

視覚特性を考慮した時空間符号化画像提示と復号観測

梅津 潤也[†] 坂上 文彦[†] 佐藤 淳[‡]名古屋工業大学[‡]

1. はじめに

安全な通信手段の確保は現代社会における最重要課題の一つであり、数多くの方法が研究・実用化されている。しかし、これらの方法の多くは主たる通信経路となるネットワーク上の脅威に対抗するものであり、ショルダーハッキングと呼ばれる画面やキーボードを直接のぞき見するような方法に対しては、対抗することができない。このような問題に対応するために、本研究では通常のディスプレイと、透過液晶などの光線の透過を制御可能なメガネを併用することで、特殊な透過率パターンの下で観測・撮影を行った場合にのみ画像の復号を行うことができる、符号化画像提示法を提案する。

2. 時空間積分の観測のモデル化

まず、本研究で用いる画像の観測モデルについて述べる。本研究ではディスプレイ上に時間的に変化する画像が提示され、これをカメラや眼球といったレンズを含む観測系により観測する。このような場合、ディスプレイから照射された光はレンズによって屈折され、図1に示すように光線が撮像面上の1点に集光する、すなわち空間的に積分されることで観測が行われる。

ここで眼鏡を装着することでレンズの焦点距離が変更された場合を考える。このような場合、ディスプレイから照射された光が撮像面上の1点に集光せず、図2のように複数の画素に入射することになる、このような光線の広がりやPSF(Point Spread Function:点広がり関数)と呼ばれる光の広がりを表す関数により表現される。いま、ディスプレイ上の提示画像 D とし、また、ディスプレイ上の一点のみが発光しているシーンを観測した画像をPSF画像 p とする。このとき、観測結果 I は以下のように表すことができる。

$$I = D * p$$

ここで $*$ は畳み込み積分を表す。この式により、光の空間的な広がりを表現することができる。

次に、観測における光の時間積分について述べる。一般的な観測においては、上記の空間積分によりある一瞬の光のみにより観測を行って

Image Encoding Display and Decoding Observation for Restricted Image Presentation based on Time varying Coded Aperture

[†] Junya Umetsu, Fumihiko Sakaue and Jun Sato,

[‡] Nagoya Institute of Technology

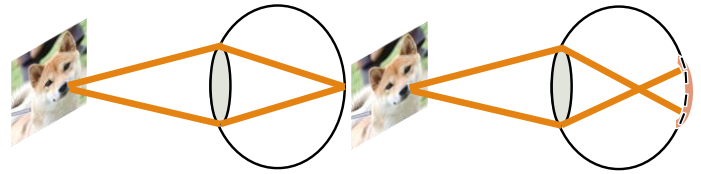


図1 正視

図2 焦点距離がずれた場合

いるのではなく、ある露光時間中に入射した光の時間的に積分することで観測が行われている。そのため、この時間よりも短い時間で光が明滅していたとしても人間の目ではこれを捉えることができず、連続的に光が発せられているものととらえられる。いま、時刻 t におけるディスプレイ上の提示画像を D_t とした場合、観測結果 I_t は以下のように表現できる。

$$I_t = \int_{t-T}^t D_{t'} * p dt' \quad (1)$$

ここで、 T は観測系の露光時間であり、 D_t は時刻 t にディスプレイ上に表示された画像である。この式は、光の時間的、空間的な積分を表現しており、本研究ではこの式により表される観測モデルに従って画像を符号化・復号する方法を検討する。

3. 時空間符号化を用いた多重画像提示

次に、眼鏡上のマスクパターンを高速に変化させることで、符号化を行いながら観測を行う場合を考える。本研究では眼鏡が十分に眼球に近い位置に装着されているものとし、マスク上に提示したパターンがそのまま画像ボケのPSFとして観測されると仮定する。まず正視、すなわち眼球に入射した光が1点に集光される場合を考える。このような場合ディスプレイ上に提示されたパターンに対して時間積分のみが適用され、以下の式により観測 I'_t が決定される。

$$I'_t = \sum_{t'=t-T+1}^t D_{t'} \quad (2)$$

一方、マスクと屈折レンズを用いる場合、時間積分と同時に空間積分が発生する。ここで時刻 t における透過率パターンにより得られるPSFを p_t 、ディスプレイに提示する画像を D_t とすると、観測画像 I_t は以下の式により計算される。

$$I_t = \sum_{t'=t-T+1}^t D_{t'} * p_{t'} \quad (3)$$

このような観測は、時間的に符号化された画像

を高速にマスクが変化する眼鏡により復号しているものと捉えることができるため、本研究ではこれを復号観測と呼ぶ。したがって、復号観測を行った場合と正視で観測を行った場合にそれぞれ異なる画像を観測されることが本研究の目的となる。

