圧縮センシングによる
動画像の高速撮像に関する研究

提出先　　大阪大学大学院情報科学研究科

提出年月　2023 年 1 月

吉田　道隆

業績リスト

関連発表論文

論文誌

1. Michitaka Yoshida, Toshiki Sonoda, Hajime Nagahara, Kenta Endo, Yukinobu Sugiyama, Rin-ichiro Taniguchi, “High-Speed Imaging Using CMOS Image Sensor With Quasi Pixel-Wise Exposure”, IEEE Transactions on Computational Imaging, Vol.6, pp.463-476, 2019

国際会議 (査読あり)

1. Yoshida Michitaka, Torii Akihiko, Okutomi Masatoshi, Endo Kenta, Sugiyama Yukinobu, Taniguchi Rin-ichiro, Nagahara Hajime, “Joint optimization for compressive video sensing and reconstruction under hardware constraints”, Proceedings of the European Conference on Computer Vision (ECCV2018), Munich, Germany, pp.634-649, 2018.09

国内会議 (査読なし)

1. 吉田 道隆，鳥居 秋彦，奥富 正敏，遠藤 健太，杉山 行信，谷口 倫一郎，長原 一，“DNN により最適化されたピクセルコーディング CMOS イメージセンサによるハイスピード撮像”, 第 22 回画像の認識・理解シンポジウム (MIRU2019), 大阪, 2019 年 7 月

2. 吉田 道隆，鳥居 秋彦，奥富 正敏，遠藤 健太，杉山 行信，谷口 倫一郎，長原 一，“ハードウェアの制約を考慮した圧縮ビデオセンシングにおける圧縮と再構成の同時最適化”, 第 21 回画像の認識・理解シンポジウム (MIRU2018), 札幌, 2018 年 8 月

その他論文

論文誌

1. Sudhakar Kumawat, Tadashi Okawara, Michitaka Yoshida, Hajime Nagahara, Yasushi Yagi,“Action Recognition From a Single Coded Image”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, doi: 10.1109/TPAMI.2022.3196350.

国際会議 (査読あり)

1. Ryoya Mizuno, Keita Takahashi, Michitaka Yoshida, Chihiro Tsutake, Toshiaki Fujii, Hajime Nagahara, “Acquiring a Dynamic Light Field Through a Single-Shot Coded Image”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, U.S.A, 2022.06

2. Tadashi Okawara, Michitaka Yoshida, Hajime Nagahara, Yasushi Yagi,“Action Recognition from a Single Coded Image”, Proceedings of the International Conference on

Computational Photography (ICCP2020), Saint Louis, U.S.A, 2020.04

国内会議 (査読あり)

1. 水野 良哉，高橋 桂太，坂井 康平，都竹 千尋，藤井 俊彰，吉田 道隆，長原 一，“動的光線空間のシングルショット撮影”, 画像の認識・理解シンポジウム (MIRU2021), 名古屋, 2021 年 7 月

概要

光線には空間や時間，色など様々な情報が含まれるため，カメラを用いることで実世界の様々な情報を取得することができる．しかし，通常のイメージセンサは画素が 2 次元に配置されており，すべての情報を取得することは困難である．そこで，従来はカメラの性能を向上させるか，画像取得後の処理によって補完または推定している．しかし，近年ではセンサの性能は物理的な限界に近付きつつあり，また，画像処理においても近年では大量の画像を用意できるようになったためこれ以上統計処理により取り出せる情報量を劇的に増やすことは難しい．しかし，撮影後に画像処理を施すことを前提とするなら従来のように撮像画像の画質や情報量を最適化するのではなく，後処理後の画像の画質や精度を最適化することを考える．この考え方をコンピュテーショナルフォトグラフィといい，近年注目を集めている．

従来のデバイス開発は人間に近い観測を目指し，従来の画像処理では入力は人間が目で見ているような画像であることを前提としている．しかし，そもそも撮影後に画像処理を行う前提なら撮影画像はシーンを均一に撮影する必要はなく，復元に必要な情報を含んでいればよい．そこで，センサ上で必要な情報を圧縮して取得し，撮影後に再構成することで効率のよい情報の取得が行える．この考え方を圧縮センシングと呼び，本研究では従来のカメラシステムの一部を自由に制御可能なものへ変更することで，実用性と汎用性を持たせながら 2 次元のイメージセンサで取得できる情報の拡張に取り組む．

通常のイメージセンサは画素が 2 次元に配置されており，読み出し回路の帯域が制限さ

れているため，動画を撮影する際には空間解像度とフレームレートの間にはトレードオフの関係がある．そのため，高空間解像度で高フレームレートな動画はハイスピードカメラで撮影されが，このような特殊なセンサは非常に高価であり，回路が複雑になることからフォトトランジスタの面積が減少するため感度が悪くなる問題もある．また，通常の撮影画像からフレーム間を補完して高フレームレートな動画を生成する手法が提案されているが，近年では性能が頭打ちになってきている．そこで，本研究では高空間解像度で高フレームレートな動画像を取得する手法として圧縮センシングを用いる．圧縮センシングを用いた動画像撮像は，画素ごとに露光タイミングをずらした画像を撮影することで，より密な時間の情報をスパースに単一画像に畳み込む．この符号化露光画像から再構成処理により，撮像センサの時空間サンプリングを超えた画像を再構成する．

しかしながら，このような画素ごとに露光を制御できる市販センサは存在しない．そこで，本研究では準画素毎露光制御を実現する CMOS イメージセンサを用い，符号化露光の自由度を向上させる露光制御信号と読み出しによる歪みを抑えた圧縮センシングによる動画像の高速撮像手法を提案した．また，露光タイミングをずらすことによる符号化露光は時空間情報の畳み込み，撮影画像からの動画の再構成は逆畳み込みと表現できるため，符号化露光と再構成をどちらもニューラルネットワークで表現することで同時最適化することができる．しかし，提案するセンサに限らず，符号化露光可能なイメージセンサには露光制御に制約があることが多い．そこで，本研究では制約のある符号化露光と動画像再構成の同時最適化手法を提案する．

目 次

図目次

表 目 次

第 1 章

序論

1.1　研究背景

光線には空間や時間，色など様々な情報が含まれるため，カメラを用いることで実世界の様々な情報を取得することができる．しかし，通常のイメージセンサは画素が 2 次元に配置されており，すべての情報を一度に取得することは困難である．例えば，高空間解像度で高フレームレートな動画像は実世界の現象を分析する際に有用であるが，空間解像度とフレームレートの間にはトレードオフの関係があり空間情報と時間情報を同時に取得することは困難である (図 1.1)．これは，カメラの読み出し回路の帯域の制限により，単位時間あたりに読み出せる情報量に限りがあるためである．そのため，空間解像度の高いスチルカメラではフレームレートの高い撮影は行えず，時間解像度の高いビデオカメラでは空間解像度を高くすることは難しい．このトレードオフをハードウェアを開発することで解決しているものがハイスピードカメラである [1]．ハイスピードカメラはセンサからの読出しを高速に行うため画素ごとにバッファを設けるほか，AD 変換の時間を短縮するために並列の AD 変換器を搭載している．しかし，このような特殊なセンサは非常に高価であり，回路が複雑になることからフォトトランジスタの面積が減少するため感度が悪

くなる問題もある．そこで，撮影後に時間情報を復元する手法として，多くの画像を統計分析して通常の撮影画像からフレーム間を補完して高フレームレートな動画を生成する手法 [2–6] が提案されている．これらの手法は，撮影された動画から大量の学習データで訓練されたニューラルネットワークを用いて，オプティカルフローや位相，奥行やモーションブラーを推定し，撮影された動画のフレーム間の物体の動きを補完することで高フレームレートの動画を復元する．しかし，近年ではセンサの性能は物理的な限界に近付きつつあり，また，画像処理においても大量の画像の用意が容易になったためこれ以上統計処理により取り出せる情報量を劇的に増やすことは難しい．これはハードウェアはハードウェアのみで，ソフトウェアはソフトウェアのみで研究されていることが一因である (図 1.2-(a))．しかし，撮影後に画像処理を施すことを前提とするならば，後処理に必要な情報を積極的に取得すると効率や精度の向上が見込める．このように撮影から後処理までを一貫して設計する手法をコンピュテーショナルフォトグラフィといい，近年注目を集めている (図 1.2-(b))．光によって取得できる実世界の情報には様々あるが，本論文では 2 次元の画像に時空間情報を畳み込み，撮影後に動画像として再構成する手法について論ずる．

従来のデバイス開発ではいかに画質良く画像を撮影できるかを目指し，従来の画像処理では入力は人間が目で見ているような画像であると盲目的に受け入れてきた．しかし，この固定観念を捨て，撮影画像に画像処理することを前提に撮影と画像処理を共同設計することで劇的に性能を向上できる．例えば，通常のカメラで撮影された画像には時間や距離などの情報は限定的にしか含まれておらず，撮影後の後処理のみによる復元には限界がある．そこで，通常のイメージセンサを複数組み合わせて時空間解像度のトレードオフを克服しようとする手法として，多数の低フレームレートのカメラを用いてそれぞれの露光タイミングを少しずつずらして撮影し，後処理により高フレームレートの動画を得る手法 [7] や，高空間解像度のカメラと高フレームレートのカメラで同時にシーンを撮影し，高空間解像度の画像を手掛かりに低空間解像の高フレームレートの画像を高空間解像化する手法 [8] が提案されている．しかし，これらの複数のカメラを用いる手法は撮影装

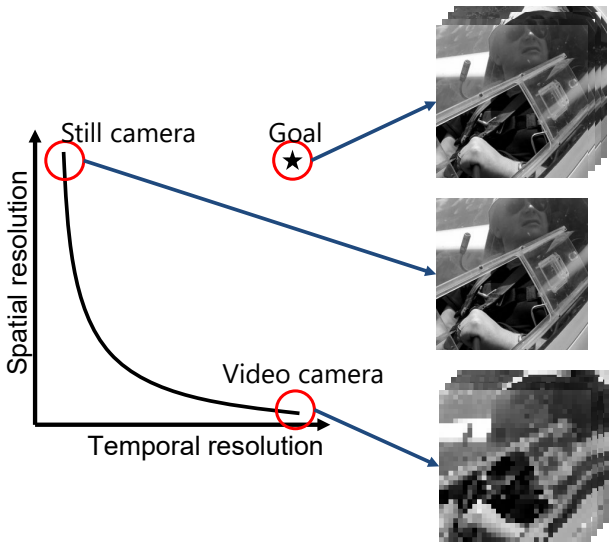


図 1.1: **動画像撮影における時空間解像度のトレードオフ**．スチルカメラでは空間解像度が高いが時間解像度が低い．一方，ビデオカメラでは時間解像度が高いが空間解像度が低い．本論文では時空間解像度の高い動画の取得を目指す．

置が大型化するほか，撮影データの冗長性が高い．また，複数のカメラを横に並べて撮影する場合は視点が異なるためそれぞれのカメラの画像が少し異なるためそのずれの修正が必要になり，ビームスプリッタ等を用いて同軸に複数のカメラを配置する場合は光量が落ちるためノイズが増えるなどの問題がある．そこで，フレームレートの高くないカメラを反復可能なシーンに同期させ，露光タイミングを少しずつずらしながら複数回撮影することで，センサのフレームレートを大きく超える動画を再構成する手法が提案されている [9]．しかし，この手法はカメラと同期できる反復可能なシーンにしか適用できず，繰り返し撮影するため撮影時間が大幅にかかるといった問題がある．また，このような複数の撮影データから再構成する手法は，冗長性が高く計算や記録のコストも高くなる．しかし，そもそも撮影後に画像処理を行う前提なら撮影画像は必ずしも人間が理解できる必要はなく，復元に必要な情報を含んでいればよい．また，画像などのデータはサイズが大きいため保存の際には圧縮されるが，これは取得後にデータの大部分を捨てているという

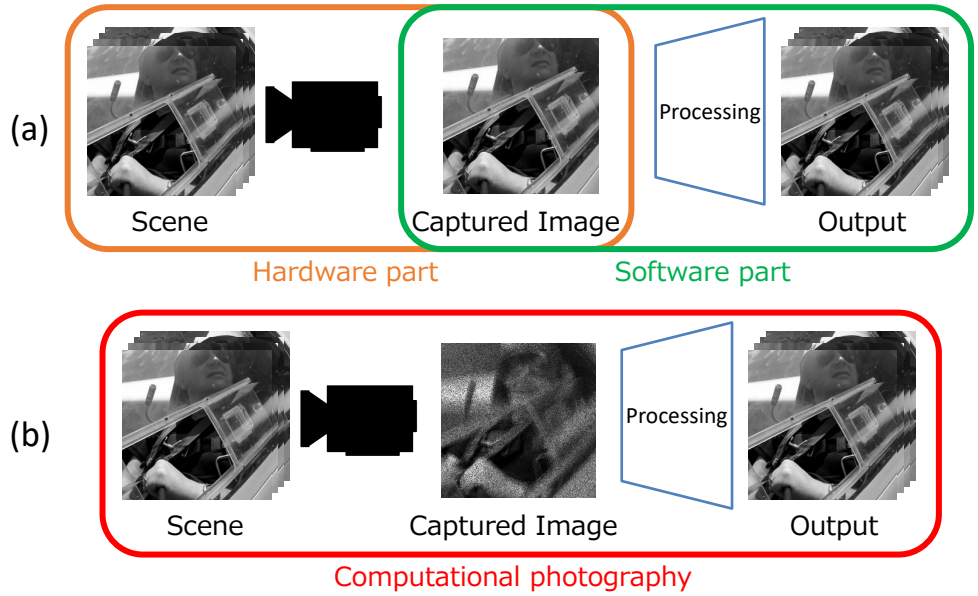


図 1.2: **一般的なコンピュータビジョンのアプローチとコンピュテーショナルフォトグラフィのアプローチ**．(a): 一般的なコンピュータビジョンのアプローチでは撮影と後処理を個別に設計する．(b): コンピュテーショナルフォトグラフィでは撮影と後処理を共同で設計することにより，精度を向上させることができる．

ことであり，非効率である．(動画の非可逆圧縮形式である H.264 の圧縮率はおよそ 1/100 である．) そこで，センサ上で必要な情報を圧縮して取得し，撮影後に再構成することで効率のよい撮影が行える．この考え方を圧縮センシングと呼び，本研究ではこの圧縮センシングにより 2 次元のイメージセンサで取得できる情報の拡張に取り組む．

画像などのデータはサイズが大きく多くの冗長性を含んでおり，圧縮センシングではこの冗長性を利用し必要な情報を選択して観測する．圧縮センシングでは，観測対象の情報が疎であると仮定し，観測行列により間引かれた，標本化定理よりも少ない観測から元の情報を復元する．元信号を疎な表現へ変換することが容易である，X 線 CT [10, 11] や MRI [12, 13]，ブラックホールの撮影 [14–19] などは圧縮センシングが用いられる典型的

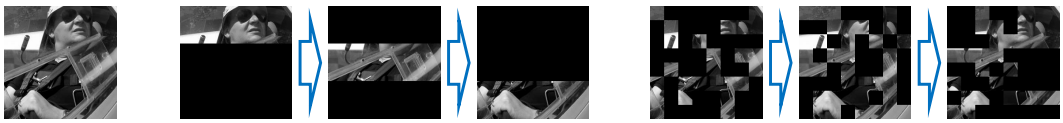
な例である．ここで観測行列を $\boldsymbol{\phi}$ とすると，観測 $\boldsymbol{y}$ は元信号 $\boldsymbol{x}$ より以下の式で表される．

$$\boldsymbol{y} = \boldsymbol{\phi}\boldsymbol{x}, \tag{1.1}$$

ここで $\boldsymbol{y} \in \mathbb{R}^M$，$\boldsymbol{x} \in \mathbb{R}^N$ とすると，$M \geq N$ の場合は $\boldsymbol{y}$ から $\boldsymbol{x}$ 一意に求めることができる．しかし，圧縮センシングでは $\boldsymbol{\phi}$ によって観測を間引き $M < N$ となるため，劣決定問題となる．そこで，圧縮センシングでは $\boldsymbol{x}$ がスパースである，つまり $\boldsymbol{x}$ の N 個の成分のうち高々K 個 $(K < M)$ しか値を持たないと仮定することにより，$\boldsymbol{x}$ を再構成する．伝統的にはスパース最適化 [20–24] や混合ガウスモデル [25] などを用いて再構成されてきたが，近年では深層学習を用いた手法 [26–30] が主流となってきている．

本研究では従来のカメラシステムの一部をより自由に制御可能なものへ変更することで，圧縮センシングにより実用性と汎用性を持たせながら，高空間解像度で高フレームレートな動画像を取得する手法を提案する．通常の動画撮影は，すべての画素が同時に露光するグローバルシャッタのセンサ (図 1.3-(a)) や，センサの上部から行ごとに順次露光するローリングシャッタのセンサ (図 1.3-(b)) を用いて静止画を連続撮影することで実現されるのに対して，圧縮センシングを用いた動画像撮像は画素ごとに露光タイミングをずらした符号化露光により画像を撮影することで，より密な時空間の情報をスパースに単一画像に畳み込む (図 1.3-(c))．この符号化露光画像から露光の符号化パターン (露光パターン) の逆畳み込みにより，イメージセンサの時空間サンプリング性能を超えた画像を再構成する．しかしながら，このような画素ごとに露光を制御できる市販センサは存在しない．既存研究ではシミュレーション実験にとどまるか [23,26,31]，反射光学系を用いた疑似実装を用いることが多い [20,28,29,32]．本研究では High Dynamic Range (HDR) 撮影のために画素ごとに露光時間を制御できるよう設計されたセンサを利用し，センサのみで符号化撮像を実現する．

圧縮センシングにおいて，信号処理の観点からシーンについての事前知識の無い場合の最適な露光の符号化は一様分布からのランダムな観測である．これは，周期的な観測では



(a) Global shutter　(b) Rolling shutter　(c) Coded shutter

図 1.3: **シャッタ方式の違い**．(a): グローバルシャッタは全画素同時に露光する．(b): ローリングシャッタはセンサ上部から順に露光する．(c): 符号化シャッタは画素ごとに異なるタイミングで露光する．

エイリアシングなどによって特定の周波数の情報が観測できないが，ランダムに観測することで様々な周波数成分に対応することができるからである．しかし，事前にシーンについての知識がある場合，ランダムよりも適した符号化がある．例えば，一般的な動画においてシーンのテクスチャは大きく変わらない，急激に変化しない，などの傾向がある．そのため，事前にその知識を使って符号化の最適化を行うことができる．また，監視カメラなどではシーンがほぼ変わらないため，そのシーンに特化した符号化を行うことでさらに効率の良い観測が行える．しかし，圧縮する際には元の情報を復元する際に必要となる情報を残しながら圧縮する必要があるが，復元する際に必要となる情報は復元しないとわからない．よって，圧縮と再構成を同時に最適化することで効率の良い圧縮と高品質な再構成を達成できる (図 1.4)．ここで，圧縮は時空間の畳み込み，再構成は逆畳み込みと表現することができる．このデータを特徴量に畳み込み，逆畳み込みにより元のデータに復元するというタスクはオートエンコーダ [33] と同様の枠組みである．よって，本研究ではニューラルネットワークを用いて符号化露光と撮影画像からの動画再構成を同時に最適化する．

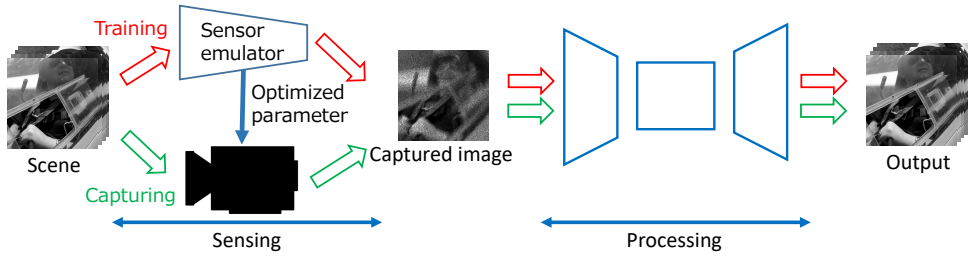


図 1.4: **撮影と再構成の同時最適化**．訓練時にはセンサで符号化を模倣するネットワークと再構成を同時最適化し，撮影時には最適化した符号化をカメラに実装し再構成ネットワークにて元のシーンを再構成する．

1.2　研究目的

本研究では圧縮センシングによる動画像撮像について，実際のセンサによる撮影と再構成の同時最適化の手法を提案する．イメージセンサの電子シャッタを利用し，画素ごとに露光タイミングをずらすことで時間情報を単一の画像に畳み込み，再構成により時間情報を取り出す．

符号化露光センサを用いた動画像の高速撮像

圧縮センシングによる動画像撮像を実現するため，時空間情報を符号化して撮影する必要がある．そのため，画素ごとに露光タイミングを制御可能なイメージセンサを用いて圧縮センシングによる動画像の高速撮像手法を提案する．従来研究の多くがシミュレーション実験に終始するか，反射光学系などを用いた疑似実装により実験を行っているが，提案手法ではイメージセンサのみで符号化露光を行う．本研究で用いるセンサは HDR 撮影のために，グローバルシャッタセンサの信号線に画素ごとに露光制御するための回路を追加したセンサである．元々HDR 撮影のために設計されているため完全に画素ごとに露光を制御することができないが，露光のリセット信号と電荷の転送信号を組み合わせることで疑似的に画素ごとに露光を制御することができる．提案手法ではこの疑似的な画素ごとの露光制御により圧縮センシン

グによる動画像撮像が可能であることを示す.

