋め込んだ。果実画像の色域は、阿部らが提案した従来の方法で57-198に絞り、音色数を均等にした。測定は、 $\theta = 0 \sim, 45 \sim, 90 \sim, 135 \sim$ の4つの振動角条件で行い、振動半径は $r = 5$ から40まで5刻みで行った。 $\theta = 0 \Delta$ と 90Δ の条件では、振動方向に対応する成分のみを復号に使用した。各条件下で50回の反復でデコーディングを連続して行い、このプロセスを20回の反復で繰り返し、平均デコーディング率を求めた。キャプチャレートは、阿部らによる従来の方法と同じ24fpsであり、残りのキャプチャパラメータは、セクション4で説明したデフォルトのパラメータに設定された。なお、 $d = 35$ と40の復号率は、 $\theta = 0 \sim$ では $d = 30$ までの条件でしか振動させることができないため、測定していない。

図2に各条件で測定した復号化レートを示す。振動半径が最も小さい条件 $d = 25$, $\theta = 135 \Delta$ を用いることにしたのは、復号率が 95 % を超える条件のうち、最も小さい条件とした。

References

- 1) P. H. Pathak, X. Feng, P. Hu, and P. Mohapatra : "Visible light communication, networking, and sensing: A survey, potential and challenges", IEEE communications surveys & tutorials 17(4):2047–2077 (2015)
- 2) International Organization for Standardization :, "Information technology — Automatic identification and data capture techniques — QR Code bar code symbology specification" (2015)
- 3) S. Abe, T. Hiraki, S. Fukushima, and T. Naemura : "Imperceptible color vibration for screen-camera communication via 2D binary pattern", ITE Transactions on Media Technology and Applications 8(3):170–185 (2020)
- 4) Y. Kakui, K. Araki, C. Han, S. Fukushima, and T. Naemura : "Using a dual-camera smartphone to recognize imperceptible 2d barcodes embedded in videos", In Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology UIST '22 Adjunct pp. 1–3 (2022)
- 5) C. Plain-Jones : "Data Matrix identification", Sensor Review



(a) Hyperlink to the artist's website in music video



(b) Instant closed captions on smartphone screen (left: English, right: Japanese)



(c) Landmark information spatially embedded in the sightseeing video.

Fig. 12 Application examples of extracting information embedded in videos

invisible information into videos, thereby demonstrating the feasibility of our proposed system.

Moreover, we believe that exploring the adaptability of this system under various lighting conditions and with different displays and smartphone combinations will further demonstrate its versatility in real-world scenarios. In addition, the viewing experience can be improved by expanding the gamut of modulated videos.

Acknowledgements

This project is supported by Nakayama Future Factory.

Appendix

A. Parameters used in the encoding process

As mentioned in Section 4, the vibration radius (r) and angle (θ) are the parameters used for color modulation. In particular, a large vibration radius facilitates decoding; however, if this parameter becomes excessively large, the modulation becomes perceptible to the human eye. Experiments were conducted to determine these two parameters prior to determining the capturing parameters because the amount of modulation must be sufficiently large for stable decoding.

We used a still image of fruit (Fig. 1(a)) as the original content, with a QR code (Fig. 1(b)) embedded as invisible information. The gamut of the fruit image was narrowed to 57–198 to equalize the number of tones using the previous method proposed by Abe et al. The measurements were performed under four vibration angle conditions: $\theta = 0^\circ, 45^\circ, 90^\circ$, and 135° , and the vibration radius ranged from $r = 5$ to 40 in five increments. In the $\theta = 0^\circ$ and 90° conditions, only the component corresponding to the vibration direction was used for decoding. Decoding was consecutively conducted over 50 iterations under each condition, and this process was repeated over 20 iterations to obtain the average decoding rate. The capture rate was 24 fps, the same as the previous method by Abe et al. The remaining capture parameters were set to the default parameters described in Section 4. Note that the decoding rates for $d = 35$ and 40 were not measured for $\theta = 0^\circ$, because the image could only be vibrated under conditions up to $d = 30$.

Fig. 2 shows the decoding rates measured for each condition. We decided to use the condition with the smallest vibration radius, $d = 25$ and $\theta = 135^\circ$, from among the conditions that achieved a decoding rate greater than 95 %.

