

動画に重畳した不可視マーカの頑健な認識—デュアルカメラによる同時撮影—

Robust Recognition of Imperceptible Markers Embedded in Videos: A Dual Camera System for Simultaneous Capturing

覚井 優希* 荒木 航太* 韓 燦教* 福嶋 政期† 苗村 健*
 Yuki Kakui Kota Araki Changyo Han Shogo Fukushima Takeshi Naemura

1 はじめに

近年、ディスプレイの増加やスマートフォンの普及に伴い、スクリーン-カメラ間通信が注目されている。中でも QR コード [1] に代表される可視のコードが広く用いられているが、これらは人間にとっては映像コンテンツの視聴体験の妨げとなりうる。これに対し、ユーザーには普段通りの映像視聴体験を提供し、同時にサイドチャンネルとして情報を送ることを可能にする、不可視のスクリーン-カメラ間通信の研究が進められている。応用先としては、不可視の情報が埋め込まれたビデオ広告にユーザーがスマートフォンをかざすことで、商品の販売ページに飛ぶようなものが想定される。

不可視のスクリーン-カメラ間通信の実現方法の一つに、高速な色度変化を利用したものがある。人間は 25Hz 程度以上の色度変化を知覚できないという性質を持ち [2]、一般的なディスプレイでも不可視性を実現することができる。この特性を利用して、不可視の色振動 [3] では、情報ビット 1 を埋め込みたい画素値に対して、XYZ 色空間中の XZ 平面で対称の位置にある 2 色を交互に高速に表示した。復号の際は、色度変調した 2 つのフレームの差分を取ることで、埋め込まれたコードを抽出することができる。24 [fps]、露光時間 1/120 [s] に設定したカメラで連続撮影した画像について、2 フレーム前との絶対差分を取ったものを 2 つ足し合わせることで、スマートフォンでのリアルタイムの復号およびローリングシャッターへの対処を実現した。荒木らは、不可視の色振動の動画対応を目的として、埋め込む情報の 0/1 の画素をそれぞれ逆位相で振動させる手法を提案した [4]。受信側では符号付きの差分を用いて復号を行うことで、0/1 の間により大きな色差を持たせることができるため、動画コンテンツでの復号率が向上した。

本研究では、情報埋め込み手法は従来の不可視の色振動と変えず、受信側でスマートフォン背面の 2 つのカメ

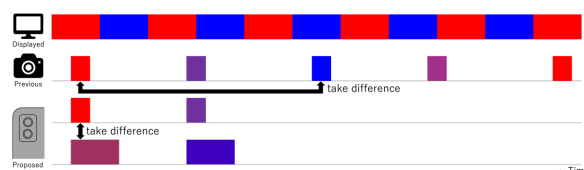


図 1: 従来手法と提案手法の撮影方法。提案手法ではスマートフォンのデュアルカメラによる同時撮影を行う。

ラを利用し、露光時間を異ならせて同時に撮影を行う手法を提案する。PC 上でのシミュレーション実験に基づき、提案手法の有効性と限界を論じる。

2 提案

本研究では、図 1 に示すようにスマートフォン背面の 2 つのカメラを利用し、異なる露光時間で同時に撮影を行うことで、同一時刻の撮影で異なる見えの画像をキャプチャする手法を提案する。露光時間を異ならせることで、2 つのカメラ間で異なる色度変調量の画像がキャプチャされる。これらの差分をとることにより、従来よりも時間軸上で近いフレーム同士で差分を取ることができる。そのため、元の映像コンテンツの時間変化に影響を受けにくい復号が可能になると期待できる。提案手法の流れは次の通りである。

1. カメラ間の色味の差の補正
2. 露光時間を変えた 2 カメラでの同時撮影
3. カメラ間の画角の差の補正
4. 復号処理

4 の復号処理は、従来手法に倣い、撮影した 2 フレームの輝度を除いた成分の絶対差分を足し合わせ、ノイズ軽減のためにぼかしをかけてから、大津の閾値判別法を用いて二値化を行う。