露光の符号化と動画像再構成の同時最適化

圧縮センシングでは観測行列を用いて観測を間引くことで効率の良い観測を実現するが，観測行列を計測対象に最適化することで再構成品質を向上させることができる．しかし，実際に計測する際に使用する装置の制約によっては理想的な計測を行えない場合がある．本研究で用いるイメージセンサは，露光を縦または横の画素間で共有された信号線で制御するため，露光する画素が 1 列に並ぶという制約がある．また，最適な観測行列は再構成のアルゴリズムにも依存する．よって，撮影に用いるイメージセンサの制約を考慮しながら符号化と再構成の同時最適化により再構成品質を向上させる.

本論文の次章以降の構成は次のとおりである．まず，第 2 章で圧縮センシングによる実世界計測の関連研究を述べる．第 3 章で本研究に用いるイメージセンサの露光制御の仕組みとそのセンサによる符号化露光の実現について述べる．第 4 章では制約のある露光の符号化と動画像再構成の同時最適化手法について述べる．最後に，第 5 章で本論文の結論と今後の展開について述べる.

第 2章

圧縮センシングによる実世界計測

本章では圧縮センシングための撮影装置，圧縮センシングによる動画像の再構成，圧縮センシングにおける観測と再構成の同時最適化の先行研究について述べる．

2.1　圧縮センシングのための撮影装置

イメージセンサで動画を撮影する際には空間解像度とフレームレートの間にトレードオフの関係であり，近年ではイメージセンサにより直接観測できる情報は限界に近付きつつある．そのため，符号化により観測できる情報を拡張する圧縮センシングのためのイメージセンサが開発されている [34]．Robucci ら [35] は同じ行の画素に対し同じ信号によって制御することにより，センサ上で行ごとに露光制御を行うイメージセンサを設計した．このイメージセンサは一般的なイメージセンサである Complementary Metal-Oxide-Semiconductor (CMOS) センサのようなアクティブ画素ではなく，パッシブ画素を採用しているため，ノイズに対して頑健でないという特徴がある．Dadkhah ら [36] の設計したイメージセンサには画素の外から露光を制御する信号線が追加されている．このセンサの露光制御は画素ブロックごとに行われ，ブロック内の画素は個々に制御することができる

が，すべてのブロックで露光の符号化は同一となる．これにより実現できる露光の符号化は制限されるが，センサの画素数に対して信号線の数が少ないため，符号化露光しない一般的なセンサと比較して画素面積に対する受光面積 (フィルファクタ) の減少が最小限で抑えられる．Majidzadeh ら [37] や Wei [38] らは画素内にランダムパターン生成器を搭載した CMOS センサを提案した．画素内に低ビットメモリを含む素子を追加することにより，画素ごとにランダムなパターンでの露光が可能となる．しかし，画素内に必要な素子が増えるため，フィルファクタは著しく低くなる．この欠点はセンサの設計が複雑になればなるほど顕著となる．このように，露光の自由度とセンサの感度はトレードオフの関係にある．Oike ら [39] の設計したイメージセンサは，一般的なセンサと同様に全画素を一様に露光し，信号は順次読み出される．読み出された信号はランダムに選択され，足し合わされたのち量子化される．この手法は，主に空間的な超解像を実現するために採用される [35–37,39]．そのため，計測される信号の空間解像度を下げることで，一定の帯域幅の中でフレームレートを上げることができる．Spinoulas ら [40] はオンチップ圧縮センシングを提案した．Spinoulas らは読み出し設定の自由度が高く安価な商用の開発ツールキットを利用することにより，読み出しの反転，関心領域位置の選択，画素の間引きの組み合わせにより不均一な観測を実現した．このセンサは我々のセンサと同様にデジタル回路で符号化を行うが、符号化方式が大きく異なる．このセンサでは読み出しの際に符号化が行われるが，我々のセンサでは画素において符号化する．

Hitomi ら [20] は圧縮動画センシングに用いるセンサとして，一般的な CMOS センサと同様に 3 トランジスタのセンサを想定した．このセンサでは画素ごとに露光の制御が可能であり，センサのみで符号化が可能である．しかし，このセンサは画素内にバッファを持たないため，1 フレームの間に複数回露光すると情報が壊れてしまうため，露光を分割することはできない．また，イメージセンサのダイナミックレンジは限られるため，各画素の露光時間は同じであることが望ましい．よって，このセンサでは各画素の露光時間はすべて同じにそろえて撮影される．また，Hitomi らはこのようなセンサを想定して実験して

いるが，実際のセンサは試作しておらず，反射光学系を用いた疑似実装にとどまっている．このようなセンサを本論文では Single-Bump Exposure (SBE) センサと呼ぶ (図 4.2-(b)).

Antipa [41] らはローリングシャッタを利用し，1 フレームに時間情報を畳み込み動画の再構成を行った．多重化光学系を用いてシーンを空間的に圧縮し，1 列の画素でシーン全体の情報を観測する．CMOS センサのローリングシャッタにより 1 列ずつ読み出すことで高速にシーンを撮影することができる．この手法は一般的な CMOS センサと拡散板を用いるため容易に実装可能である．

2.2　圧縮センシングによる動画像の再構成

画像処理において深層学習が主流となる以前は圧縮センシングにおける元の信号の再構成に，L0 ノルム正則化 [20, 21] や L1 ノルム正則化 [22–24]，混合ガウス分布を用いた手法 [25] や Total Variation (TV) 正則化による手法 [42] などが使われてきた．しかし，これらモデルベースの手法は繰り返し計算が必要であり計算時間が長いという欠点がある．そこで，画像処理における深層学習の発展とともに，圧縮センシングの再構成においても深層学習が取り入れられるようになった．Iliadis ら [26] は全結合層によって構成されたネットワークにより，観測から元の動画への写像を学習することにより再構成を行った．しかし，全結合層のみで構成されたネットワークはパラメータ数が多く，また，ネットワークの入力の符号化パターンの位置が固定されるため，8×8 などの小さなパッチ単位でしか再構成できず，ブロック状のアーティファクトが現れるなどの問題がある．そのため，近年では畳み込みニューラルネットワークを用いた手法が多く提案されている．Ma ら [27] は交互方向乗数法 (Alternating Direction Method of Multipliers: ADMM) における正則化項に CNN を用いることで再構成を行う手法を提案した．ADMM に CNN を適用することにより，TV などの伝統的な手法と比べ大幅な品質向上を達成した．Cheng ら [29] は 3 次元畳み込みニューラルネットワークをグループ化し可逆とした RevSCI-net を提案した．こ

れにより，メモリの使用量を抑え訓練時間の短縮を達成した．Sun ら [30] は残差ネットワークを用いることにより，再構成時間と再構成品質のバランスを取っている．このネットワークはビデオフレーム間の時空間的な相関を利用することで，再構成品質を向上させることを目的としている．

2.3　観測と再構成の同時最適化

深層学習による圧縮センシングの再構成が一般的になる中でエンコーダデコーダネットワークのように，観測と再構成をニューラルネットワークで実装し，End-to-End 学習により観測パラメータと再構成のためのデコーダを同時に最適化する研究が増えてきている．このような手法は Deep sensing や Deep optics と呼ばれ，近年さまざまな分野に適用されている．Inagaki ら [43] は符号化開口を用いてライトフィールドの圧縮センシングを提案した．カメラによる観測と再構成をニューラルネットワークで表現し，2 枚の符号化開口パターンとデコーダを End-to-End 学習により同時最適化した．これにより，$5 \times 5 = 25$ 視点の画像を 2 枚の画像として観測し，デコーダにより再構成することを可能とした．Wu ら [44] はカメラの絞り位置に位相マスクを置くことで単眼，単一の画像から受動的な 3 次元計測を提案した．位相マスクと再構成を同時最適化することにより，従来手法よりも優れた位相マスクを設計することを可能とした．Nie ら [45] はフィルタの分光透過率をエンコーダネットワークのパラメータで表現し，観測と再構成を同時最適化した．これにより，少数の観測画像から精度よくハイパースペクトル画像を推定した．同様の取り組みは圧縮センシングによる動画像の撮影へ適用され [46,47]，ランダムな観測を用いた従来手法と比べ高品質な再構成を達成している．

第 3章

符号化露光センサを用いた圧縮センシングによる動画像の高速撮像

イメージセンサで動画を撮影する際，アナログ・デジタル (AD) 変換器と読み出し回路の帯域幅がボトルネックとなるため，空間解像度とフレームレートの間にトレードオフの関係がある．よって，AD 変換器の高速化や帯域幅の拡張といったイメージセンサの改善が行われているが，近年ではイメージセンサにより直接観測できる情報は限界に近付きつつある．ここで，動画は撮影後保存や転送する際には圧縮されることが一般的であり，労力をかけて取得した情報の多くを捨てている．これは動画が多くの冗長性を持っているからであり，この冗長性を利用して撮影時に圧縮して情報を取得する圧縮センシングによる動画像の撮影が行われてきた．圧縮センシングを利用した動画撮像の手法は，露光中に全画素同じタイミングでシャッタを開け閉めすることで時間情報を畳み込むフラッターシャッター [48,49] や周期的運動に対するフラッターシャッター [50]，読み出し回路を工夫することによりローリングシャッタの機構を利用して各行の読み出しタイミングや露光時間を変えることによる時間情報の符号化を実現した手法 [51]，画素ごとに露光タイミングをオフセットすることでフレームレート以上の時間情報を取得する時間ピクセル多重化手

法 [52]，シーンの動きが多い部分はフレームレートを上げ動きの少ない部分は空間解像度を上げるハイブリッドシャッタ [53]，画素ごとにシャッタタイミングや露光時間を変えることにより空間解像度を保ったままより多くの時間情報を畳み込む手法 [20,25,26,54] などがある．圧縮センシングにおいて理想的な観測であるランダム露光を実現するため，全画素の露光を自由に制御できることが理想である．そのため，近年では画素ごとにシャッタを制御する手法が主流であるが，センサのみで画素ごとに自由に露光を制御することができるものは市販されていない．よって多くの手法がシミュレーション実験 [23,26,31] または反射光学系などを用い光学的にシミュレートされた実装 [20,28,29,32] を用いている．

そこで，本研究では準画素毎露光制御を実現する CMOS イメージセンサ [55] に対し，露光の自由度を向上させる露光制御信号を適用し，読み出しによる歪みを除去した圧縮センシングによる動画像の高速撮像手法を提案する．本研究で用いる CMOS イメージセンサは図 3.2 のように，各画素の露光を 2 種類の露光制御信号線により制御する．この信号線は露光のリセットと電荷の転送を制御しており，これらの信号のタイミングを制御することによってサブフレームの時間情報を畳み込んで撮影する．これらの露光制御信号線はグローバルシャッタセンサにも搭載されているものであり，露光のリセット信号線は縦列の画素間で，電荷の転送信号線は横列の画素間で共有されている．グローバルシャッタセンサではすべての信号線に同時に信号を入力することで全画素一斉露光を実現しているが，本センサではそれぞれ別の信号を入力することで符号化露光を実現する．図 3.2 のようにそれぞれの信号線を 8 つに分割することで 8 列ごとに異なる信号を入力することが可能である．全画素自由に露光を制御することができるセンサと比べ画素数分の信号線が必要なく，フォトダイオードの受光面積を犠牲にしないため，感度は通常の CMOS センサーと同等である．本章では限られた信号線による準画素毎露光制御を可能としたイメージセンサを用い，この限られた信号線によって符号化露光を実現するための信号パターンを提案する．学習型の再構成デコーダを用いて，CMOS センサで撮影した画像から読み出しによる歪みを補正した高フレームレートの動画を再構成する．本論文ではこれ以降，本

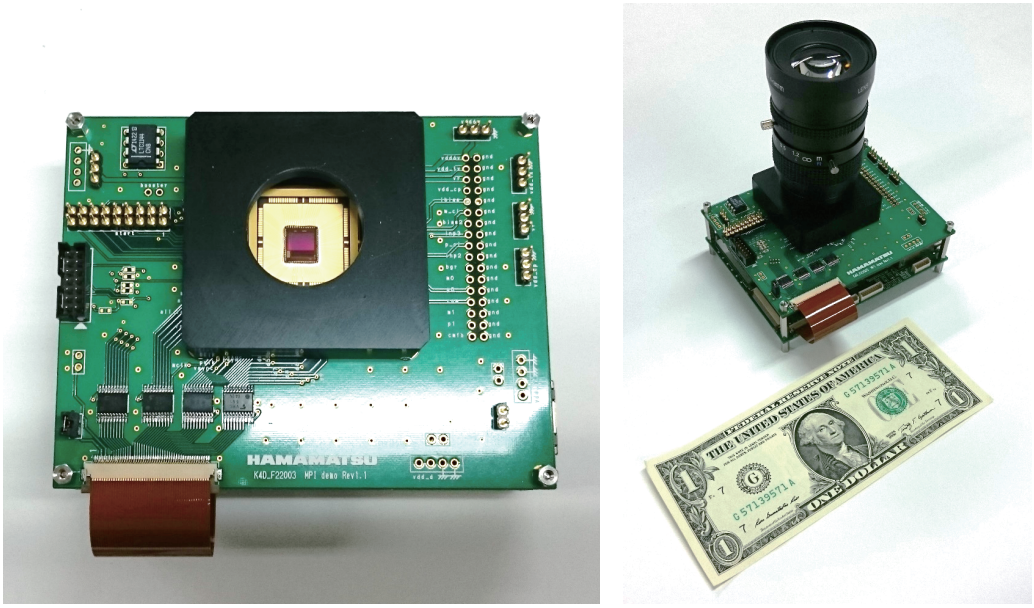


図 3.1: **準画素毎露光制御可能な試作イメージセンサ.**

研究で用いる準画素毎露光制御可能なイメージセンサを Quasi Pixel-Wise Exposure (QPE) センサと呼ぶ.