References

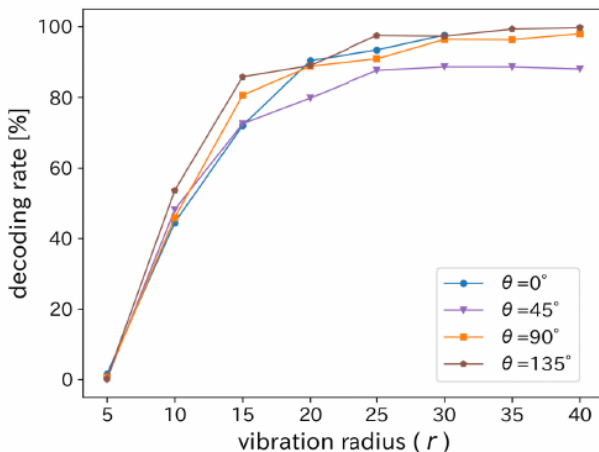
- 1) P. H. Pathak, X. Feng, P. Hu, and P. Mohapatra : "Visible light communication, networking, and sensing: A survey, potential and challenges", *IEEE communications surveys & tutorials* 17(4):2047–2077 (2015)
- 2) International Organization for Standardization :, "Information technology — Automatic identification and data capture techniques — QR Code bar code symbology specification" (2015)
- 3) S. Abe, T. Hiraki, S. Fukushima, and T. Naemura : "Imperceptible color vibration for screen-camera communication via 2D binary pattern", *ITE Transactions on Media Technology and Applications* 8(3):170–185 (2020)
- 4) Y. Kakui, K. Araki, C. Han, S. Fukushima, and T. Naemura : "Using a dual-camera smartphone to recognize imperceptible 2d barcodes embedded in videos", In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology UIST '22 Adjunct* pp. 1–3 (2022)
- 5) C. Plain-Jones : "Data Matrix identification", *Sensor Review*



(a) fruits [30]



(b) QR code ("test", version 1, error correction level L) 変調パラメータ決定実験に使用した画像とQRコード



app.Fig. 2 振動半径と振動角度を変化させた場合のデコーディング率

- (1995)
- 6) T. Hao, R. Zhou, and G. Xing : "COBRA: Color barcode streaming for smartphone systems", In Proceedings of the 10th international conference on Mobile systems, applications, and services pp. 85–98 (2012)