ここで、色振動の振幅値を ΔI として、撮影時に得られる色差について述べる。簡単のため、フレーム間差分のない静止画であると仮定する。また、色振動による変調後の画像をそれぞれプラス変調画像、マイナス変調画

* 東京大学 The University of Tokyo

† 九州大学 Kyushu University

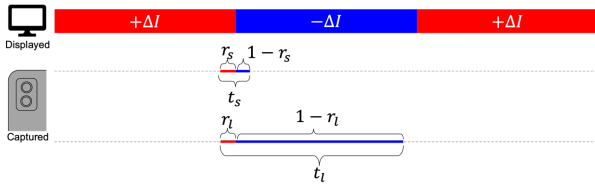


図 2: 同時撮影によって得られる静止画における色度差

像と呼ぶ。従来手法では、プラス変調画像からマイナス変調画像を引くため、色差は常に $2\Delta I$ となる。一方、提案手法では、図 2 に示すように、露光時間が短いカメラと長いカメラについて、プラス変調画像の表示中に撮影が始まるとして、プラス変調画像の撮影時間の割合をそれぞれ r_s, r_l とすると、色差は次のように表される。

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

このように、提案手法の撮影方式で得られる色差は、露光時間と撮影開始タイミングによって決まる、2つのカメラそれぞれの変調画像を撮影する時間の比に依存する。

3 シミュレーション実験

カメラ間の色味や画角の補正の必要性、また、ノイズやローリングシャッターの影響のない理想的な条件を想定したシミュレーション実験を行った。これは、2章の提案手法の流れのうち、2の同時撮影と4の復号処理を再現するものである。

3.1 実験方法

映像コンテンツとして、[5], [6] で提供されているビデオシーケンスの中から図 3(a)~(g) に示す 7 本の動画を使用した。埋め込んだ情報は図 3(i) に示す QR コード (文字列: “<https://www.nae-lab.org/>”, 誤り訂正レベル: L) である。

最初に、動画の全体的な復号率を調べるための実験を行った。ここで、復号率は次のように求めるものとする。

$$\text{復号率} = \frac{\text{正しくデコードされたフレームペアの数}}{\text{復号処理を行なったフレームペアの数}}$$

従来手法では、実質的には送信側で 5 フレーム前との差分をとっていることになる。そこで、動画中の始めの 5 フレームを除くすべてのフレームに対し、図 4(a) に示すように 5 フレーム前との差分をとり、復号処理を行い、QR コードを読み取れたら復号成功として復号率を計測した。ただし、図 4 は色の変調がわかりやすいように強調して表示している。なお、従来、提案ともにディスプレイのガンマ補正の影響を組み込んで実験を行った。一方、提案手法については、図 4(b) のように、2つのカ



(a) big_buck_bunny_scene [5] (b) football [5] (c) walking [5]



(d) MINECRAFT [6] (e) GTAV [6] (f) highway [5]



(g) old_town_cross [6] (h) baboon (i) QR コード

図 3: シミュレーション実験に用いた動画像と埋め込んだ QR コード

メラの露光時間を指定したうえで、露光時間と撮影開始時刻に従って隣接する各フレームの撮影時間を求め、それに応じて各フレームを混合したフレームを作成することで同時撮影を再現した。撮影開始時刻は、送信側と受信側で同期が取れないことからランダムに設定する必要がある。そのため、送信側の何フレーム目で撮影を行うか、また、そのフレームに切り替わってから撮影が開始されるまでの時間 t [s] ($0 < t < 1/60$) を一様分布の乱数で決定した。今回設定した露光時間のパラメータでは最大 3 フレームが混合されるため、最後の 2 フレームを除くすべてのフレームから始まる 3 連続のフレームセットを対象とした。そして、露光時間が短いカメラのキャプチャ画像と長いカメラのキャプチャ画像をそれぞれ再現した混合フレーム同士の差分をとり、復号処理を行った。2つのカメラの露光時間のパラメータは、短い方が $1/250$ [s] または $1/120$ [s] の 2 条件、長い方が $1/60$, $1/50$, $1/45$, $1/40$, $1/30$ [s] の 5 条件とした。また、計測は一条件につき 1000 回の計測を 5 度実施した。