3.1　準画素毎露光制御可能なイメージセンサ

本研究で用いる QPE センサは既存のグローバルシャッタセンサ (Hamamatsu Photonics, S13102) を元に設計されている．図 3.2 に示すように，QPE センサの露光制御信号線は元となったグローバルシャッタセンサと同じリセット信号線と転送信号線であり，このうちリセット信号線を水平方向から垂直方向へ変更したのみである．このため，信号線数やトランジスタ数はグローバルシャッタセンサと同じであり，受光面積や画素サイズはグローバルシャッタセンサとほぼ同じである．よって，グローバルシャッタから符号化露光を可能としたことによる感度や解像度に対する影響は限定的である．元となったグローバルシャッタセンサはマイクロレンズが装着されていないため，画素面積に対する受光面積

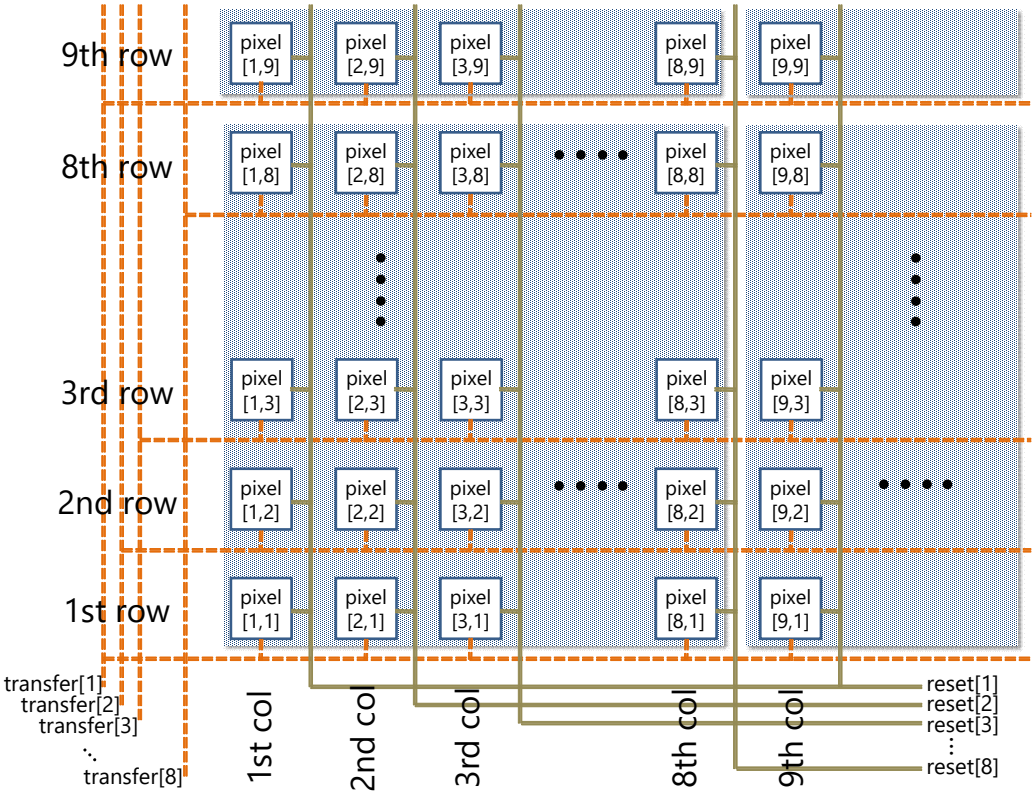


図 3.2: **準画素毎露光制御可能な CMOS イメージセンサの構造**．これはセンサの左下を拡大表示したものである．

(フィルファクタ) が 39%，センササイズが $7.4\mu m \times 7.4\mu m$ である．一方，QPE センサはフィルファクタが 36%，センササイズが $7.4\mu m \times 7.4\mu m$ である．水平に並んだ画素は転送 (transfer) 信号線を共有し，垂直に並んだ画素はリセット (reset) 信号線を共有しているため，これらの画素には同じ信号が入力される．また，図 3.2 に示すように，8 本のリセット信号線と転送信号線があり，それぞれ 8 列ごとに接続されている．そのため，露光は 8×8 画素単位で制御され，8×8 画素のブロック内の画素を 8 本のリセット信号線と転送信号線で制御する．

ここから，QPE センサの 1 画素の動作について説明する．図 3.3 に画素から AD コンバータまでの等価回路を示す．画素にはリセット/転送スイッチがあり，リセット/転送信

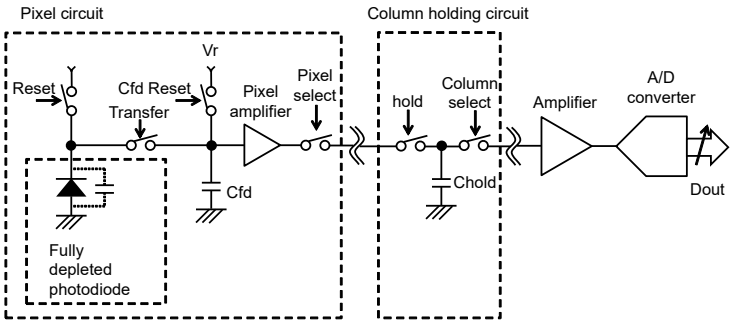


図 3.3: **画素から AD コンバータまでの等価回路**．画素の回路構成は一般的な CMOS センサと同様であるが，リセット/転送スイッチとキャパシタ C_{fd} が画素内に追加されている．制御信号は画素の外から入力される．

号線と接続されている．各フォトダイオードは光電効果により電荷を発生させ，この電荷をフォトダイオード内に保持する．また，フォトダイオードに光が当たっている間は常に電荷を発生させ続ける．そのため，露光を開始するためにはあらかじめフォトダイオードに蓄積された電荷を捨てる必要があるため，露光リセット信号を ON にしてフォトダイオード内の電荷を捨て，露光リセット信号を OFF にすることでフォトダイオード内に電荷が貯まる，つまり露光が開始される．フォトダイオード内の電荷は露光転送（露光固定）信号が ON になるとキャパシタ C_{fd} に転送され，転送信号が OFF になると露光処理が終了する．この露光リセット信号と転送信号のタイミングを様々組み合わせることにより，多くの露光パターンを設定することができる．1 フレームの露光が終了すると，各画素のキャパシタ C_{fd} 内の電荷が順次読み出される．読み出し直後，拡散容量リセット信号によりキャパシタ C_{fd} をリセットし，次のフレームの露光準備が完了する．

本研究で用いるセンサの入力信号の系列とそれに対応する露光パターンの例を図 3.4 に示す．本研究で用いる QPE センサーが様々なアプリケーションに柔軟に対応できることを示すために，6 つの異なる種類の露光パターンを示す．図 3.4-(a)，(b)，(c) はそれぞれグローバルシャッタ，フラッタシャッタ [48,49]，エピポーラ撮像 [56] の例である．図 3.4-(a)

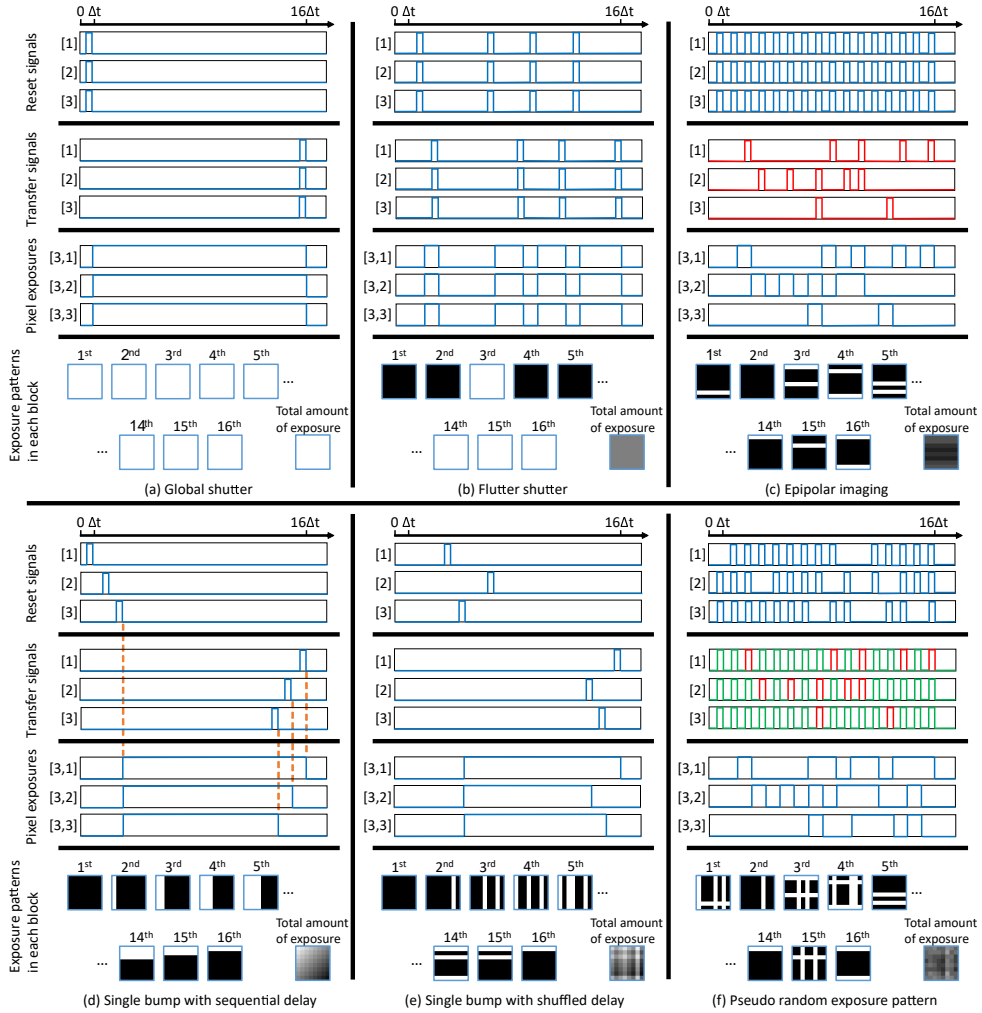


図 3.4: **入力信号列とそれに対応する画素の露光量、およびブロック内の露光パターンの例**．(a): グローバルシャッタの例．(b): フラッタシャッタ [48,49] の例．(c): エピポーライメージング用 [56] のシャッタパターンの例．(d)，(e)，(f): 高速撮像用の符号化パターンの例．1 行目と 2 行目にそれぞれ 3 つのリセット信号と転送信号を示し，3 行目に 3 つの画素のシャッタタイミングを示す．4 行目は露光画素を白，非露光画素を黒とした露光パターンの 8×8 画素ブロックを示す．右端のパターンは露光パターンの全露光時間に対して正規化した積算露光量を示す．各信号のグラフの横軸は時間軸を表し，ここでは露光時間を $16\Delta t$ とする．制御信号の縦軸は信号の ON/OFF を表す．(c) と (f) の転送信号は $-\delta$ で転送する位置を赤，$+\delta$ で転送する位置を緑で表現している (詳細は第 3.2 章に記述している)．

と (b) では，すべての信号線に同じ信号を入力し，全画素同じタイミングで露光している．一般的なグローバルシャッタセンサでも図 3.4-(a) と (b) は実装可能であるが，本研究で用いる QPE センサは図 3.4-(a)，(b)，(c) のすべての露光パターンを実装可能である．本研究で用いる QPE センサーでは，プロジェクタと同期することでエピポーラ幾何により直接反射光のみを観測可能なエピポーラ撮像 [56] が可能である．しかし，QPE センサーでエピポーラ撮像を行うためには，プロジェクタとカメラのエピポーラ線を水平または垂直に限定する必要がある．また，提案する高速度撮像のための露光パターンを図 3.4-(d)，(e)，(f) に示す．これらの例では，1 フレームの中でリセット信号と転送信号の組を単位 (1 回のリセット信号と 1 回の転送信号が対応している) として示している．この露光の単位は 1 フレームの中で何回も繰り返すことができるが，一部の従来手法 [20] のセンサでは 1 回の露光しかできない．露光パターンは垂直方向のリセット信号と水平方向の転送信号の組み合わせで決定され，隣接画素で異なるタイミングでの露光を実現できる．

3.2　準画素毎露光制御による符号化露光

第 3.1 章のような構造のセンサを用いることで，隣接する画素に異なるタイミング露光を設定し不均一な露光を実現できる．QPE センサで実現できる 1 フレームで 1 回露光のパターンは図 3.4-(d) のように水平方向，垂直方向，時間方向の各次元で一定の依存性がある．また，各信号の順序をランダムに入れ替えても露光パターンの水平方向，垂直方向，時間方向の依存性は図 3.4-(e) のようになる．本章では疑似ランダムサンプリング を実現するためのセンサ制御法を提案する．基本的な考え方は以下の通り．1) 露光パターンは 1 フレーム中の露光の総和をサブフレームで分割したものとして表現される (図 3.5)．2) 各露光はリセット信号と転送信号を用いて制御される．各露光の ON/OFF 状態は，リセット信号の ON/OFF 状態と転送タイミングの $\pm\delta$ 遅延の組み合わせで制御される．提案する露光制御処理の詳細な説明は以下の通りである．

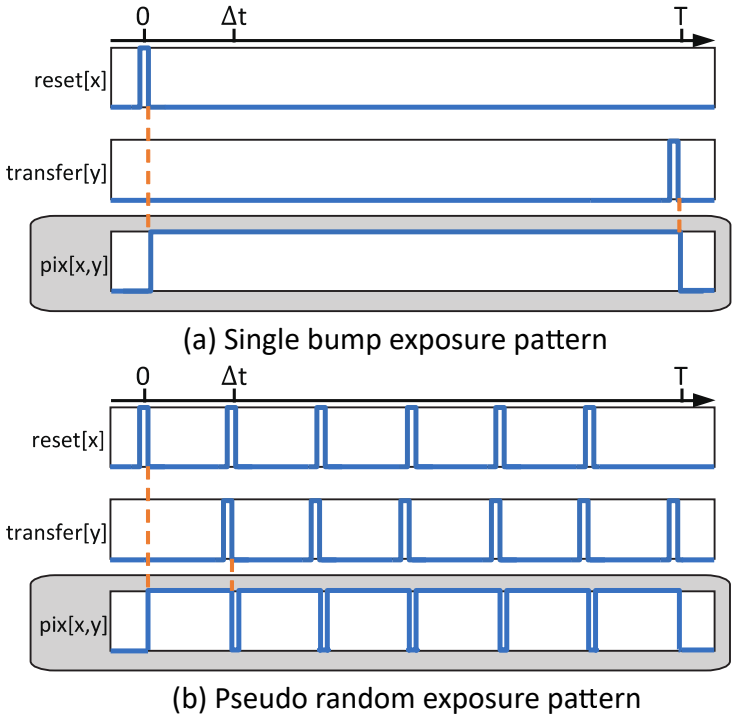


図 3.5: **通常露光と準画素毎露光制御に必要な信号の違い**．1 フレームを撮影するための制御信号と露光タイミングを示す．(b) の例では，露光が 6 つに分割されているが，ここでは (a) の露光と同等であるとする．

撮影したフレームの時間長 T を N 個のサブフレームに分割すると考える．これは，動画再構成後のフレーム数に相当する．一つのサブフレームの露光時間を区間 $\Delta t = \frac{T}{N}$ で表す．各サブフレーム n には $1, 2, 3, ..., N$ の番号を振る．各画素の露光制御はリセット信号の ON/OFF と転送タイミングをリセット信号から δ だけずらして行う．ここで δ は極めて小さく，サブフレーム長に対して無視できる．サブフレーム n において，$n\Delta t - \delta$ で C_{fd} に電荷を転送すると，$n\Delta t$ でリセット信号が ON か OFF かにかかわらず，その画素は露光される (図 3.6-(a),(b))．$n\Delta t$ でリセット信号が OFF であれば，転送タイミングに関係なく画素が露光される (図 3.6-(b),(d))．しかし $n\Delta t$ でリセット信号が ON の場合，$n\Delta t + \delta$ の時点では直前にフォトダイオードがリセットされているため C_{fd} に電荷が転送されず，

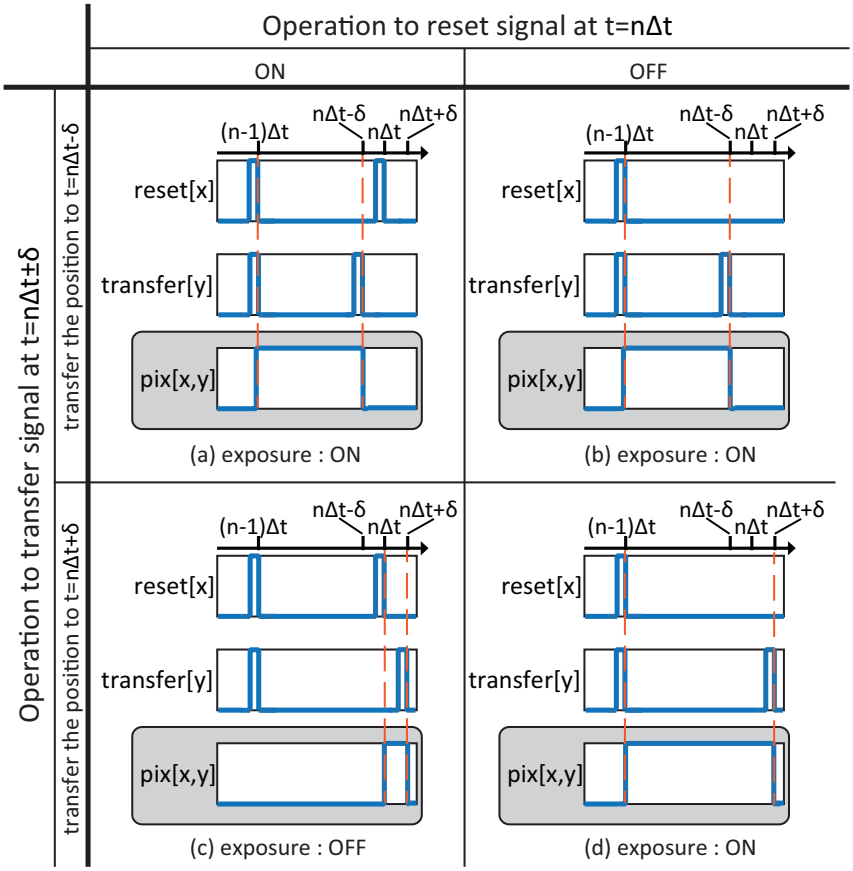


図 3.6: **リセット信号と転送信号の組み合わせ**．サブフレーム n における制御信号と対応する露光状態を示す．左右の差は $t = n\Delta t$ におけるリセット信号である (ON,OFF)．上下の違いは転送信号のタイミングである (つまり，転送信号のタイミングが $n\Delta t - \delta$ または $n\Delta t + \delta$ である).