4. 提示画像とマスクパターンの推定

前節で述べた復号観測を実現するため、符号化された提示画像と復号のためのマスクパターンを推定する方法について述べる。いま、正視でディスプレイを観測する場合には画像 I'_o を、復号観測を行う場合には画像 I_o (秘匿画像) を観測させることを考える。この場合、それぞれの観測画像と目標画像 I_o , I'_o との差分が最小となるように、提示画像とマスクパターンを決定する。つまり、最適なパターン D_t , p_t を以下の式により求める。

$$\{\widehat{D}_t, \widehat{p}_t\} = \operatorname{argmin}_{D_t, p_t} \|I_o - I_t\|^2 + \|I'_o - I'_t\|^2 \quad (4)$$

しかし、式(4)で示す最小2乗基準により画像の推定を行った場合、推定画像がやや不鮮明になるといった性質がある。そのため、人間の目がエッジなどの輝度変化に敏感に反応する性質があることを利用し、目標画像と観測画像の高周波成分が近くなるように以下の評価式 F を導入する。

$$F = \|H(I_o) - H(I_t)\|^2 + \|H(I'_o) - H(I'_t)\|^2 \quad (5)$$

ここで、 $H(I)$ は画像 I の周波数成分を抽出するハイパスフィルタである。

さらに、提示画像に含まれる秘匿情報が提示画像中に直接的に表されないようにするために、提示画像が目標画像から遠ざかるような評価式 N を導入する。この評価式では画像識別ニューラルネットワーク (VGG16) に入力し、その中間層出力 $V(I)$ を画像の意味を表す特徴ベクトルと考え、以下の様に定義する。

$$N = \sum_{t'} \gamma S(V(I_o), V(D_{t'})) - \delta \|Z(V(I_o), V(D_{t'}))\|^2 \quad (6)$$

ここで、 $Z(X, Y)$ はZNCC, $S(X, Y)$ はScale Invariant MSE[1]である。以上の評価式を含め、最適なパターン $\widehat{D}_t, \widehat{p}_t$ は以下のように決定される。

$$\{\widehat{D}_t, \widehat{p}_t\} = \operatorname{argmin}_{D_t, p_t} \alpha (\|I_o - I_t\|^2 + \|I'_o - I'_t\|^2) + \beta F - N \quad (7)$$

このようにして決定された提示画像およびマスクパターンを用いることで、正視と復号観測にそれぞれ異なる画像を観測させることができる。

4. 実験

4.1 実験環境

提案法により画像を時空間的に符号化・復号が可能かどうかを確認する実験を行った。本実

験では定量的な評価を行うために、人間の目の代用に望遠カメラを用いて撮影を行い、これを評価した。透過マスクはOHPフィルムにあらかじめ印刷しておき、これを取り換えながら撮影を行うことで、マスクパターンの時間変動を表現した。また、マスクパターンとレンズを併用する代わりに、カメラのピントを調整することで画像ボケを発生させた。正視の状態についてはマスクを外してディスプレイにピントを合わせた状態で撮影を行うことで再現した。1度の観測には4枚の画像が観測されるものとし、提示画像およびそれに対応するマスクパターンを提案法により推定した。正視の時に観測させる通常画像および秘匿画像は図5に示す画像を用いた。

4.2 実験結果

提案法により決定された提示画像およびマスクパターンを図3, 図4に示す。また、これらの画像を正視および復号観測により観測して得られた結果を図6に示す。これらの結果を見ると、正視、復号観測どちらの場合でも、目標画像に近い画像を観測できていることが分かる。また、図3に示した提示画像を見ても、この画像から秘匿画像を推測することは困難であり、秘匿画像を適切に秘匿できていることが確認できる。この結果から、提案法により秘匿画像を適切に秘匿しつつ、観測状態に応じて異なる画像を観測させられることが確認できた。

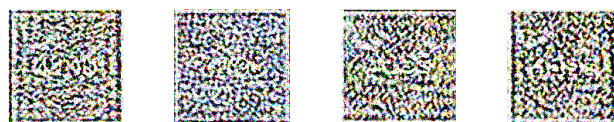


図3 提示画像

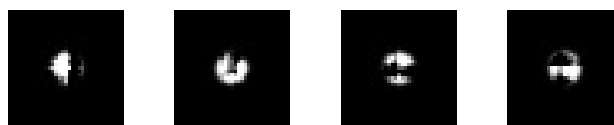


図4 マスク画像



(a) 秘匿

(b) 通常

(a) 秘匿

(b) 通常

図5 目標画像

図6 観測画像

参考文献

[1] David Eigen, Christian Pirsch, Rob Fergus, Depth Map Prediction from a Single Image using a Multi-Scale Deep Network, CVPR(2014)