3.2 実験結果

まず、長い方の露光時間が $1/50$ [s] の場合における、短い方の露光時間を変えたときの復号率を計測した。結果は図 5 の通りである。エラーバーは標準偏差を示す。各動画の結果について検定を行ったところ、全てのペアについて、有意水準 5% で有意差が見られ、露光時間 $1/250$ [s] の場合の方が復号率が高いことがわかった。

次に、短い方の露光時間を $1/250$ [s] に固定し、長い方の露光時間を $1/60$ [s] から $1/30$ [s] まで変化させた時の復号率の結果を図 6 に示す。エラーバーは標準偏差である。但し、従来手法では復号率にばらつきは生じない

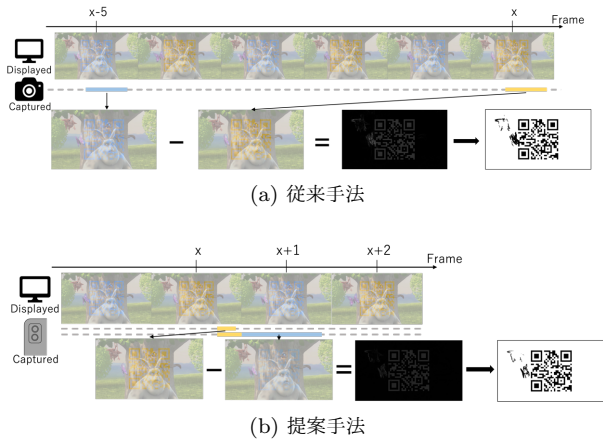


図 4: シミュレーション実験における撮影方法の再現

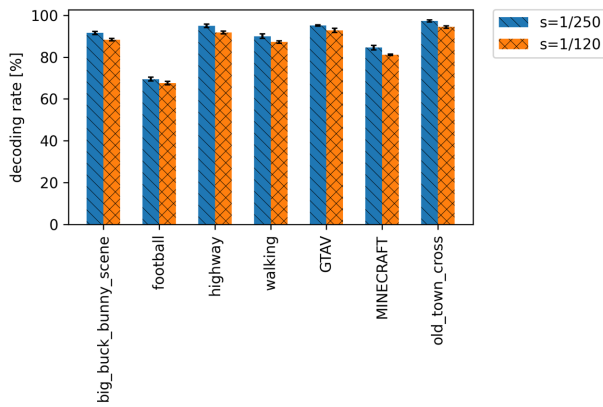


図 5: 短い方の露光時間と復号率の関係。エラーバーは標準偏差を示す。

ため、エラーバーの表示はない。図 6(h) に示す凡例のように、従来手法と比較して有意水準 5% で有意に高い場合は斜線、低い場合はドットで表示した。7 本の動画のうち (a), (b), (c), (d) の 4 本では、長い露光時間がどの値であっても従来手法よりも高い復号率であった。(e), (f) についても露光時間の設定によっては従来手法に勝る復号率となった。一方、(g) では従来手法よりも常に復号率が低かった。また、今回実験した条件の中では、いずれの動画についても、露光時間が 1/50 [s] の時よりも有意水準 5% で有意に復号率が高くなるような露光時間は存在しなかった。

4 考察

4.1 露光時間に関する考察

図 5 に示す結果から、短い方の露光時間については、短ければ短いほど復号率が高くなると考えられる。これは、露光時間が短いほど、露光時間の長いカメラが撮影する画像との情報埋め込み部分の色差が大きくなる傾