結果その画素は露光されない (図 3.6-(c)). したがって，各サブフレーム n ごとに露光するかしないかが決まるため，サブフレーム n の露光を独立に制御できる．これにより，時間方向の画素の依存性を除去することができる．さらに，水平方向/垂直方向の空間依存性も上記の操作により緩和することができる．図 3.6 に示すように，あるリセット/転送信号に対し 2 つの応答，つまり同じ水平方向/垂直方向に並ぶ画素に対して異なる露光状態を設定することができる．

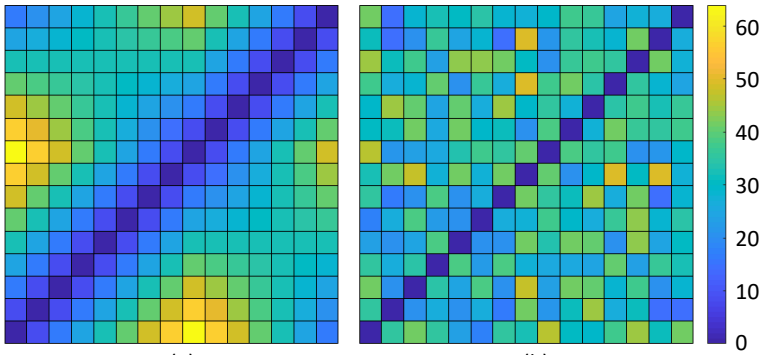


図 3.7: **サブフレーム間の露光パターンのハミング距離**．図の (m, n) の位置は m 番目のサブフレームと n 番目のサブフレームの露光パターンのハミング距離を表し，色が黄色に近いほど距離が遠く青に近いほど距離が遠いことを表す．(a): 1 フレームの間に 1 回露光する露光パターンにおけるハミング距離．(b): 提案する準画素毎符号化露光パターンにおけるハミング距離．

上記の通り準画素ごとに露光を制御制御することができる．このような操作による提案センサの露光パターン例を図 3.4-(f) に示す．図 3.4-(d)，(e) のような 1 フレーム内にリセット信号と転送信号が 1 組しかないセンサの露光制御と比較して，図 3.4-(f) の露光パターンは時間方向の自由度が高く，列依存性が緩和されていることがわかる．また，各露光パターンのフレーム間におけるハミング距離の推移を図 3.7 に示す．1 フレームに一回のみ露光するパターンではハミング距離が線形に変化するのに対し (図 3.7-(a))，提案する準画素毎露光制御では非線形でありよりランダム性が高いことが確認できる (図 3.7-(b))．このように，提案する露光制御法による疑似ランダムパターンは他の露光パターンと比べランダムであり動画像の符号化露光に使用することができる．また，本研究で用いる QPE センサでは図 3.4 に示すように，通常のグローバルシャッタセンサとしても使用することができる．露光制御信号と露光パターンの関係の一例を図 3.8 に示す．露光パターンは 8×8 ブロックの繰り返しで構成され，露光される画素は水平方向もしくは垂直方向に

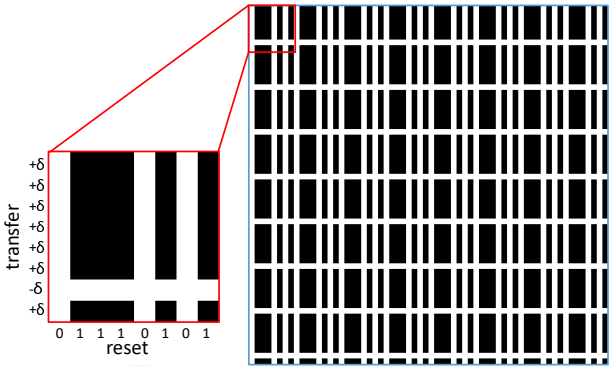


図 3.8: **露光パターンと制御信号の関係**．右側は 64×64 画素の露光パターン，左側は 8×8 の露光パターンとそれに対応するリセット信号と転送信号を表す．白色は露光する画素を，黒色は露光しない画素を表す．リセット信号では 0 が OFF を，1 が ON を表す．$\pm\delta$ は転送信号のタイミングを表す．

並ぶ．これは，リセット信号が OFF または転送信号のタイミングが $-\delta$ の画素が露光するからである．

3.3　符号化露光からの動画像再構成

従来より様々な方法で，符号化露光画像から空間分解能を損なうことなくセンサのフレームレートよりも高いフレームレートの動画像が再構成されてきた．本論文では以下の方法で動画像を再構成する．

Hitomi ら [20] はシーンのスパース性を仮定し，圧縮動画センシングの観測とシーンの再構成をモデル化した．$\boldsymbol{x}(w,h,t)$ をターゲットシーン，$\boldsymbol{\phi}(w,h,t)$ を符号化パターンとすると，符号化露光した観測画像 $\boldsymbol{y}(w,h)$ は以下のように表すことができる．

$$\boldsymbol{y}(w,h) = \sum_{t=1}^{N} \boldsymbol{\phi}(w,h,t) \odot \boldsymbol{x}(w,h,t), \tag{3.1}$$

ここで，N は観測画像におけるサブフレーム数，$\odot$ は要素ごとの積である．

さらに，$\boldsymbol{x}(w,h,t)$ のサブフレームが基底となる動きの疎な線形結合で表現できると仮定すると，この基底 $\mathbf{D} = [D_1(w,h,t), D_2(w,h,t), \cdots, D_K(w,h,t)]$ を用いて $\boldsymbol{x} = \mathbf{D}\boldsymbol{\alpha}$ と表現できる (α は線形結合係数)．この表現を用いると，式 (3.1) は以下のように書き換えることができる．

$$\boldsymbol{y} = \boldsymbol{\phi}\boldsymbol{x} = \boldsymbol{\phi}\mathbf{D}\boldsymbol{\alpha}. \tag{3.2}$$

基底 $\mathbf{D}$ は多くの動画を用いて k-means 法により学習される．$\mathbf{D}$，$\boldsymbol{\phi}$，$\boldsymbol{y}$ を用いて，OMP [57] などのスパース再構成法により $\hat{\boldsymbol{\alpha}}$ を推定することで元のシーンを再構成する．

$$\arg\min \parallel \alpha \parallel_0 \; s.t. \parallel \boldsymbol{y} - \boldsymbol{\phi}\mathbf{D}\boldsymbol{\alpha} \parallel_2^2 \leq \varepsilon. \tag{3.3}$$

再構成されたシーンの動画像 $\hat{\boldsymbol{x}}$ は $\hat{\boldsymbol{x}} = \mathbf{D}\hat{\boldsymbol{\alpha}}$ と表される．再構成は $M \times M$ 画素のパッチごとに行われ，それぞれ位置を変えながら M^2 回再構成を行う．最終的な再構成結果 $x(w,h,t)$ は M^2 回の再構成の平均を取ったものである．

Yang ら [25] は混合ガウス分布 (Gaussian Mixture Model: GMM) をもとにした再構成手法を提案した．動画をパッチに分け，このパッチ $\{\boldsymbol{x}_p\}$ が以下のように表されると仮定した．

$$\boldsymbol{x}_p \sim \sum_{k=1}^{K} \lambda_k \mathcal{N}(\boldsymbol{x}_p \mid \mu_k, \Sigma_k) \tag{3.4}$$

ここで，$\mathcal{N}$ ガウス分布を表し，K，$\mu_{\mathtt{k}}$，Σ_k と λ_k はそれぞれ，GMM の要素数，平均，共分散行列，k 番目の要素の重み ($\lambda_k > 0$ and $\sum_{k=1}^{K} \lambda_k = 1$) である．よって，$\boldsymbol{x}_p$ の条件付き期待値を計算することにより，動画像を再構成することができる．Yang ら [25] は再構成するパッチを水平/垂直方向へ $\frac{M}{2}$ 画素ずらして再構成し，平均を取ることで再構成品質向上させた．

Iliadis ら [26] は近年主流となっている深層学習を用いた再構成手法を提案した．全結合層より構成されるネットワークにより，元のシーンである動画と符号化露光画像間の非線形写像を学習する．ネットワークは入力層，4 層の隠れ層，出力層で構成され，入力層のサイズは符号化露光画像のサイズであり，隠れ層と出力層のサイズは再構成する動画のサ

イズである．訓練動画と再構成された動画の誤差を最小化することにより，ネットワークを訓練する．平均二乗誤差と密接な関係にあるピーク信号対雑音比 (Peak Signal-to-Noise Ratio: PSNR) を品質指標とするため，損失関数として平均二乗誤差 (Mean Squared Error: MSE) を用いる．スパース再構成や GMM を用いた手法と比べ，再構成の際に繰り返し計算が必要ないことから高速に再構成を行うことができる．

3.3.1　読み出しゆがみの補正

本研究で用いる QPE センサはグローバルシャッタセンサを元に設計されているが，準画素毎露光制御を行うと読み出す際にゆがみが生じる．図 3.9 と 3.10 にグローバルシャッタと符号化露光の違いを示す．また，図 3.9-(a)，3.10-(a) に各画素の，サブフレーム番号，露光パターン番号，転送タイミング，読み出しタイミング，露光タイミングを示し，図 3.9-(b)，3.10-(b) に図 3.9-(a)，3.10-(a) の赤線のタイミングで読み出された電荷が露光されたタイミングを示す．図 3.11-(b)，(c) はシーン (a) を 3.9-(a) と 3.10-(a) の露光方法で観測した画像である．図 3.9-(a)，(b) と 3.10-(a)，(b) は y–t スライスを示しているが，図 3.9-(c) と 3.10-(c) は x–y スライスを示す. グローバルシャッタの場合，(図 3.9)，転送信号は 1 フレームに 1 回入力され，電荷はすべての画素で同時に C_{fd} へ転送される．そのため，C_{fd} から読み出される電荷は全画素で同じ時間情報を持っている (図 3.9-(b))．よって，読み出された画像はゆがまない．

一方，準画素毎露光制御により画像を観測すると (図 3.10)，サブフレームの最後にフォトダイオードから C_{fd} に電荷が転送され (図 3.10-(a) の青矢印の位置)，行ごとに C_{fd} から電荷が読み出される (図 3.10-(a) の赤線の位置)．そのため，転送信号の後に読み出される行は前の行の 1 サブフレーム後の時間情報を持つこととなる．図 3.10-(b) において，3 行目と 4 行目のように同じサブフレーム内で読み出される画素は同じ時間情報を持つが，4 行目と 5 行目のように違うサブフレームで読み出される画素は 1 サブフレームずれた時間

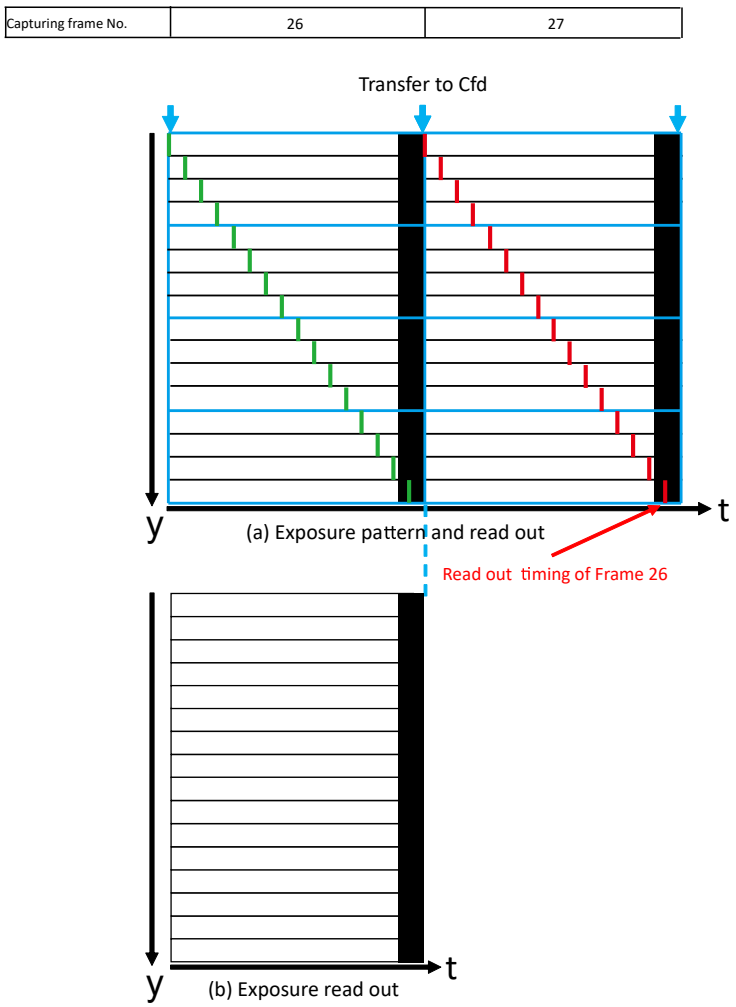


図 3.9: **QPE センサを用いたグローバルシャッタによる撮像**．(a): 観測フレーム番号，C_{fd} への電荷の転送タイミング (青矢印)，それぞれの画素の露光時間，フレーム 25 と (緑線) とフレーム 26(赤線) の読み出しタイミング．ここでは，簡単のため 16 行のみの 2 フレームの撮像について示している．(b): 各 C_{fd} へ転送された電荷が実際にフォトダイオードでとらえられたタイミング．

情報を持つ．(例えば，1〜4 行目は 101〜104 サブフレームの情報を持ち，5〜8 行目は 102〜105 サブフレームの情報を持つ．) よって，読み出された画像は (センサの行数/サブフレーム数) 行のブロックごとに 1 サブフレームずれることによりゆがむ．また，読み出しタ

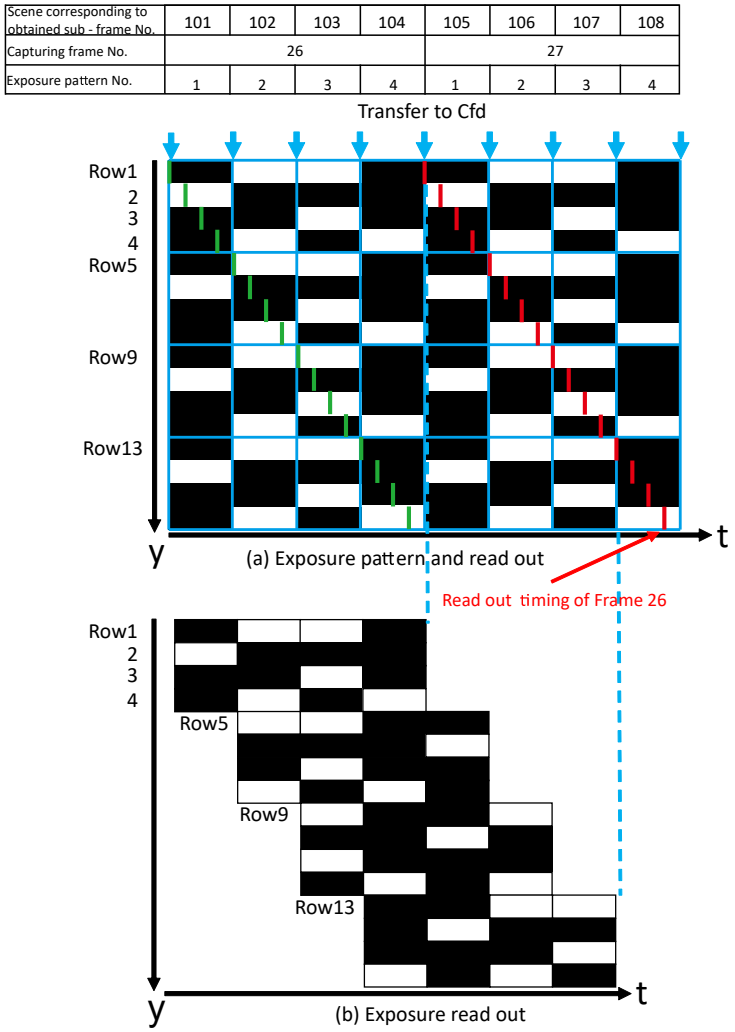


図 3.10: **QPE センサを用いた準画素毎露光制御による撮像**.(a): サブフレーム番号，観測フレーム番号，露光パターン番号，C_{fd} への電荷の転送タイミング (青矢印)，それぞれの画素の露光時間，その時点での露光パターン，フレーム 25 と (緑線) とフレーム 26(赤線) の読み出しタイミングを示す．ここでは，8 サブフレームの 2 フレームのみ示す．(b): 撮影タイミングに合わせた露光パターン．

イミングが異なるため，ブロックごとに露光パターンの順序が変わる (図 3.10-(b))．そのため，露光パターンの順序を考慮して再構成する必要がある．図 3.11 にグローバルシャッタと準画素毎露光制御により撮影した画像の例を示す．図 3.11-(a) のような白い円が移動

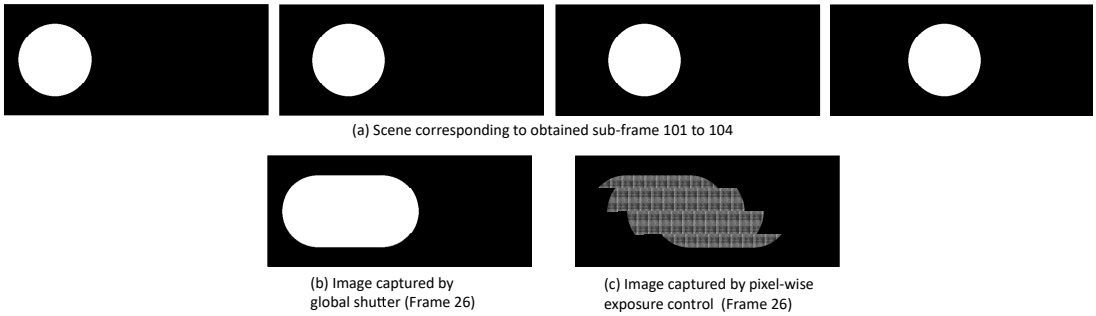


図 3.11: **グローバルシャッタと準画素毎露光制御により撮影した画像の例**．(a): 白い円が左から右へフレーム 101 から 104 にかけて移動するシーン．(b): グレーバルシャッタにより撮影した例．(c): 準画素毎露光制御により撮影した例．

しているシーンをグローバルシャッタで撮影すると図 3.11-(b) のような画像となる．一方，準画素毎露光制御により撮影すると図 3.11-(c) のようにブロックごとに撮影されるタイミングが異なり，ゆがんで撮影される．図 3.12 は図 3.11-(c) の観測画像から動画像を再構成した結果と，再構成した結果から読み出しゆがみを補正したものである．観測画像から再構成された動画像は図 3.12-(a) のようにゆがんでいるが，ブロックごとにサブフレームがずれていることを考慮することにより，そのゆがみを補正することができる (図 3.12-(b))．

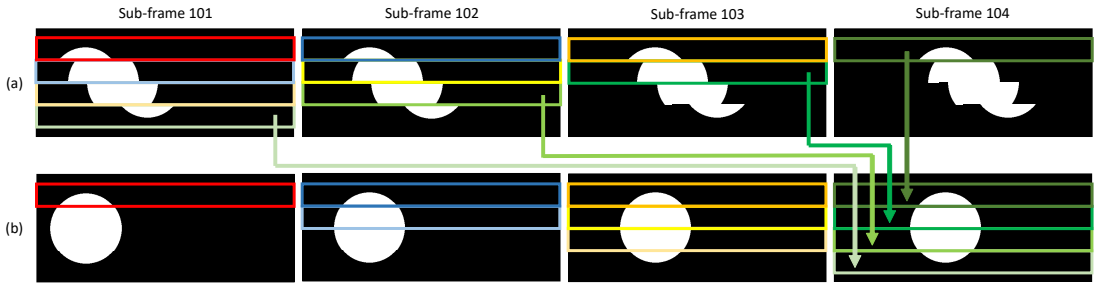


図 3.12: **読み出しゆがみの補正**．(a): QPE センサにより撮影した画像から再構成した動画．これは読み出しの過程でゆがんでいる．(b): 読み出しゆがみを補正した動画．(a) の近傍サブフレームを読み出しタイミングを考慮しずらすことで歪みを補正する．同じ色の領域は同じサブフレームであることを示す．

図 3.12-(b) において，同じ色の範囲は同じサブフレームを表し，同じ色の部分をつなぎ合わせることで正しいサブフレームを再構成することができる．

3.4　実験

3.4.1　シミュレーション実験

図 3.13 に，本研究で用いる QPE センサによる準画素毎露光により撮影された画像から再構成されたサブフレームの品質を示す．画素毎ランダム露光 [20]，QPE センサによる通常符号化露光，最適化画素毎露光による撮影結果と比較した．本シミュレーション実験ではサブフレーム数 $N = 16$ とし，再構成手法には深層学習を用いた手法 [26] を用いた．図 3.13 には再構成された 16 サブフレームのうち 3 フレームのみを示している．また，高時空間解像度の理想的な動画，異なる 4 種類の露光で撮影された符号化露光画像，およびそれらから再構成されたサブフレームを示す．

画素毎ランダム露光によって撮影された画像は十分なランダム性を持っているため，シーンを復元することができる．また，QPE センサによる通常符号化露光画像からもシーンを復元することができるが，エッジ部分が不鮮明であるなどランダム露光と比べ再構成品質が良くない．一方，提案する準画素毎露光では通常符号化露光と比べエッジが鮮明であるなど画素毎ランダム露光から再構成されたシーンに近く，理想に近い再構成が行えている．この結果から，準画素毎露光はシーンを再構成するために必要なランダム性を持っていることがわかる．このように，本研究で用いる QPE センサは高空間解像度で高時間解像度のシーンを観測することができるといえる．各露光パターンの特徴は以下の通り．