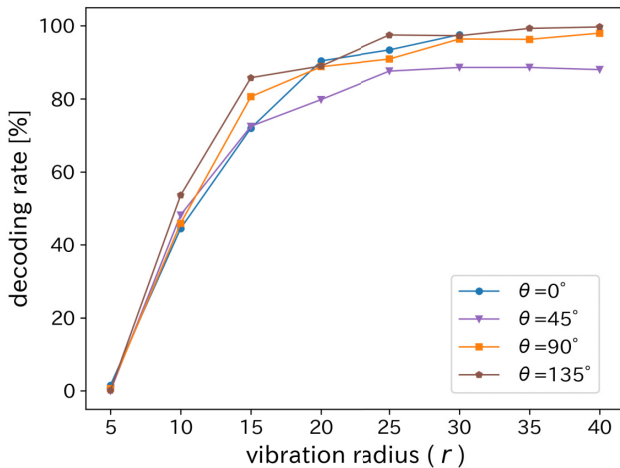
- 7) Q. Wang, M. Zhou, K. Ren, T. Lei, J. Li, and Z. Wang : "Rain bar: Robust application-driven visual communication using color barcodes", In 2015 IEEE 35th International Conference on Distributed Computing Systems pp. 537–546 (2015)
- 8) Hao Han, Kunming Xie, Tongyu Wang, Xiaojun Zhu, Yanchao Zhao, and Fengyuan Xu : "Rescqr: Enabling reliable data recovery in screen-camera communication system", IEEE Transactions on Mobile Computing pp. 1–13 (2023)
- 9) H. K. Chu, C. S. Chang, R. R. Lee, and N. J. Mitra : "Halftone QR codes", ACM Transactions on Graphics (TOG) 32(6):1–8 (2013)
- 10) M. Xu, Q. Li, J. Niu, H. Su, X. Liu, W. Xu, P. Lv, B. Zhou, and Y. Yang : "ART-UP: A novel method for generating scanning-robust aesthetic QR codes", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 17(1):1–23 (2021)
- 11) C. Chen, W. Huang, B. Zhou, C. Liu, and W. H. Mow : "PiCode: A new picture-embedding 2D barcode", IEEE Transactions on Image Processing 25(8):3444–3458 (2016)
- 12) C. Chen, B. Zhou, and W. H. Mow : "RA code: A robust and aesthetic code for resolution-constrained applications", IEEE Transactions on Circuits and Systems for Video Technology 28(11):3300–3312 (2017)
- 13) C. Chen, W. Huang, L. Zhang, and W. H. Mow : "Robust and unobtrusive display-to-camera communications via blue channel embedding", IEEE Transactions on Image Processing 28(1):156–169 (2018)
- 14) Z. Yang, Y. Bao, C. Luo, X. Zhao, S. Zhu, C. Peng, Y. Liu, and X. Wang : "ARTcode: preserve art and code in any image", In Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing pp. 904–915 (2016)
- 15) W. Yuan, K. Dana, A. Ashok, M. Gruteser, and N. Mandayam : "Dynamic and invisible messaging for visual MIMO", In 2012 IEEE Workshop on the Applications of Computer Vision (WACV) pp. 345–352 (2012)
- 16) T. Li, C. An, X. Xiao, A. T. Campbell, and X. Zhou : "Real-time screen-camera communication behind any scene", In Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services pp. 197–211 (2015)
- 17) M. Izz, Z. Li, H. Liu, Y. Chen, and F. Li : "Uber-in-light: Unobtrusive visible light communication leveraging complementary color channel", In IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications pp. 1–9 (2016)
- 18) E. Simonson and J. Brozek : "Flicker fusion frequency: background and applications", Physiological reviews 32(3):349–378 (1952)
- 19) A. Wang, C. Peng, O. Zhang, G. Shen, and B. Zeng : "In-frame: Multiflexing full-frame visible communication channel for humans and devices", In proceedings of the 13th ACM Workshop on Hot Topics in Networks pp. 1–7 (2014)
- 20) A. Wang, Z. Li, C. Peng, G. Shen, G. Fang, and B. Zeng : "In-Frame++: Achieve simultaneous screen-human viewing and hidden screen-camera communication", In Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services pp. 181–195 (2015)
- 21) V. Nguyen, Y. Tang, A. Ashok, M. Gruteser, K. Dana, W. Hu, E. Wengrowski, and N. Mandayam : "High-rate flicker-free screen-camera communication with spatially adaptive embedding", In IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications pp. 1–9 (2016)
- 22) Kai Zhang, Yi Zhao, Chenshu Wu, Chaofan Yang, Kehong Huang, Chunyi Peng, Yunhao Liu, and Zheng Yang : "Chroma-code: A fully imperceptible screen-camera communication system", IEEE Transactions on Mobile Computing 20(3):861–876 (2021)
- 23) Y. Jiang, K. Zhou, and S. He : "Human visual cortex responds to invisible chromatic flicker", Nature neuroscience 10(5):657–662 (2007)
- 24) G. Woo, A. Lippman, and R. Raskar : "VRCodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter", In 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) pp. 59–64 (2012)
- 25) G. Yamamoto, L. Sampaio, T. Taketomi, C. Sandor, H. Kato, and T. Kuroda : "Imperceptible on-screen markers for mobile interaction on public large displays", IEICE TRANSACTIONS on Information and Systems 100(9):2027–2036 (2017)
- 26) J. Klein, J. Xu, C. Brauers, J. Jochims, and R. Kays : "Investi-



(a) fruits [30]



(b) QR code ("test", version 1, error correction level L)

app.Fig. 1 Image and QR code used in the experiment of modulation parameter decision**app.Fig. 2** Decoding rates at varying vibration radii and angles

- (1995)
- 6) T. Hao, R. Zhou, and G. Xing : "COBRA: Color barcode streaming for smartphone systems", In Proceedings of the 10th international conference on Mobile systems, applications, and services pp. 85–98 (2012)