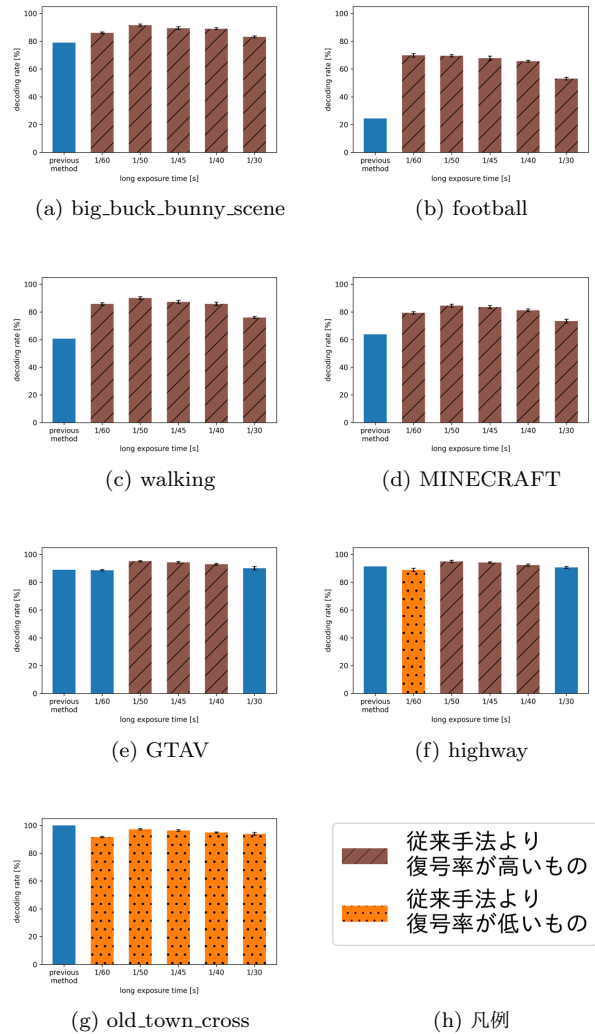


図 6: 長い方の露光時間と復号率の関係。エラーバーは標準偏差を示す。

向があるためである。一方、長い方の露光時間は長ければ長いほど良いとはいえない。1/30 [s] が送信側の 2 フレームに相当する長さであるため、それを超えて得られる色差はそれ以下の露光時間で得られるものより大きくはならず、単にフレーム間差分による悪影響が大きくなることその理由である。逆に 1/60 [s] よりも短くなると、露光時間の長い方のカメラの撮影がフレーム間にまたがらず、短い方のカメラと同一の画像を撮影する確率が高くなるため、復号率は低下する。そのため、復号率のグラフが図 6 に示したような山なりの形状になったと考えられる。以降、長い方の露光時間 t_l は $1/60 < t_l \leq 1/30$ [s] であるとして考察を進める。

4.2 撮影開始タイミングに関する考察

(1) 式から、プラス変調画像とマイナス変調画像の撮影時間の比が 2 つのカメラで等しくなるような撮影開始時刻では、色差が 0 となることわかる。これは、長

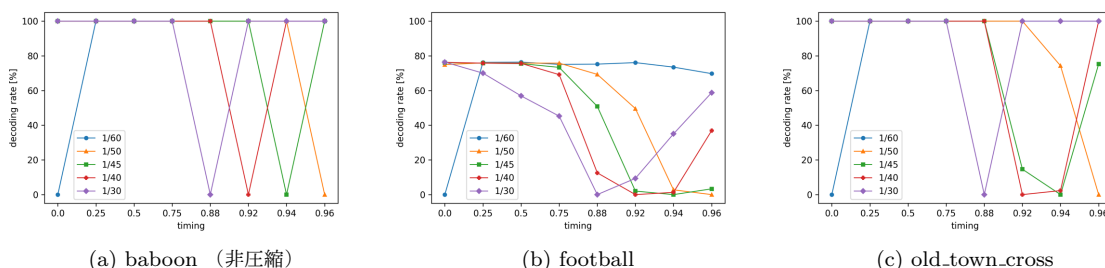


図 7: 撮影開始時刻と復号率の関係

い方のカメラの撮影が 3 フレームにまたがる場合に生じる。このとき、 t_s, t_l をそれぞれ短い方と長い方の露光時間とすると、露光時間が長い方のカメラがプラス変調画像とマイナス変調画像を撮影する時間はそれぞれ $t_l - 1/60$ [s], $1/60$ [s] である。露光時間が短い方のカメラが、長い方のカメラと同じ比率で変調画像を撮影する場合、プラス変調画像を撮影する時間は、 $t_s \cdot \frac{t_l - 1/60}{t_l}$ である。すなわち、ディスプレイ側のフレーム切り替え時刻から、1 フレームの表示時間に対して、