- Iliadis ら [46]: 時空間方向にランダムであるため，時空間方向の様々な周波数の情報を観測できる．

- Hitomi ら [20]: 空間方向はランダムであるため，空間方向では様々な周波数の情報

を観測できるが，時間方向にはセンサの物理的な制約からランダム性に制限があるため，観測できる情報に限りがある．

- 提案する準画素毎露光: 時間方向はランダムであるため様々な周波数の情報を観測できるが，空間方向には制約があるため観測できる空間周波数には限りがある．

Hitomi ら [20] の時間方向に制約のある露光パターンのほうが我々の空間方向に制約のある露光パターンより再構成品質が良い理由は，動画に含まれる時間情報は空間情報と比べバリエーションが限られており，時間情報の周波数が少ないため，空間方向の制約が時間方向の制約より影響を受けやすいからである．

辞書ベースのスパース最適化手法 (OMP, LASSO) [20,58]，GMM を用いた手法 [25]，および深層学習を用いた手法 [26] の再構成品質及びノイズに対する頑健性を比較した．すべての手法で同じ訓練データを用い，予備実験の結果より，辞書ベースのスパース最適化手法 (OMP, LASSO) では辞書の要素数を 5000 に，GMM の要素数を 20 に設定した．シミュレーションにより作成した符号化露光画像に対し，平均 0，分散 0，0.01，0.05 の白色ガウスノイズを加えた．従来手法 [20,25,26] の多くは撮影画像をパッチに切り分け，パッチごとに再構成を行っていた．しかし，パッチごとに再構成を行うとそれぞれのパッチは独立に再構成されるため，パッチの境界で画素値が不連続となりブロック状のアーティファクトが現れる．そこで，深層学習以前の繰り返し計算による再構成手法の多くはパッチのウィンドウをスライドさせながら再構成を行い，重なり合った部分を平均化することでこの問題を解決した．全結合層のみを用いた深層学習で再構成を行った Iliadis ら [26] のネットワークでは，入力画像の露光パターンの位置とネットワークの入力層の位置が一対一対応するため再構成するウィンドウは露光パターンの位置に固定される．再構成ウィンドウをスライドさせると露光パターンの位置もずれるため，その位置に合わせた再構成ネットワークを用意する必要がある．Iliadis ら [26] は露光パターンを再構成パッチのサイズ 8×8 画素の半分である 4×4 画素のパターンの繰り返しとすることで，1 つの再構成ネッ

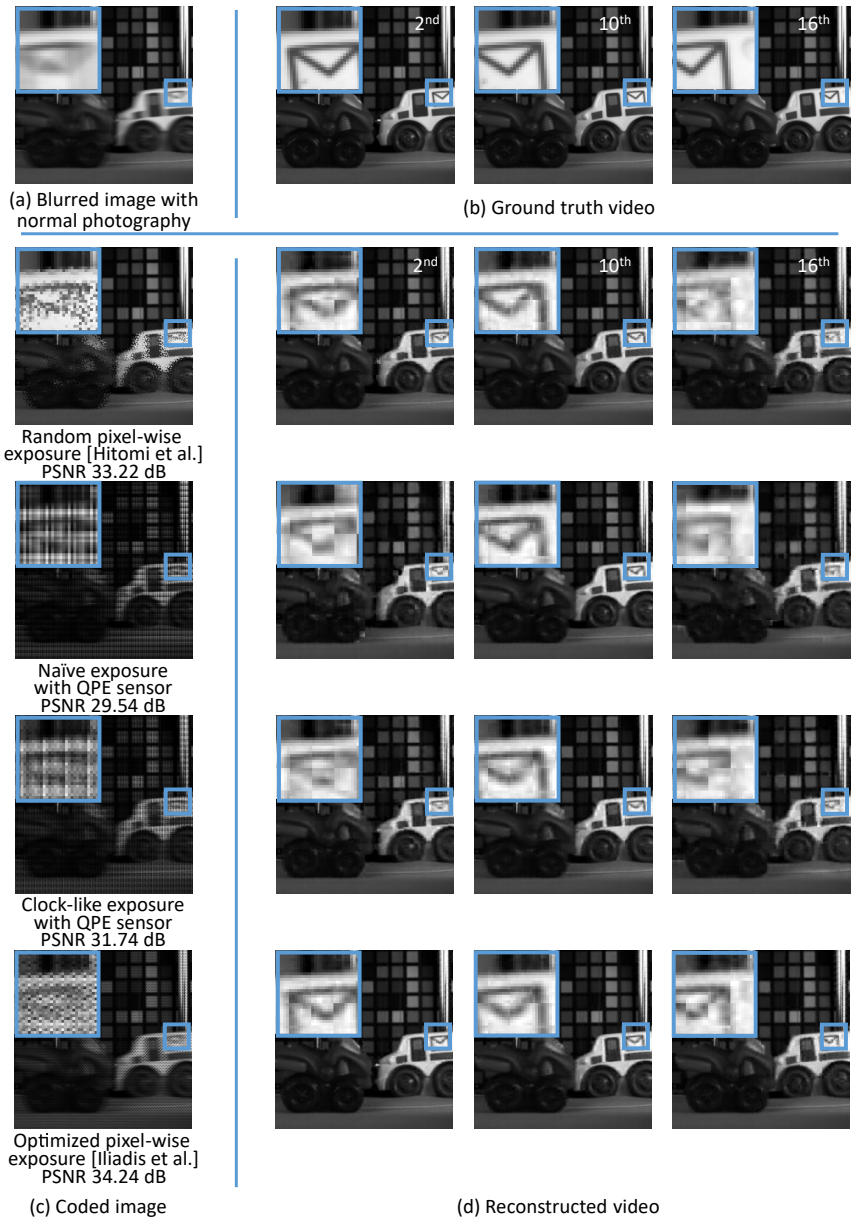


図 3.13: **露光パターンの比較**．(b): 高時空間解像度の理想的な動画．ここでは 16 フレームのうち 3 フレームのみ示す．(a): 時間解像度の低いスチルカメラで撮影された画像の例 (16 フレームの平均を取った画像)．(c): ランダム露光 [20] による撮影，QPE センサによる通常撮影，QPE センサによる準画素毎露光制御による撮影，最適化した画素毎露光制御による撮影 [46] をシミュレートした画像．(d): 深層学習により再構成された動画．(b) と同様のサブフレームのみ示す．

トワークでパッチサイズの半分である 4×4 画素のスライド幅でウィンドウをスライドすることを可能とした．本論文ではそれぞれの論文に基づきウィンドウのスライド幅を，辞書ベースのスパース最適化手法 (OMP, LASSO) [20, 58] では 1×1 画素，GMM を用いた手法 [25] では 4×4 画素とした．また，深層学習を用いた手法では Iliadis ら [26] 以外の露光パターンは 8×8 画素の繰り返しのため，Iliadis ら [26] のようにウィンドウのスライド幅を 4×4 画素とすることはできないため，ウィンドウのスライド幅を 8×8 画素とした．OMP，LASSO，GMM，深層学習の各手法で再構成されたノイズレベルが異なる結果を図 3.14 に示す．どの手法もノイズレベルが上がるにつれて画質が劣化し，PSNR が低下した．また，ノイズを加えた場合でも各手法間で再構成品質の順位に変化はなかった．表 3.1 に各手法による再構成品質 (PSNR) を示す．画素毎ランダム露光 [20] では深層学習による再構成が最も品質が良く，その他 3 つの露光パターンにおいては GMM による再構成が最も品質が良かった．しかし，深層学習による再構成品質は GMM に近く再構成速度は圧倒的に早いため，深層学習による再構成が最も現実的であるといえる．

再構成するパッチウィンドウをスライドさせることにより，再構成品質を向上させることができる．Hitomi ら [20] はウィンドウを縦横に 1 画素ずつずらしながら再構成し，Yang ら [25] はウィンドウのスライド幅をウィンドウサイズの半分に設定した．ウィンドウのスライド幅を小さくすると再構成品質が上がるが，その分計算コストが増えるため再構成品質と計算コストはトレードオフの関係にある．そこで，スライディングウィンドウの幅が再構成の頑健性に与える影響を評価した．OMP，LASSO，GMM の各手法においてスライディングウィンドウ幅を変えたときの再構成品質 (PSNR) と再構成時間を表 3.2 に示す．実験に用いた計算機のプロセッサは Intel Core i7-6900K，メインメモリは 64GB，OS は Ubuntu 16.04 LTS である．図 3.15 に各手法においてスライディングウィンドウの幅を変化させたときの再構成結果を示す．図 3.15 左列の OMP ではスライディングウィンドウ幅を狭くすることにより再構成品質が向上している．図 3.15 中列の LASSO ではごま塩ノイズのようなアーティファクトが減少している．図 3.15 右列の GMM ではスライディン

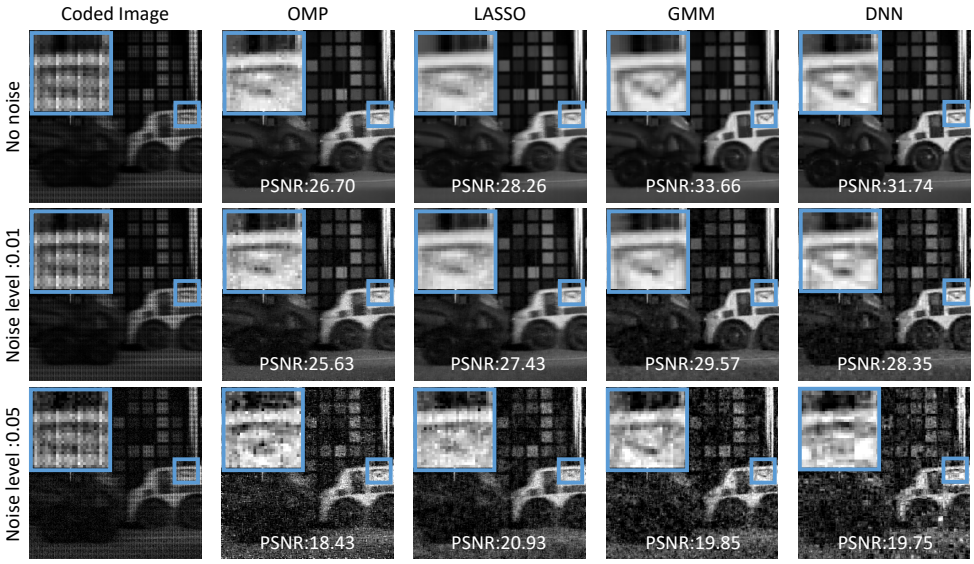


図 3.14: **ノイズ頑健性の比較**. パッチサイズを 8×8 画素とし，スライディングウィンドウ幅を辞書ベースのスパース最適化手法 (OMP, LASSO) では 1×1 画素，GMM では 4×4 画素，深層学習 (DNN) では 8×8 画素とした.

グウィンドウ幅を狭くすることによりブロックノイズが低減している．OMP では画質が大きく改善されたが，LASSO と GMM では大きな改善はしなかった．また，各パッチの再構成結果は OMP では不安定であるが，他の方法では安定していた.

3.4.2　試作センサによる実験

図 3.1 に示す試作 CMOS センサにより準画素毎露光を行うことで提案手法の実現性を実証した．試作センサの仕様を表 3.3 に示す．センサの構造については第 3.1 章で述べたとおりである．このセンサのダイナミックレンジは 60dB であり，露光は Field-Programmable Gate Array (FPGA) によって制御されている．試作センサにて撮影した画像から深層学習 [26] を用いてサブフレームを再構成した.

試作センサを用いて走っている列車を撮影した．図 3.18 に撮影した画像と対応する露

表 3.1: 異なるノイズレベルにおける再構成品質 (PSNR)

	ノイズ レベル	画素毎 ランダム露光	QPE センサ 通常符号化露光	QPE センサ 準画素毎露光	最適化 画素毎露光
OMP	0	24.66	22.41	22.96	26.04
[20]	0.01	24.46	21.30	22.46	25.62
	0.05	21.56	13.81	17.59	20.98
LASSO	0	25.01	23.04	23.71	26.93
[58]	0.01	24.91	22.40	23.38	26.51
	0.05	23.35	16.29	19.52	22.36
GMM	0	27.69	**26.89**	**28.18**	**29.72**
[25]	0.01	27.54	**24.84**	**26.27**	**28.15**
	0.05	24.58	18.22	**19.18**	**21.73**
深層学習	0	**28.65**	26.57	27.81	29.19
[26]	0.01	**28.33**	24.72	25.99	28.10
	0.05	**24.82**	**18.56**	18.88	21.65

光パターン，再構成したサブフレームと読み出しゆがみを除去したサブフレームを示す．図 3.18-(a) は符号化された撮影画像である．画像は連続して撮影されているが，ここではそのうち 3 フレームのみを示す．また，すべてのフレームでセンサーの有効画素 656×496 すべてを使って撮影を行ったが，図 3.18(b)-(f) では動きのある領域 (赤枠で囲まれた領域) に注目した．センサのフレームレートを 15FPS に設定し，サブフレーム数を 16 に設定したため，再構成されるサブフレームは 240FPS 相当となる．高速で変化するシーンをカメラで撮影すると，被写体がブレて写る．我々はこのブレを符号化するために図 3.18-(b) のような露光パターンを設定した．よって，図 3.18-(a) には符号化されたブレが見受けられる．読み出しゆがみ (第 3.3.1 章参照) と本研究で用いるセンサの駆動方式 (第 3.2 章参照)

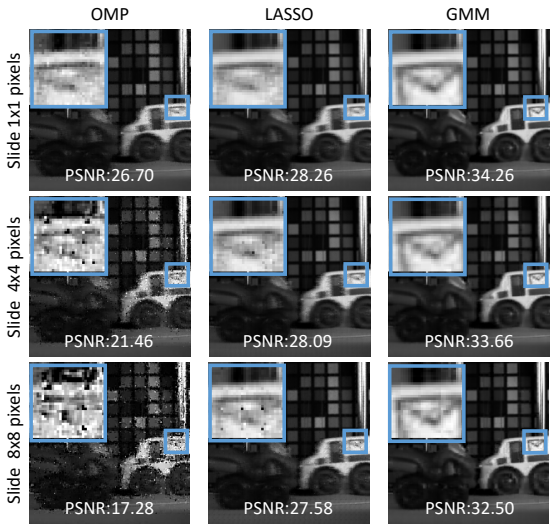


図 3.15: **スライディングウィンドウ幅による再構成の頑健性の比較**. 辞書ベースの手法 (OMP, LASSO) および GMM について，再構成時のスライディングウィンドウ幅による頑健性を調査した.

により，設定した露光パターンと実際にセンサ上で適用される露光パターンが異なる．露光パターンは図 3.18-(b) のようにセンサ全体で同じであるが，読み出しタイミングが図 3.10 のように列ごとに異なるため，図 3.18-(c) のようにゆがんだ露光パターンで撮影することとなる．この例では総画素 512 行に対して 16 サブフレームで撮影しているため，32 行ごとに異なる露光パターンで構成されている．この露光パターンを用いてフレーム f からサブフレームを再構成すると読み出しゆがみを含むため，図 3.18-(d) のような結果となる．シーンの動きは復元されているが，読み出しゆがみにより列車の全面が斜めにゆがんでいる．最初と最後のサブフレームは他のサブフレームと比べ再構成エラーが多い傾向にある．これはすべての再構成手法に共通する特徴であり，これは端のサブフレームは中間のサブフレームと比べ前後のサブフレームの情報が少ないからである．図 3.18-(e) は再構成歪みを補正した結果である．図 3.18-(e) のサブフレームはフレーム f と $f-1$ の情報から再構成されている．シーンの動きが復元され，また，ゆがみが除去されたことにより列

表 3.2: 異なるスライディングウィンドウ幅による再構成品質と再構成時間の比較

		OMP	LASSO	GMM
1×1	PSNR(dB)	22.96	23.71	**28.48**
	Time(Sec)	7.04×10^2	1.02×10^4	2.40×10^3
4×4	PSNR(dB)	20.00	23.59	**28.18**
	Time(Sec)	4.66×10^1	6.17×10^2	1.58×10^2
8×8	PSNR(dB)	17.06	23.41	**27.44**
	Time(Sec)	1.38×10^1	1.78×10^2	4.43×10^1

表 3.3: 試作した QPE センサーの仕様

画素サイズ	7.4×7.4 μm
総画素数	672×512 pixels
有効画素数	656×496 pixels
フレームレート	7, 15, 30, 71 fps
受光エリアサイズ	4.8544×3.6704 mm
フィルファクタ	約 36%
画素構成	5 トランジスタ, 1 キャパシタ

車前面が垂直になっている．また，図 3.18-(f) はゆがみ除去後のサブフレームのうち，フレーム f に含まれる部分を示したものである．図 3.18-(f) の黒い部分をフレーム $f-1$ から補うことで図 3.18-(e) が得られる．

図 3.16，3.17 にほかの撮影結果を示す．図 3.16 上段はコーヒーを注ぐ様子を，図 3.16 下段は水中へ硬貨を落とす様子を撮影したものである．図 3.16-(a) で符号化された動きが図 3.16-(b) で再構成されていることがわかる．また，図 3.17-(a) は瞬きを撮影した例で，図 3.17-(b) では動きが再構成されていることがわかる．図 3.17-(c)，(d) は試作センサを用

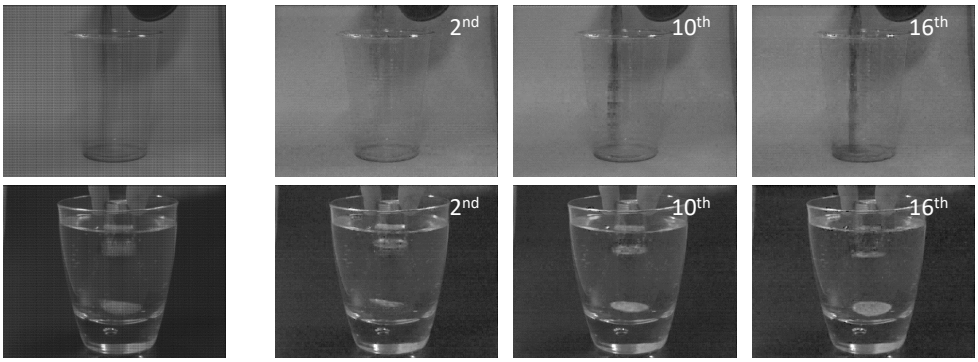


図 3.16: **プロトタイプ QPE センサによる撮影結果**．プロトタイプセンサにより，コーヒーを注ぐシーンおよび水中へ硬貨を落とすシーンを撮影した．(a): プロトタイプセンサにより実際に撮影したシーン．(b): 再構成した動画のうちの 3 フレーム．

いて通常のグローバルシャッタで撮影した例で，図 3.17-(c) は露光時間を $1/15$ 秒 (15 FPS 相当) とし，図 3.17-(d) は露光時間を $1/240$ 秒 (240 FPS 相当) とした．図 3.17-(c) は時間分解能が低いため動きのある部分がブレており不鮮明である．また，図 3.17-(d) は鮮明な画像であるが，連続した画像ではないため動きが飛び飛びである．

3.5　まとめ

本研究では準画素毎露光可能な CMOS センサ (QPE センサ) を用い，符号化露光の自由度を向上させる露光制御信号と読み出しによる歪みを抑えた圧縮センシングによる動画像の高速撮像手法を提案した．標準的な CMOS センサと比べわずかな設計の変更のため，感度，フレームレート，解像度などの性能は標準的な CMOS センサと同等である．提案した信号パターンは，露光制御の観点から空間的な制約があるにもかかわらず，サブフレームの再構成に必要な画素毎ランダム露光を実現した．また，学習型の再構成手法を用いた動画像の再構成を行い，読み出しゆがみの除去手法を提案した．提案手法の実現可能性と効率性を，試作した CMOS センサによる実実験により実証した．

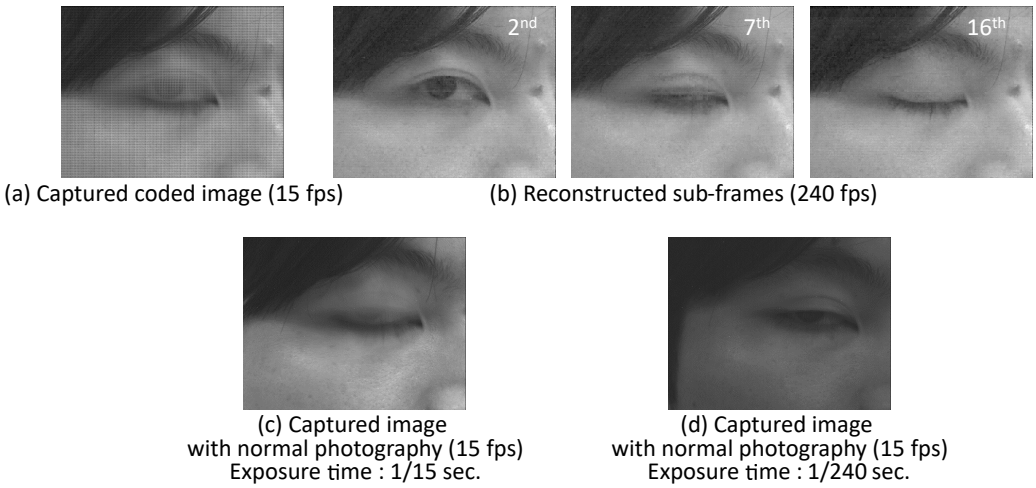


図 3.17: **プロトタイプ QPE センサによる撮影モードの違い**．(a): QPE センサで準画素毎露光制御により撮影した画像．(b): 準画素毎露光制御により撮影した画像から再構成されたサブフレーム．(c): シャッタ速度を 1/15 秒として QPE センサでグローバルシャッタにより撮影した画像 (15FPS 相当)．(d): シャッタ速度を 1/240 秒 QPE センサでグローバルシャッタにより撮影した画像 (240FPS 相当).