- 7) Q. Wang, M. Zhou, K. Ren, T. Lei, J. Li, and Z. Wang : "Rain bar: Robust application-driven visual communication using color barcodes", In 2015 IEEE 35th International Conference on Distributed Computing Systems pp. 537–546 (2015)
- 8) Hao Han, Kunming Xie, Tongyu Wang, Xiaojun Zhu, Yanchao Zhao, and Fengyuan Xu : "Rescqr: Enabling reliable data recovery in screen-camera communication system", IEEE Transactions on Mobile Computing pp. 1–13 (2023)
- 9) H. K. Chu, C. S. Chang, R. R. Lee, and N. J. Mitra : "Halftone QR codes", ACM Transactions on Graphics (TOG) 32(6):1–8 (2013)
- 10) M. Xu, Q. Li, J. Niu, H. Su, X. Liu, W. Xu, P. Lv, B. Zhou, and Y. Yang : "ART-UP: A novel method for generating scanning-robust aesthetic QR codes", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 17(1):1–23 (2021)
- 11) C. Chen, W. Huang, B. Zhou, C. Liu, and W. H. Mow : "PiCode: A new picture-embedding 2D barcode", IEEE Transactions on Image Processing 25(8):3444–3458 (2016)
- 12) C. Chen, B. Zhou, and W. H. Mow : "RA code: A robust and aesthetic code for resolution-constrained applications", IEEE Transactions on Circuits and Systems for Video Technology 28(11):3300–3312 (2017)
- 13) C. Chen, W. Huang, L. Zhang, and W. H. Mow : "Robust and unobtrusive display-to-camera communications via blue channel embedding", IEEE Transactions on Image Processing 28(1):156–169 (2018)
- 14) Z. Yang, Y. Bao, C. Luo, X. Zhao, S. Zhu, C. Peng, Y. Liu, and X. Wang : "ARTcode: preserve art and code in any image", In Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing pp. 904–915 (2016)
- 15) W. Yuan, K. Dana, A. Ashok, M. Gruteser, and N. Mandayam : "Dynamic and invisible messaging for visual MIMO", In 2012 IEEE Workshop on the Applications of Computer Vision (WACV) pp. 345–352 (2012)
- 16) T. Li, C. An, X. Xiao, A. T. Campbell, and X. Zhou : "Real-time screen-camera communication behind any scene", In Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services pp. 197–211 (2015)
- 17) M. Izz, Z. Li, H. Liu, Y. Chen, and F. Li : "Uber-in-light: Unobtrusive visible light communication leveraging complementary color channel", In IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications pp. 1–9 (2016)
- 18) E. Simonson and J. Brozek : "Flicker fusion frequency: background and applications", Physiological reviews 32(3):349–378 (1952)
- 19) A. Wang, C. Peng, O. Zhang, G. Shen, and B. Zeng : "In-frame: Multiflexing full-frame visible communication channel for humans and devices", In proceedings of the 13th ACM Workshop on Hot Topics in Networks pp. 1–7 (2014)
- 20) A. Wang, Z. Li, C. Peng, G. Shen, G. Fang, and B. Zeng : "In-Frame++: Achieve simultaneous screen-human viewing and hidden screen-camera communication", In Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services pp. 181–195 (2015)
- 21) V. Nguyen, Y. Tang, A. Ashok, M. Gruteser, K. Dana, W. Hu, E. Wengrowski, and N. Mandayam : "High-rate flicker-free screen-camera communication with spatially adaptive embedding", In IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications pp. 1–9 (2016)
- 22) Kai Zhang, Yi Zhao, Chenshu Wu, Chaofan Yang, Kehong Huang, Chunyi Peng, Yunhao Liu, and Zheng Yang : "Chroma-code: A fully imperceptible screen-camera communication system", IEEE Transactions on Mobile Computing 20(3):861–876 (2021)
- 23) Y. Jiang, K. Zhou, and S. He : "Human visual cortex responds to invisible chromatic flicker", Nature neuroscience 10(5):657–662 (2007)
- 24) G. Woo, A. Lippman, and R. Raskar : "VRCodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter", In 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) pp. 59–64 (2012)
- 25) G. Yamamoto, L. Sampaio, T. Taketomi, C. Sandor, H. Kato, and T. Kuroda : "Imperceptible on-screen markers for mobile interaction on public large displays", IEICE TRANSACTIONS on Information and Systems 100(9):2027–2036 (2017)
- 26) J. Klein, J. Xu, C. Brauers, J. Jochims, and R. Kays : "Investi-

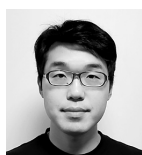
- gations on temporal sampling and patternless frame recovery for asynchronous display-camera communication”, IEEE Transactions on Circuits and Systems for Video Technology 32(6):4004–4015 (2021)
- 27) H. Cui, H. Bian, W. Zhang, and N. Yu : “UnseenCode: Invisible on-screen barcode with image-based extraction”, In IEEE INFOCOM 2019-IEEE Conference on Computer Communications pp. 1315–1323 (2019)
- 28) N. Otsu : “A threshold selection method from gray-level histograms”, IEEE Transactions on Systems, Man, and Cybernetics 9(1):62–66 (1979)
- 29) Johannes Kiess :, “Test sequences” (2011)
- 30) G. Bradski : “The OpenCV Library”, Dr. Dobb’s Journal of Software Tools (2000)



Kakui Yuki received B.S. and M.S. degrees from the University of Tokyo, Japan, in 2021 and 2023, respectively. Her research interests include ubiquitous computing and human-computer interaction.