$$1 - \frac{t_s \cdot (1 - \frac{1}{60t_l})}{\frac{1}{60}} \quad (2)$$

の割合だけ過ぎた時刻から撮影が開始された場合、2 つのカメラのキャプチャ画像の差分をとった時に得られる色差は 0 となる。そこで、撮影開始タイミングを一定にして復号率の計測を行った。撮影開始タイミングはフレーム投影時間の 0, 0.25, 0.5, 0.75, 0.88, 0.92, 0.94, 0.96 だけ過ぎたタイミングの計 8 条件である。計測にあたり、短い方の露光時間は $1/250$ [s] に固定し、3 章のシミュレーション実験と同様に、長い方の露光時間を 5 条件として混合フレームを作成した。埋め込み元のコンテンツとして、football, old_town_cross に加え、図 3(h) に示す静止画 baboon を用いた。baboon は非圧縮の状態で行なっている。

結果を図 7 に示す。静止画である (a)baboon について、(2) 式によって計算されるように、長い方の露光時間が $1/60, 1/50, 1/45, 1/40, 1/30$ [s] の場合、それぞれ撮影開始タイミングが 0, 0.96, 0.94, 0.92, 0.88 の場合に 2 つのカメラがキャプチャする画像の色差が 0 となり、復号不可となる。それ以外の場合は 0 より大きい色差が得られ、フレーム間差分の影響が全くない静止画では 100% 復号が可能である。一方、(b), (c) に示す通り、埋め込み元の映像コンテンツが動画の場合には、色差が 0 になるような撮影開始タイミングの周辺でも、復号率が下がる傾向にあることがわかる。これは、色振動による色差が映像コンテンツのフレーム間差分よりも小

さくなるからだと考えられる。このように、提案手法には原理的に色差が 0 になるような撮影開始時刻が存在するため、復号率 100% を達成することができない。そのため、元から動画の動きが少なく、従来手法の復号率が 100% であった old_town_cross においては提案手法の復号率の方が小さくなる結果となった。

5 おわりに

本研究では、不可視の色振動の動画対応を目的とし、映像コンテンツのフレーム間差分が復号を阻害するという課題に取り組んだ。スマートフォン背面の 2 つのカメラを利用し、露光時間を変えて同時に撮影を行う手法を提案し、同時撮影と復号処理を PC 上で再現したシミュレーション実験を行なった。従来手法と比較して多くの動画で復号率が向上することを確認した一方で、提案手法では原理的に色差が得られない場合があり、復号率 100% を達成することができないことも明らかになった。今後は受信側のスマートフォンのアプリケーションを実装し、実機実験を行う予定である。

参考文献

- [1] “Automatic identification and data capture techniques — QR code bar code symbology specification”, <https://www.iso.org/standard/62021.html>. 2022 年 5 月 19 日参照.
- [2] Y. Jiang, K. Zhou and S. He: “Human visual cortex responds to invisible chromatic flicker”, *Nature neuroscience*, **10**, 5, pp. 657–662 (2007).
- [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, pp. 170–185 (2020).
- [4] 荒木航太, 田中稜太郎, 覚井優希, 韓燦教, 福嶋政期, 苗村健: “動画に不可視マーカーを埋め込むためのフレーム間差分に頑健な色振動方式の基礎検討”, 映像メディア処理シンポジウム (IMPS 2021), P2-21 (2021).
- [5] “Test sequences”, https://pi4.informatik.uni-mannheim.de/~kiess/test_sequences/download/. 2022 年 5 月 19 日参照.
- [6] “Xiph.org video test media [derf’s collection]”, <https://media.xiph.org/video/derf/>. 2022 年 5 月 19 日参照.