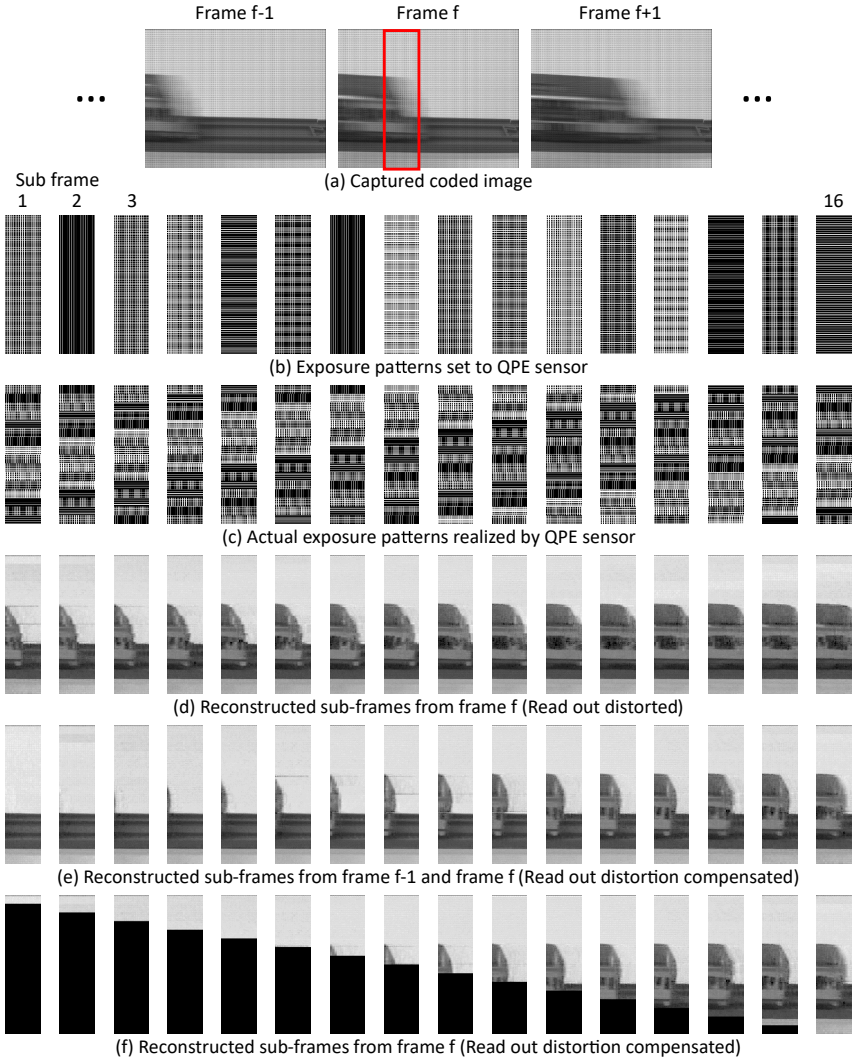


図 3.18: **試作センサによる実実験: 列車走行シーンと使用した露光パターン及び再構成したサブフレーム**．図 3.1 に示す試作センサを用いて撮影を行った．(a): 3 つの連続した撮影フレーム．(b)-(f) ではフレーム f において動きのある領域である赤枠で囲まれた領域のみを示す.(b): 試作 QPE センサへ設定した露光パターン．ここでは 16 サブフレームすべてを示す．(c): 実際にセンサ上で適用される露光パターン．(d): 読み出しゆがみを除去する前の再構成結果．(e): 読み出しゆがみ除去後の再構成結果．(f): ゆがみ除去後のサブフレームのうち，フレーム f に含まれる部分のみ．

第 4章

制約のある露光の符号化と動画像再構成の同時最適化

デジタルカメラは光を利用して周囲の環境の情報を観測する．しかし，イメージセンサは通常 2 次元に配置されたフォトダイオードで構成されているが，光によって取得できる情報は空間，時間，色など高次元であるため，環境の情報を直接観測することはできない．そこで，高次元な情報を観測行列を用いて 2 次元に写像して観測する必要がある．観測行列を用いた画像の撮影は式 1.1 のように表せる．時空間情報を観測する場合は以下のように書き換えることができる．

$$\boldsymbol{y}(w,h) = \sum_{t}^{N} \boldsymbol{\phi}(w,h,t) \odot \boldsymbol{x}(w,h,t), \tag{4.1}$$

ここで，$\boldsymbol{y} \in \mathbb{R}^{W \times H}$ は 2 次元の撮影画像，$\boldsymbol{\phi} \in \mathbb{R}^{W \times H \times N}$ は観測行列，$\boldsymbol{x} \in \mathbb{R}^{W \times H \times N}$ は対象のシーン，$\odot$ は要素ごとの積である (W: 画像の幅，H: 画像の高さ，N: フレーム数)．撮影した画像 $\boldsymbol{y}$ から元のシーン $\boldsymbol{x}$ を再構成するが，$\boldsymbol{y}$ は $\boldsymbol{x}$ よりも次元が低いため劣決定問題である．本研究で扱う時空間情報の観測では，観測行列は露光の制御により実装するため，露光パターンと表現する．従来，時空間情報は隣接画素の露光タイミングをずらし空間解像度を犠牲にすることで取得されてきた [59, 60](図 4.1-(a))．この撮像方式では空

間解像度が $\frac{1}{N}$(N: フレーム数) に低下する．超解像手法を用いて元の空間解像度まで復元することもできるが，そのような手法は均一な撮影にて取得した少数のデータを用いた推定のため，復元性能には限界がある [61]．Gupta ら [60] は，領域ごとに空間情報を取得するか時間情報を取得するかを選択できる手法を提案したが，空間次元と時間次元のトレードオフを根本的に解決したわけではない．また，Shen ら [62] は長時間露光フレームと短時間露光フレームを交互に露光することで，低ノイズ，高フレームレートの動画を再構成する手法を提案した．これに対し，圧縮センシングによる動画像撮像は空間解像度を低下させることなく空間情報と時間情報を同時に取得する (図 4.1-(b))．時空間符号化露光により画像を撮影することで，時空間情報を 1 枚の画像に畳み込み，後処理で動画を再構成する．深層学習が登場する以前は L0 ノルム正則化 [20, 21]，L1 ノルム正則化 [22–24]，動きの推定とカルマンフィルタを組み合わせた手法 [31] や混合ガウスモデル [25] を用いた手法など反復最適化によって動画像の再構成が行われていた．近年，深層学習の導入により，再構成の品質と速度が向上している [26, 27, 63–65].

動画の再構成品質に影響を及ぼすものとして，符号化露光による観測の品質と観測画像からの再構成の品質があげられる．符号化露光の品質向上を考える際には，撮影に使用するセンサの制約を考慮する必要があるが，センサの制約は非常に複雑であり，手作業で最適化するためには職人技が必要となる．また，撮影するシーンによっても最適な符号化というのは変わる．一方，再構成手法の性能を左右するものとして，使用するモデルや学習データと再構成しようとするシーンの分布の類似度も関係してくる．ここで，符号化露光と再構成の手法は全く無関係でなく，互いに依存関係にある．また，符号化露光による撮影は時空間情報の撮影画像への畳み込み，再構成は撮影画像から動画への逆畳み込みと表現できる．このデータを特徴量に畳み込み，逆畳み込みにより元のデータに復元するというタスクはオートエンコーダ [33] と同じである．そこで，オートエンコーダのように再構成だけでなく露光パターンも深層学習のネットワークで表現することで，深層学習を用いて露光パターンと再構成を同時最適化する手法が提案されている [46, 47, 66, 67]．この

ようなアプローチはディープセンシングと呼ばれ，観測と再構成を別々に最適化する手法よりも良い結果を得ることができる．

露光タイミングをずらすことによる符号化露光は時空間次元の畳み込み，撮影画像からの動画の再構成は逆畳み込みと表現できるため，ディープセンシングを用いて露光パターンと再構成を同時最適化することができる [46,47](図 4.1-(b))．しかし，時空間情報の撮影画像への畳み込みはセンサでの符号化露光により表現する必要があるため，様々な制約がある．まず，通常のオートエンコーダでは特徴量は任意のベクトルであるが，本研究の時空間符号化ではカメラで畳み込む必要があるため，畳み込まれた結果は画像として取得できる形でなければならない．また，光や画像は 0 以下の値を持たないため，通常のニューラルネットワークと異なり，ネットワークの重みや出力はすべて 0 以上である必要がある．さらに，センサでの符号化には露光制御のメカニズムによる制約があるが，それをニューラルネットワークで表現することは困難である．例えば，画素ごとに露光タイミングをずらすためには画素ごとに露光を制御する必要があるが，これを実現する CMOS センサはハードウェアの制約上現実的ではない．通常のイメージセンサは，グローバルシャッターやローリングシャッターのように一様に画像を撮影するように設計されており，画素ごとに露光を制御するためには制御線など新たな部品を追加する必要がある．これにより，各画素の回路が複雑になるためフォトダイオードのサイズを小さくせざるを得なくなり，このような特殊なセンサーでは感度と解像度を両立させることが困難である [38](図 4.2-(a))．さらに，3 トランジスタ CMOS センサのような標準的な CMOS センサは画素ごとのバッファを持たず，非破壊で多重露光することはできない [20]．そのため，このようなセンサでは 1 回の読み出しで 1 回の露光しか行えない (図 4.2-(b))．そこで，第 3 章で提案したようなより柔軟に露光を制御できるセンサがあるが，このセンサは空間的な制御に制約がある (図 4.2-(c))．そのため，使用するイメージセンサの制約を考慮しながら露光の符号化を最適化する必要がある．また，CMOS センサで露光のタイミングを切り替える際は電子シャッタを利用するため，ON/OFF の 2 値でしか制御できない．そのため CMOS センサに

よる露光制御では露光パターンは (0,1) の 2 値となるが，深層学習では微分可能な関数しか扱えないため，直接露光パターンを最適化することはできない．これも圧縮動画センシングにおいて，深層学習による符号化露光のための露光パターン最適化の際のハードウェア制約の一つである．

そこで，本研究では制約のある露光パターンと動画像再構成の同時最適化手法を提案する．本研究の貢献は以下の通り．

- 深層学習を用いて符号化露光のための観測行列 (露光パターン) と動画像再構成の同時最適化手法を提案する．撮像に使用するハードウェアの制約を考慮し，露光パターンと再構成デコーダを End-to-End で同時最適化する．
- 提案手法では制約なしを含む様々なハードウェア制約のある露光パターンを最適化することができる．深層学習では微分可能な関数しか扱えないが，露光パターンは 2 値であり，様々なハードウェア制約を持つため直接最適化することは困難である．そこで，深層学習におけるこのような制約のある重みの更新手法を提案する．
- 提案手法による同時最適化により，既存のランダムな露光パターンによる符号化露光画像と比べ高品質な再構成が行えることを実証した．また，最適化された露光パターンを実際にセンサに実装し，符号化露光画像の撮影と高フレームレートな動画の再構成が行えることを確認した．

4.1　制約のある露光の符号化と動画像再構成の同時最適化

本章では圧縮センシングによって動画像を撮影するタスクにおいて，符号化露光のための露光パターンと再構成の同時最適化手法を提案する．提案する深層学習のネットワークは大きく分けて 2 つの部分より構成される．1 つ目は観測 (符号化) 層であり，露光パター

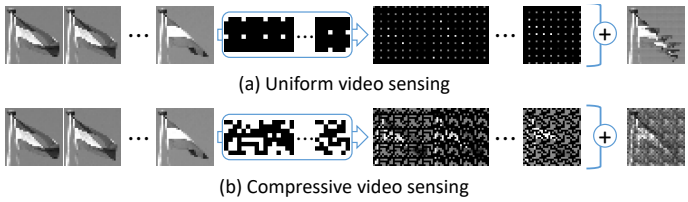


図 4.1: **露光パターンによる撮影画像の違い**．撮影画像は露光パターンとシーンの要素ごとの積の和として表される．(a): 均一な露光パターンによる均一な動画撮像．(b): 不均一な符号化露光パターンによる動画の圧縮センシング.

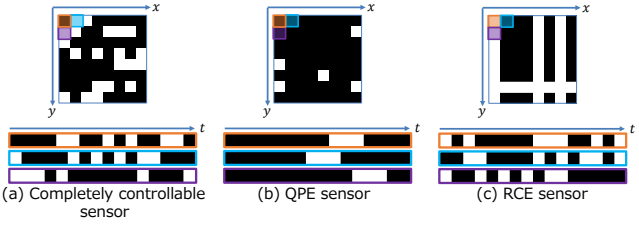


図 4.2: **ハードウェア制約下における露光パターンの例**．(a): 完全に画素ごとに露光制御可能なセンサ．(b): 時間方向の制御に制約のあるセンサ [20](SBE センサ)．(c): 空間方向の制御に制約のあるセンサ (QPE センサ).

ンを使用するハードウェアの制約の中で最適化する．2 つ目は再構成 (復号) 層であり，最適化された露光パターンで符号化された 1 枚の撮影画像から複数のサブフレームを再構成する．全体的な枠組みを図 4.3 に示す.

4.1.1　観測層

再構成 (復号) 層から誤差を伝播させ観測 (符号化) 層の重みを更新することで，高品質な動画を再構成できるよう露光パターンを最適化する．ここで，観測層の重みは実際のセンサに実装するために制約がある．CMOS センサに露光パターンを実装するためには，電子シャッタによって露光の ON/OFF を切り替える必要があるため，露光パターンは (0,1) の 2 値である必要がある．また，露光制御の信号線やトランジスタの配置により，各セン

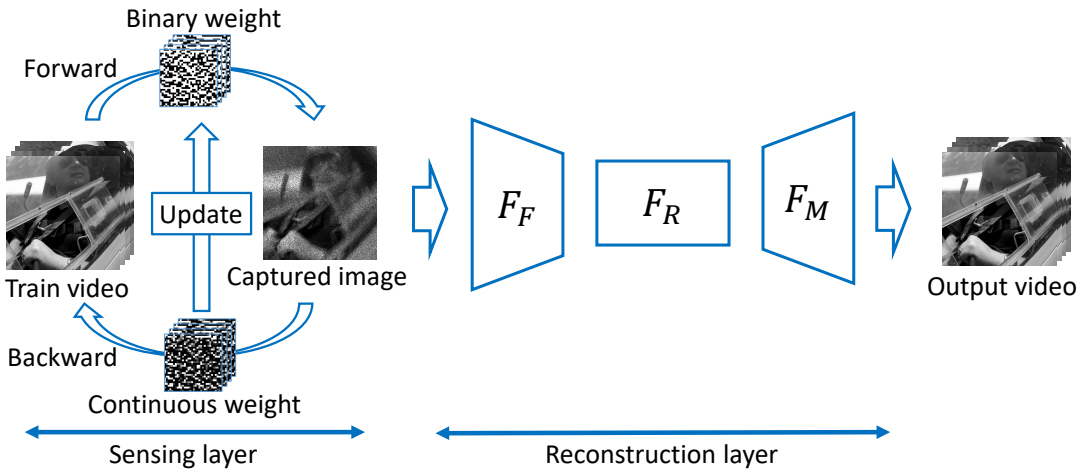


図 4.3: **提案手法のネットワーク構造**．提案するネットワークは圧縮センシングにおける露光パターンと再構成を深層学習によって同時最適化する．左側は露光パターンによって動画を 1 枚の画像に圧縮する観測層，右側は 1 枚の撮影画像と複数のサブフレームの間の写像を学習する再構成層である．(F_F: 特徴抽出層．F_R: 特徴量レベルの非線形写像層．F_M: 再構成層．)

サに依存した時間方向または空間方向に制約がある [20,21]．そこで，本研究ではニューラルネットワークの重みを 2 値化することで計算コストやメモリ使用量を抑えることを目的とした Binaryconnect [68] や，露光パターンの様々な時間的制約を考慮した最適化手法 [69] を参考に観測層を設計した．本研究では以下の 3 つのセンサの制約を考慮する．

1. 電子シャッタによって露光を制御するため，露光パターンは 2 値であるが，時空間的な制約はない理想的なセンサ．(制約なしセンサ).

2. 露光パターンは 2 値であり，時間的な制約のあるセンサ (SBE センサ).

3. 露光パターンは 2 値であり，空間的な制約のあるセンサ (QPE センサ).