Kota Araki received a B.S. degree from the University of Tokyo, Japan, in 2022, where he is currently pursuing a master’s degree at the Graduate School of Information Science and Technology. His research interests include ubiquitous computing and augmented reality.



Changyo Han received a Ph.D. in Information Science and Technology from the University of Tokyo, Japan, in 2020. He is a Japan Society for the Promotion of Science (JSPS) research fellow at the University of Tokyo, Japan. He is currently an assistant professor with the Interfaculty Initiative in Information Studies at the University of Tokyo, Japan. His research interests include human-computer interaction, digital fabrication, and ubiquitous computing. He is a member of ACM.



Shogo Fukushima is an Associate Professor at the Graduate School and Faculty of Information Science and Electrical Engineering at Kyushu University. He received a Ph.D. degree in Engineering from the University of Electro-Communications in 2013. His research interests include intelligence amplification (IA), virtual reality, entertainment computing, and human emotions.



Takeshi Naemura received a Ph.D. in Electrical Engineering from the University of Tokyo, Japan, in 1997. He is currently a Professor with the Interfaculty Initiative in Information Studies, the University of Tokyo. He was a Visiting Assistant Professor of computer science with Stanford University, USA, supported by the “Japan Society for Promotion of Science (JSPS) Postdoctoral Fellowships for Research Abroad” from 2000 to 2002. His research interests include virtual reality and human interfaces. He is a member of ITE, IEICE, and ACM.

- gations on temporal sampling and patternless frame recovery for asynchronous display-camera communication”, IEEE Transactions on Circuits and Systems for Video Technology 32(6):4004–4015 (2021)
- 27) H. Cui, H. Bian, W. Zhang, and N. Yu : “UnseenCode: Invisible on-screen barcode with image-based extraction”, In IEEE INFOCOM 2019-IEEE Conference on Computer Communications pp. 1315–1323 (2019)
 - 28) N. Otsu : “A threshold selection method from gray-level histograms”, IEEE Transactions on Systems, Man, and Cybernetics 9(1):62–66 (1979)
 - 29) Johannes Kiess :, “Test sequences” (2011)
 - 30) G. Bradski : “The OpenCV Library”, Dr. Dobb’s Journal of Software Tools (2000)



Kakui Yuki received B.S. and M.S. degrees from the University of Tokyo, Japan, in 2021 and 2023, respectively. Her research interests include ubiquitous computing and human-computer interaction.



Kota Araki received a B.S. degree from the University of Tokyo, Japan, in 2022, where he is currently pursuing a master’s degree at the Graduate School of Information Science and Technology. His research interests include ubiquitous computing and augmented reality.



Changyo Han received a Ph.D. in Information Science and Technology from the University of Tokyo, Japan, in 2020. He is a Japan Society for the Promotion of Science (JSPS) research fellow at the University of Tokyo, Japan. He is currently an assistant professor with the Interfaculty Initiative in Information Studies at the University of Tokyo, Japan. His research interests include human-computer interaction, digital fabrication, and ubiquitous computing. He is a member of ACM.



Shogo Fukushima is an Associate Professor at the Graduate School and Faculty of Information Science and Electrical Engineering at Kyushu University. He received a Ph.D. degree in Engineering from the University of Electro-Communications in 2013. His research interests include intelligence amplification (IA), virtual reality, entertainment computing, and human emotions.



Takeshi Naemura received a Ph.D. in Electrical Engineering from the University of Tokyo, Japan, in 1997. He is currently a Professor with the Interfaculty Initiative in Information Studies, the University of Tokyo. He was a Visiting Assistant Professor of computer science with Stanford University, USA, supported by the “Japan Society for Promotion of Science (JSPS) Postdoctoral Fellowships for Research Abroad” from 2000 to 2002. His research interests include virtual reality and human interfaces. He is a member of ITE, IEICE, and ACM.