このような制約下でニューラルネットワークを学習することは困難である．ニューラルネットワークでは誤差逆伝播によって勾配を計算し重みを更新するが，勾配を計算するた

めにはネットワークの重みが連続値である必要がある．しかし，露光パターンは 2 値である必要があり，勾配計算が困難である．そのため，通常のニューラルネットワークでは露光パターンを最適化することはできない．そこで，露光パターンを最適化する観測層において，順伝播では露光パターンと同様に 2 値の重みを用い，逆伝播では勾配計算が可能なように連続値重みを使用することでハードウェアの制約下で露光パターンの最適化を行う．観測層の学習順序は以下の通り．

1. ハードウェアの制約を満たすランダムなパターンで順伝播重み (2 値重み)，逆伝播重み (連続値重み) を初期化する．
2. 順伝播では露光パターンと同様に 2 値の重みにより実際のカメラを模擬し，シーンを 1 枚の画像へ畳み込む．
3. 逆伝播では重みを連続値のものへ切り替え，元のシーンと再構成後の動画の誤差を逆伝播させることにより，連続値重みを更新する．
4. 手順 3 で更新した連続値重みを各ハードウェアに対応した関数に入力することで，ハードウェアの制約を満たす 2 値に変換し，順伝播用 2 値重みを更新する．
5. ネットワークの更新が収束するまで手順 2 から 4 を繰り返す．

露光制御の制約がないセンサでは，各画素が独立してサブフレーム単位でシャッタの ON/OFF を制御できる．露光は電子シャッタによって制御されるため，露光パターンは 2 値となる．しかし，露光パターンに空間的，時間的な制約がないため，2 値重みの更新関数は以下のような閾値処理を行う．

$$\boldsymbol{w}_b(i,j,t) = \begin{cases} 1\ (\boldsymbol{w}_c(i,j,t) \geq Threshold) \\ 0\ (\boldsymbol{w}_c(i,j,t) < Threshold) \end{cases} \tag{4.2}$$

ここで，$\boldsymbol{w}_b(i,j,t)$ は画素位置 (i,j) のサブフレーム t における 2 値重み，$\boldsymbol{w}_c(i,j,t)$ は画素位置 (i,j) のサブフレーム t における連続値重みである．

時間的な制約のあるセンサ (SBE センサ) では各画素独立にシャッタの ON/OFF を制御できるが，電荷の非破壊読み出しができないため，1 フレームの間に 1 回しか露光ができない．また，センサのダイナミックレンジが限られるため，すべての画素で露光時間を同じにする必要がある．そのため，このような制約を考慮した以下のような関数により連続値重みを 2 値重みへ変換する．SBE センサには空間的制約がないため，各画素独立に処理を行う．

$$\boldsymbol{w}_b(i,j,t) = \begin{cases} 1\ (\tau \leq t < \tau + N) \\ 0\ (other) \end{cases} \tag{4.3}$$

$$\tau = \arg\max_{\tau}[\sum_{t=\tau}^{\tau+N-1} \boldsymbol{w}_c(i,j,t)] \tag{4.4}$$

ここで，N は露光するサブフレーム数 (露光時間) である．

空間的な制約のあるセンサ (QPE センサ) では，露光の制御線が縦または横の画素で共有するため，露光する画素が縦または横に並ぶような空間的制約がある．しかし，各画素は 1 フレーム中で複数回の露光が可能であり，露光時間もそろえる必要はない．そこで，連続値の 2 値への変換関数は各列の連続値重みを畳み込むことでそれぞれの縦または横の列が露光するかしないかを決定する．RCE センサには時間的制約がないため，この処理はサブフレームごとに独立に行う．

$$\boldsymbol{w}_b(i,j,t) = \begin{cases} 1\ (\boldsymbol{w}_v(i,t) > 0\ or\ \boldsymbol{w}_h(j,t) > 0) \\ 0\ (other) \end{cases} \tag{4.5}$$

$$\boldsymbol{w}_v(i,t) = \sum_{j}^{W} \boldsymbol{w}_c(i,j,t),\ \boldsymbol{w}_h(j,t) = \sum_{i}^{H} \boldsymbol{w}_c(i,j,t) \tag{4.6}$$

ここで，$\boldsymbol{w}_v(i,t)$ は縦の露光制御信号に対する重み，$\boldsymbol{w}_h(j,t)$ は横の露光制御信号に対する重みである．

4.1.2　再構成層

再構成層として，RevSCI-net [29] を用いる．RevSCI-net は特徴抽出層 F_F，特徴量レベルの非線形写像層 F_R，再構成層 F_M の 3 つの部分より構成されている．

特徴抽出層 F_F では，3 次元 CNN 層 (カーネルサイズ: 5×5，3×3，1×1，3×3) を用いて入力の高次元特徴を抽出する．特徴抽出層 F_F の入力は以下の式によって正規化され，$W \times H$ から $W \times H \times N$ へ拡張される．

$$\bar{\boldsymbol{y}} = \boldsymbol{y} \oslash \sum_{t=1}^{N} \boldsymbol{\phi}, \quad \bar{\boldsymbol{x}} = \bar{\boldsymbol{y}} \odot \boldsymbol{\phi} \tag{4.7}$$

ここで，$\oslash$ は要素ごとの商，$\odot$ は要素ごとの積，$\bar{\boldsymbol{y}} \in \mathbb{R}^{W \times H}$ は正規化された入力，$\bar{\boldsymbol{x}} \in \mathbb{R}^{W \times H \times N}$ は拡張された入力である．3 次元へ拡張された入力 $\bar{\boldsymbol{x}}$ は特徴抽出層 F_F で 4 次元の特徴量 ($\mathbb{R}^{W \times H \times C \times N}$) へ変換される．

特徴抽出層 F_F の出力を動画特徴に変換する非線形写像層 F_R は，メモリ使用量を減らすために積層型可逆ブロックを使用している．RevSCI-net の元となった Rev-Net [70] は特徴量を 2 チャネルに分割しているが，RevSCI-net では動画の再構成性能を向上させるため，特徴量を複数のチャネルへ分割する．ResNet ブロックは誤差逆伝播で勾配を計算する際に，すべての中間層の活性度を保存するために多くのメモリを必要とするが，特徴量を複数チャネルへ分割することで最後の活性度のみを保存するだけで勾配が計算できるため，メモリ使用量を大幅に削減することができる．

再構成層 F_M は，非線形写像層 F_R の出力を入力とし，再構成された動画 (サブフレーム) を出力する．再構成層 F_M は 3 次元 CNN 層 (カーネルサイズ: 3×3，3×3，1×1，3×3) で構成されている．

4.2　実験

4.2.1　実験設定

提案するネットワークはオートエンコーダと同様の構造であるため，入力と出力の誤差を最小化するように学習した．画質の評価としてピーク信号対雑音比 (PSNR) を用いるため，損失関数には平均二乗誤差 (MSE) を用いた．

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{N}\sum_{k=1}^{N}(\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k)^2 \tag{4.8}$$

ここで，$\hat{\boldsymbol{x}}_k$ は k 番目の再構成されたサブフレーム，$\boldsymbol{x}_k$ は元のシーンの k 番目のフレームである．訓練用のデータセットは，動画におけるオブジェクトセグメンテーションのタスクのためのデータセットである DAVIS2017 [71] を用いた．このデータセットに対し，回転と反転を加えた 6864 本の動画を訓練に用いた．最適化アルゴリズムとして Adam を用い，学習率は 0.00002，バッチサイズを 8 に設定した．

4.2.2　シミュレーション実験

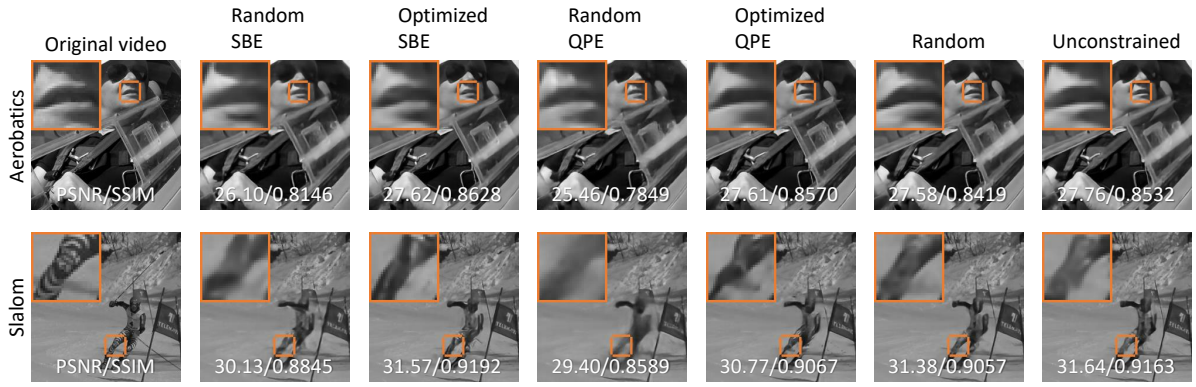


図 4.4: **再構成結果**．シーン Aerobatics の 10 番目のサブフレーム (上側) とシーン Slalom の 14 番目のサブフレーム (下側) を示す．各図の左上に枠で囲まれた範囲の拡大図を示す．

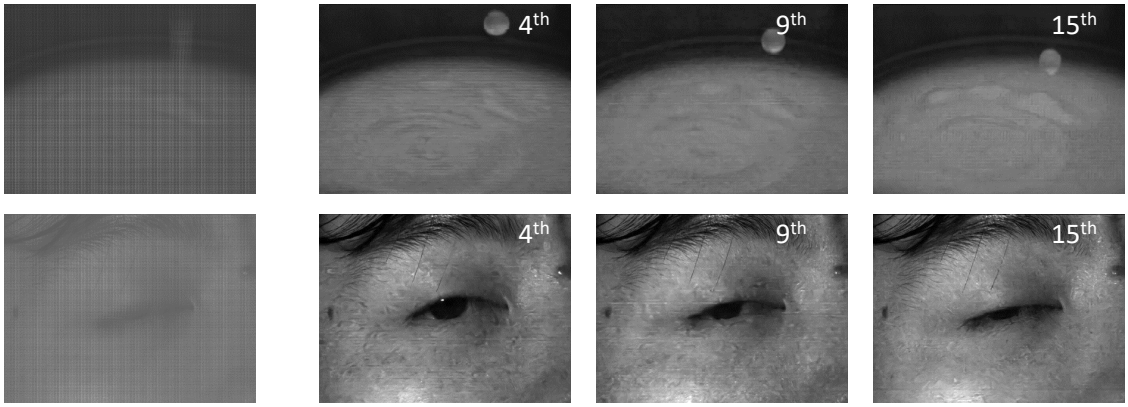
表 4.1: 33 本の検証動画の平均再構成品質 (PSNR/SSIM)

	SBE センサ	QPE センサ	制約なしセンサ
最適化あり	30.75 / 0.9033	30.23 / 0.9004	31.31 / 0.9062
最適化なし	29.05 / 0.8686	28.15 / 0.8463	31.28 / 0.9024

露光パターンと動画像再構成の同時最適化による効果を評価するために，シミュレーション実験を行った．露光パターンと動画像再構成を同時に最適化した結果と，ランダムに生成した露光パターンを使用して再構成のみを学習した結果を比較した．ハードウェアの制約として，露光パターンに時間的な制約のある SBE センサ [20](図 4.2-(b))，露光パターンに空間的な制約のある QPE センサ (図 4.2-(c))，制約なしセンサ (図 4.2-(a)) の 3 つを考慮した．制約なしセンサは画素ごとに独立して露光を制御でき，圧縮センシングにおいて理論的に理想とされる完全ランダムな露光パターンを実現することができる．本実験では $W \times H \times N$ を $256 \times 256 \times 16$ に設定した．SBE センサ，QPE センサ，制約なしセンサの制約を満たす，ランダムと最適化した露光パターンで撮影した画像をシミュレーションにより作成し，再構成ネットワークへ入力した．出力結果と元のシーンを比較することで再構成品質を定量的に評価した．評価指標には，信号がとりうる最大値とノイズの比であるピーク信号対雑音比 (PSNR) と，人間の主観評価に近い評価として画像の構造に注目した指標である SSIM を用いた．評価には 256×256 画素，16 サブフレームの 33 本の動画を用いた．図 4.4 にシーン Aerobatics とシーン Slalom の元のシーンと各再構成結果を示す．図 4.4 の上側の Aerobatics を見ると，最適化されたパターンのほうがより唇の形が元動画に近く再構成されている．また，図 4.4 の下側の Slalom では，最適化していないランダムなパターンでは膝の形状がブレて再構成されているが，最適化されたパターンではシャープに再構成されている．表 4.1 に 33 本の検証動画に対し，SBE センサ，QPE センサ，制約なしセンサの制約において，ランダムパターンと最適化パターンによって再構成した結果の平均 PSNR と SSIM を示す．すべてのハードウェア制約において，ランダムな

露光パターンを用いた結果よりも露光パターンと再構成を同時最適化した結果のほうが再構成品質が良かった.

4.2.3　プロトタイプセンサによる実験



(a) Captured coded image (71 fps)　(b) Reconstructed sub-frames (1136 fps)

図 4.5: **実シーンの再構成結果**.　(a): 撮影画像.　(b): 再構成された 4 番目，9 番目，15 番目のサブフレーム.

最適かられた露光パターンが実機へ実装可能であることを示すため，第 3 章で提案したセンサを搭載したセンサにて実際のシーンを符号化撮影した．実験に使用したカメラの外観を図 3.1 に示す．また，このセンサの仕様は表 3.3 に示す．露光は FPGA によって制御されている．画像の撮影は 71FPS で行った．また，1 フレーム当たり 16 サブフレームに設定したため，再構成後の動画は 1,136FPS 相当となる．提案手法で最適化した露光パターンをセンサの制御信号へ変換し，FPGA にてセンサの露光を制御した．撮影画像は第 3.3.1 章で述べたように読み出しゆがみにより露光パターンが場所によってずれるため，ずれた露光パターンを用いて再構成層を再学習した．再学習した再構成層に撮影画像を入力し，出力されたサブフレームを第 3.3.1 章の手法で読み出しゆがみを除去した．図 4.5 に実際に撮影した画像と再構成したサブフレームを示す．図 4.5 の上段は白い水滴が落ち

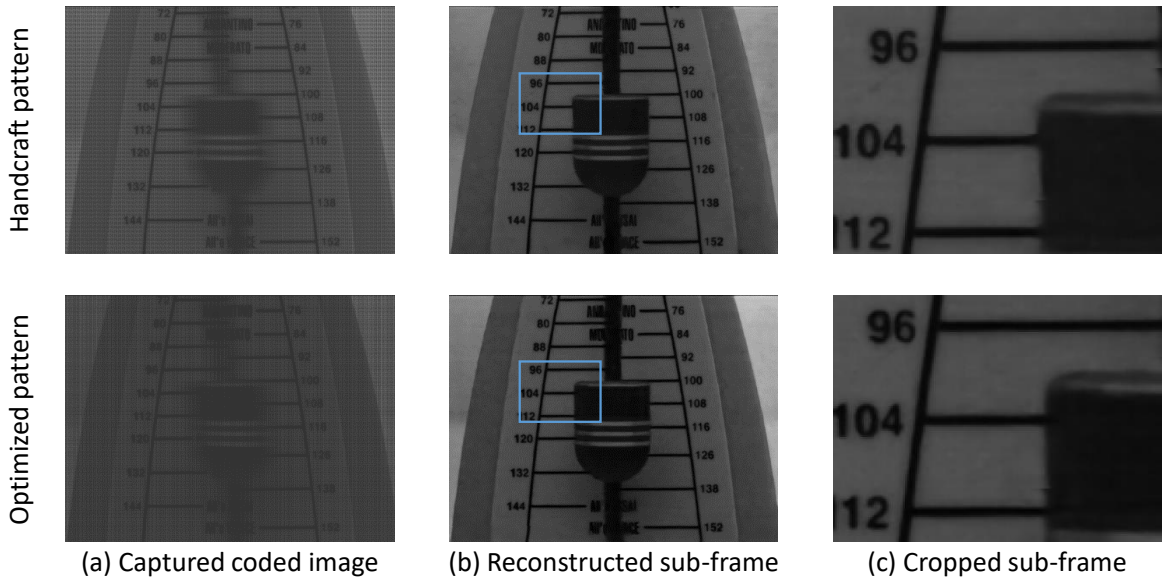


図 4.6: **ランダムパターンと最適化パターンの比較**．(a): 撮影画像．(b): 再構成されたサブフレーム．(c): (b) の枠内の拡大表示.

るシーンであり，水滴が落ちる様子が復元されていることがわかる．図 4.5 の下段はまばたきの様子であり，まぶたの閉じていく様子がわかる．このように，シミュレーション実験だけでなく，提案手法により撮影画像からサブフレームを再構成することができた．図 4.5 のシーンは訓練データに含まれるシーンとは大きく異なるものであり，再構成ネットワークの汎化性能が高いことがわかる．図 4.6 は試作センサによる露光パターンの比較である．上段がランダムパターン，下段が提案する最適化した露光パターンによる結果である．時間情報を符号化するため，符号化露光の制限として，センサのダイナミックレンジが制限される．これは，画素ごとに露光時間が異なる場合，最も露光時間の長い画素が飽和しないように撮影する必要があるためである．一部の画素が飽和するとその画素の情報が欠落するため，正しく再構成が行えない．しかし，通常の画像が 8bit であるに対し，試作センサの出力画像は 14bit であるため，この差を利用して時間情報を畳み込むことでダイナミックレンジの犠牲を最小限とすることができる．

4.3　まとめ

本章では露光制御可能なセンサの多くはその露光制御のための構造によりさまざまな制約があるため，その制約を考慮して露光パターンを最適化する手法を提案した．深層学習を用い，撮影の過程と再構成を両方ニューラルネットワークで表現することで観測と再構成を同時最適化することを可能とした．一般的なニューラルネットワークではさまざまなハードウェア制約を表現することは困難だが，提案手法ではさまざまなハードウェア制約をみたす露光パターンの最適化を可能とした．本章では SBE，QPE，制約なしの 3 種類のハードウェア制約に対し，露光パターンを最適化し，動画を再構成できることを示した．また，各ハードウェア制約において露光パターンを最適化したことにより，最適化していないランダムなパターンと比較して再構成品質が向上したことをシミュレーション実験と実機実験にて確認した．

第 5 章

結論

本研究ではカメラを用いて取得できる情報量を拡張することを目的とし，圧縮センシングの考え方により動画の符号化撮像と圧縮画像から動画を再構成を行うことで，イメージセンサで時空間情報を取得した．符号化撮像のためのセンサとその符号化の自由度を上げる手法，およびセンサの制約を満たす露光パターンと動画の再構成を同時最適化する手法を提案した．本章では本研究の成果を第 5.1 節で，今後の課題を第 5.2 節で述べる．

5.1　研究成果

実世界は様々な情報を含んでいるため高次元であり，その情報を 2 次元のイメージセンサで取得することは困難である．しかし，シーンの時空間情報は多くの冗長性を含んでおり，均一な観測では多くの無駄な情報を含む．そこで，本論文では圧縮センシングを用い，符号化露光によって時間情報を 1 枚の画像に畳み込むことが可能なセンサと，符号化露光の露光パターンと再構成を同時最適化する手法を提案した．

動画の圧縮センシングの多くの従来研究は，シミュレーション実験しかせず実際のシーンの再構成を行わない研究や，反射光学系を用いた疑似実装にとどまりセンサのみでの符

号化を行わない研究である．しかし，実際の現場で使用する際にはシミュレーションのようにうまくいくことは少なく，反射光学系を用いた実装では撮像装置が大きくなるため実用的ではない．そこで，本論文では，センサのみで動画の圧縮センシングに必要な符号化露光を行う手法を提案した．提案するセンサでは，一般的なグローバルシャッタセンサの配線の一部のみを変更することにより，センサの感度悪化を最小限に抑えながら符号化露光を可能とした．また，提案するセンサでは露光を制御する配線が限られるため露光制御の自由度が低いが，制御信号を工夫することにより準画素毎露光制御を可能とする手法を提案した．さらに，提案するセンサにて準画素毎露光制御を行うと読み出された画像にゆがみが生じるが，この読み出しゆがみは再構成後のサブフレーム単位でずれることにより生じる．そこで，サブフレームごとに画像をずらすことで読み出しゆがみを除去する手法を提案した．

符号化露光可能なイメージセンサでは，露光制御の自由度とセンサの感度やダイナミックレンジにトレードオフの関係がある．そのため，センサの感度やダイナミックレンジを確保するためには露光制御の自由度を制限せざるを得ない．よって，イメージセンサのポテンシャルを最大限発揮するために露光パターンを最適化する必要がある．しかし，どのような情報を観測すれば再構成品質が上がるかは再構成しないとわからない．そこで，ハードウェアによる露光制御の制約を考慮しながら露光パターンと動画の再構成を同時最適化する手法を提案した．提案手法により露光パターンを最適化することで，センサの構造や再構成手法を変更せずとも再構成品質を改善した．

5.2　今後の課題

本研究では，圧縮センシングにより時間情報を 1 枚の画像に畳み込み，再構成する手法を提案し，その有効性を示した．そこで，今後の課題として画像に畳み込む時間情報の分解能の向上と，画像に畳み込む情報の次元数増加があげられる．

時間情報を用いて距離を測ることのできる Time-of-Fright (ToF) センサというものがある．ToF センサは，光源が発光してから対象に反射しセンサに到達するまでの時間を計測することで，距離が取得可能なセンサである．よって，ToF センサの距離精度は時間分解能に依存するが，センサの時間分解能の向上は容易ではない．そこで，圧縮センシングを用いて符号化 ToF センサにより時間情報を圧縮した画像を撮影し，後処理により時間情報を抽出することで距離測定性能を向上させることができる．ToF センサの撮像素子は通常の CMOS センサと大きく異なるため，観測と距離推定を同時最適化するためには，センサによる観測を表現するニューラルネットワークの設計を独自に検討する必要がある．

光線により取得できる情報は空間情報だけに限らず，時間や光の波長など多岐にわたる．本研究では時空間の 3 次元情報を 2 次元のイメージセンサで観測する手法を提案したが，時空間の 3 次元と波長の合計 4 次元へ拡張することは可能である．光の波長の情報は通常一様なフィルタを用いて空間解像度を犠牲にして取得される．そこで，このフィルタの配置や分光透過率を露光パターンや再構成と同時最適化する手法を検討する．

謝辞

本研究を進めるにあたり，大阪大学データビリティフロンティア機構長原一教授には，指導教員として日ごろから多岐にわたりご指導，ご支援いただいた．ここに深く感謝申し上げる．大阪大学産業科学研究所八木康史教授には，博士後期課程進学後 1 年間研究室所属させていただき様々なご支援をいただいただけでなく，今回副査としてご助言いただいたことをここに感謝する．大阪大学データビリティフロンティア機構中島悠太准教授には，副査として論文にご助言いただくとともに，日々の研究室生活において様々なご支援をいただいたことに感謝申し上げる．大阪大学産業科学研究所中村友哉准教授には，本研究に対して活発に議論していただき，副査としてご助言いただいたことに感謝する．浜松ホトニクス株式会社杉山行信氏，遠藤健太氏には，本研究に必要なイメージセンサを提供いただき，センサの仕様や様々な疑問に答えていただいたことにより，円滑に研究を進めることができた．ここに感謝する．大阪大学データビリティフロンティア機構，大阪大学大学院情報科学研究科知能センシング講座各位には，日々の研究室生活において様々な支援をいただいた．ここに感謝の意を表する．最後に，研究に専念できるよう私生活を支えていただいた家族に感謝する．

参考文献

[1] S. Kleinfelder, S. Lim, X. Liu, and A. El Gamal, “A 10000 Frames/s CMOS Digital Pixel Sensor,” *IEEE Journal of Solid-State Circuits*, vol. 36, no. 12, pp. 2049 –2059, 2001.

[2] S. Niklaus, L. Mai, and F. Liu, “Video frame interpolation via adaptive convolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 670–679.

[3] S. Meyer, A. Djelouah, B. McWilliams, A. Sorkine-Hornung, M. Gross, and C. Schroers, “Phasenet for video frame interpolation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 498–507.

[4] W. Bao, W.-S. Lai, C. Ma, X. Zhang, Z. Gao, and M.-H. Yang, “Depth-aware video frame interpolation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3703–3712.

[5] W. Shen, W. Bao, G. Zhai, L. Chen, X. Min, and Z. Gao, “Blurry video frame interpolation,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5114–5123.

[6] H. Lee, T. Kim, T.-y. Chung, D. Pak, Y. Ban, and S. Lee, “Adacof: Adaptive collaboration of flows for video frame interpolation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5316–5325.

[7] B. Wilburn, N. Joshi, V. Vaish, M. Levoy, and M. Horowitz, “High-Speed Videography using a Dense Camera Array,” in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2004, pp. 294–301.

[8] A. Gupta, P. Bhat, M. Dontcheva, O. Deussen, B. Curless, and M. Cohen, “Enhancing and Experiencing Space-Time Resolution with Videos and Stills,” in *Proceedings of International Conference on Computational Photography (ICCP)*, 2009, pp. 1–9.

[9] A. Velten, D. Wu, A. Jarabo, B. Masia, C. Barsi, C. Joshi, E. Lawson, M. Bawendi, D. Gutierrez, and R. Raskar, “Femto-photography: capturing and visualizing the propagation of light,” *ACM Transactions on Graphics (ToG)*, vol. 32, no. 4, pp. 1–8, 2013.

[10] G.-H. Chen, J. Tang, and S. Leng, “Prior image constrained compressed sensing (piccs): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets,” *Medical physics*, vol. 35, no. 2, pp. 660–663, 2008.

[11] Y. S. Han, J. Yoo, and J. C. Ye, “Deep residual learning for compressed sensing ct reconstruction via persistent homology analysis,” *arXiv preprint arXiv:1611.06391*, 2016.

[12] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse mri: The application of compressed sensing for rapid mr imaging,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[13] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing mri,” *IEEE signal processing magazine*, vol. 25, no. 2, pp. 72–82, 2008.

[14] T. E. C. et al., “First m87 event horizon telescope results. i. the shadow of the supermassive black hole,” *ApJL*, vol. 875, p. 1, 2019. [Online]. Available: https://iopscience.iop.org/article/10.3847/2041-8213/ab0ec7

[15] ——, “First m87 event horizon telescope results. ii. array and instrumentation,” *ApJL*, vol. 875, p. 2, 2019. [Online]. Available: https://iopscience.iop.org/article/10.3847/2041-8213/ab0c96

[16] ——, “First m87 event horizon telescope results. iii. data processing and calibration,” *ApJL*, vol. 875, p. 3, 2019. [Online]. Available: https://iopscience.iop.org/article/10.3847/2041-8213/ab0c57

[17] ——, “First m87 event horizon telescope results. iv. imaging the central supermassive black hole,” *ApJL*, vol. 875, p. 4, 2019. [Online]. Available: https://iopscience.iop.org/article/10.3847/2041-8213/ab0e85

[18] ——, “First m87 event horizon telescope results. v. physical origin of the asymmetric ring,” *ApJL*, vol. 875, p. 5, 2019. [Online]. Available: https://iopscience.iop.org/article/10.3847/2041-8213/ab0f43

[19] ——, “First m87 event horizon telescope results. vi. the shadow and mass of the central black hole,” *ApJL*, vol. 875, p. 6, 2019. [Online]. Available: https://iopscience.iop.org/article/10.3847/2041-8213/ab1141

[20] Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. K. Nayar, "Video from a single coded exposure photograph using a learned over-complete dictionary," in *Proceedings of International Conference on Computer Vision (ICCV)*, 2011, pp. 287–294.

[21] T. Sonoda, H. Nagahara, K. Endo, Y. Sugiyama, and R. Taniguchi, "High-speed imaging using cmos image sensor with quasi pixel-wise exposure," in *Proceedings of International Conference on Computational Photography (ICCP)*. IEEE, 2016, pp. 1–11.

[22] Y. Liu, M. Li, and D. A. Pados, "Motion-aware decoding of compressed-sensed video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 3, pp. 438–444, 2013.

[23] M. Azghani, M. Karimi, and F. Marvasti, "Multihypothesis compressed video sensing technique," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 4, pp. 627–635, 2016.

[24] C. Zhao, S. Ma, J. Zhang, R. Xiong, and W. Gao, "Video compressive sensing reconstruction via reweighted residual sparsity," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 6, pp. 1182–1195, 2017.

[25] J. Yang, X. Yuan, X. Liao, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, "Video compressive sensing using gaussian mixture models," *IEEE Transactions on Image Processing*, pp. 4863–4878, 2014.

[26] M. Iliadis, L. Spinoulas, and A. K. Katsaggelos, "Deep fully-connected networks for video compressive sensing," *Digital Signal Processing*, vol. 72, pp. 9–18, 2018.

[27] J. Ma, X. Liu, Z. Shou, and X. Yuan, "Deep tensor admm-net for snapshot compressive imaging," in *Proceedings of International Conference on Computer Vision (ICCV)*, 2019, pp. 10 223–10 232.

[28] M. Qiao, Z. Meng, J. Ma, and X. Yuan, "Deep learning for video compressive sensing," *Apl Photonics*, vol. 5, no. 3, p. 030801, 2020.

[29] Z. Cheng, B. Chen, G. Liu, H. Zhang, R. Lu, Z. Wang, and X. Yuan, "Memory-efficient network for large-scale video compressive sensing," in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 16 246–16 255.

[30] Y. Sun, X. Chen, M. S. Kankanhalli, Q. Liu, and J. Li, "Video snapshot compressive imaging using residual ensemble network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 9, pp. 5931–5943, 2022.

[31] X. Ding, W. Chen, and I. J. Wassell, "Compressive sensing reconstruction for video: An adaptive approach based on motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 7, pp. 1406–1420, 2017.

[32] P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, "Coded aperture compressive temporal imaging," *Optics express*, vol. 21, no. 9, pp. 10 526–10 545, 2013.

[33] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[34] M. Dadkhah, M. Deen, and S. Shirani, "Compressive Sensing Image Sensors-Hardware Implementation," *IEEE Sensors Journal*, vol. 13, no. 4, pp. 4961–4978, 2013.

[35] R. Robucci, J. Gray, L. K. C., J. Romberg, and P. Hasler, "Compressive sensing on a cmos separable-transform image sensor," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1089–1101, June 2010.

[36] M. Dadkhah, M. J. Deen, and S. Shirani, "Block-based CS in a CMOS image sensor," *IEEE Sensors Journal*, vol. 14, no. 8, pp. 2897–2909, 2014.

[37] V. Majidzadeh, L. Jacques, A. Schmid, P. Vandergheynst, and Y. Leblebici, "A (256x256) Pixel 76.7mW CMOS Imager/ Compressor Based on Real-Time In-Pixel Compressive Sensing," *In Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2956–2959, 2010.

[38] M. Wei, N. Sarhangnejad, Z. Xia, N. Gusev, N. Katic, R. Genov, and K. N. Kutulakos, "Coded two-bucket cameras for computer vision," in *Proceedings of European Conference on Computer Vision (ECCV)*, September 2018.

[39] Y. Oike and A. El Gamal, "CMOS image sensor with per-column $\Sigma\Delta$ ADC and Programmable Compressed Sensing," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 1, pp. 318–328, 2013.

[40] L. Spinoulas, K. He, O. Cossairt, and A. Katsaggelos, "Video compressive sensing with on-chip programmable subsampling," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2015-Octob, pp. 49–57, 2015.

[41] N. Antipa, P. Oare, E. Bostan, R. Ng, and L. Waller, "Video from stills: Lensless imaging with rolling shutter," in *2019 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2019, pp. 1–8.

[42] X. Yuan, “Generalized alternating projection based total variation minimization for compressive sensing,” in *Proceedings of International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 2539–2543.

[43] Y. Inagaki, Y. Kobayashi, K. Takahashi, T. Fujii, and H. Nagahara, “Learning to capture light fields through a coded aperture camera,” in *Proceedings of European Conference on Computer Vision (ECCV)*, 2018, pp. 418–434.

[44] Y. Wu, V. Boominathan, H. Chen, A. Sankaranarayanan, and A. Veeraraghavan, “Phasecam3d—learning phase masks for passive single view depth estimation,” in *Proceedings of International Conference on Computational Photography (ICCP)*. IEEE, 2019, pp. 1–12.

[45] S. Nie, L. Gu, Y. Zheng, A. Lam, N. Ono, and I. Sato, “Deeply learned filter response functions for hyperspectral reconstruction,” in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4767–4776.

[46] Iliadis, M. and Spinoulas, L. and Katsaggelos, A. K., “Deepbinarymask: Learning a binary mask for video compressive sensing,” *arXiv preprint arXiv:1607.03343*, 2016.

[47] Y. Li, M. Qi, R. Gulve, M. Wei, R. Genov, K. N. Kutulakos, and W. Heidrich, “End-to-end video compressive sensing using anderson-accelerated unrolled networks,” in *Proceedings of International Conference on Computational Photography (ICCP)*. IEEE, 2020, pp. 137–148.

[48] R. Raskar, A. Agrawal, and J. Tumblin, “Coded exposure photography: motion deblurring using fluttered shutter,” in *ACM transactions on graphics*, vol. 25, no. 3, 2006.

[49] J. Holloway, A. C. Sankaranarayanan, A. Veeraraghavan, and S. Tambe, “Flutter Shutter Video Camera for Compressive Sensing of Videos,” in *Proceedings of International Conference on Computational Photography (ICCP)*. IEEE, 2012, pp. 1–9.

[50] A. Veeraraghavan, D. Reddy, and R. Raskar, “Coded Strobing Photography: Compressive Sensing of High Speed Periodic Videos,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 671–686, 2011.

[51] J. Gu, Y. Hitomi, T. Mitsunaga, and S. Nayar, “Coded Rolling Shutter Photography: Flexible Space-Time Sampling,” in *Proceedings of International Conference on Computational Photography (ICCP)*, 2010, pp. 1–8.

[52] G. Bub, M. Tecza, M. Helmes, P. Lee, and P. Kohl, “Temporal Pixel Multiplexing for Simultaneous High-Speed, High-Resolution Imaging,” *Nature Methods*, vol. 7, 2010.

[53] M. Gupta, A. Agrawal, and A. Veeraraghavan, “Flexible Voxels for Motion-Aware Videography,” in *Proceedings of European Conference on Computer Vision (ECCV)*, vol. 3, 2010, p. 6.

[54] D. Reddy, A. Veeraraghavan, and R. Chellappa, “P2C2: Programmable Pixel Compressive Camera for High Speed Imaging,” in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 329–336.

[55] Hamamatsu Photonics K.K., “Imaging device,” *Japan patent JP2015-216594A*, 2015-12-03.

[56] M. O’Toole, J. Mather, and K. N. Kutulakos, “3d shape and indirect appearance by structured light transport,” in *CVPR*, 2014.

[57] Y. Pati, R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” in *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, 1993, pp. 40–44 vol.1.

[58] E. Van Den Berg and M. P. Friedlander, “Probing the pareto frontier for basis pursuit solutions,” *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[59] G. Bub, M. Tecza, M. Helmes, P. Lee, and P. Kohl, “Temporal pixel multiplexing for simultaneous high-speed, high-resolution imaging,” *Nature methods*, vol. 7, no. 3, pp. 209–211, 2010.

[60] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan, “Flexible voxels for motion-aware videography,” in *Proceedings of European Conference on Computer Vision (ECCV)*, 2010, pp. 100–114.

[61] S. Anwar, S. Khan, and N. Barnes, “A deep journey into super-resolution: A survey,” *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–34, 2020.

[62] W. Shen, M. Cheng, G. Lu, G. Zhai, L. Chen, M. S. Asif, and Z. Gao, “Spatial temporal video enhancement using alternating exposures,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 8, pp. 4912–4926, 2022.

[63] X. Yuan, Y. Liu, J. Suo, and Q. Dai, “Plug-and-play algorithms for large-scale snapshot compressive imaging,” in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 1447–1457.

[64] X. Han, B. Wu, Z. Shou, X.-Y. Liu, Y. Zhang, and L. Kong, "Tensor fista-net for real-time snapshot compressive imaging," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 10 933–10 940.

[65] Y. Li, M. Qi, R. Gulve, M. Wei, R. Genov, K. N. Kutulakos, and W. Heidrich, "End-to-end video compressive sensing using anderson-accelerated unrolled networks," in *Proceedings of International Conference on Computational Photography (ICCP)*. IEEE, 2020, pp. 1–12.

[66] M. Yoshida, A. Torii, M. Okutomi, K. Endo, Y. Sugiyama, R. Taniguchi, and H. Nagahara, "Joint optimization for compressive video sensing and reconstruction under hardware constraints," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2018, pp. 634–649.

[67] H. Sun, A. V. Dalca, and K. L. Bouman, "Learning a probabilistic strategy for computational imaging sensor selection," in *Proceedings of International Conference on Computational Photography (ICCP)*. IEEE, 2020, pp. 81–92.

[68] M. Courbariaux, Y. Bengio, and J.-P. David, "Binaryconnect: Training deep neural networks with binary weights during propagations," *Advances in neural information processing systems*, vol. 28, 2015.

[69] J. N. Martel, L. K. Mueller, S. J. Carey, P. Dudek, and G. Wetzstein, "Neural sensors: Learning pixel exposures for hdr imaging and video compressive sensing with programmable sensors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 7, pp. 1642–1653, 2020.

[70] A. N. Gomez, M. Ren, R. Urtasun, and R. B. Grosse, “The reversible residual network: Backpropagation without storing activations,” *Advances in neural information processing systems*, vol. 30, 2017.

[71] J. Pont-Tuset, F. Perazzi, S. Caelles, P. Arbeláez, A. Sorkine-Hornung, and L. Van Gool, “The 2017 davis challenge on video object segmentation,” *arXiv:1704.00675*, 2017.