



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Per-pixel mirror-based measuring: Um novo método para aquisição para vídeo de alta velocidade

Jonathan Alis Salgado Lima

Dissertação apresentada como requisito parcial
para conclusão do Mestrado em Informática

Orientadora

Prof.^a Dr.^a Mylène C. Q. Farias

Coorientador

Prof. Dr. Cristiano Miosso

Brasília
2014

Universidade de Brasília — UnB
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Mestrado em Informática

Coordenadora: Prof.^a Dr.^a Alba Cristina M. de Melo

Banca examinadora composta por:

Prof.^a Dr.^a Mylène C. Q. Farias (Orientadora) — CIC/UnB

Prof. Dr. Bruno Luigi Macchiavello — CIC/UnB

Prof. Dr. Ricardo von Borries — University of Texas at El Paso (UTEP)

CIP — Catalogação Internacional na Publicação

Lima, Jonathan Alis Salgado.

Per-pixel mirror-based measuring: Um novo método para aquisição para vídeo de alta velocidade / Jonathan Alis Salgado Lima. Brasília : UnB, 2014.

109 p. : il. ; 29,5 cm.

Dissertação (Mestrado) — Universidade de Brasília, Brasília, 2014.

1. *compressive sensing*, 2. amostragem de vídeo, 3. eficiência luminosa, 4. independência temporal

CDU 004

Endereço: Universidade de Brasília
Campus Universitário Darcy Ribeiro — Asa Norte
CEP 70910-900
Brasília-DF — Brasil



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Per-pixel mirror-based measuring: Um novo método para aquisição para vídeo de alta velocidade

Jonathan Alis Salgado Lima

Dissertação apresentada como requisito parcial
para conclusão do Mestrado em Informática

Prof.^a Dr.^a Mylène C. Q. Farias (Orientadora)
CIC/UnB

Prof. Dr. Bruno Luigi Macchiavello
CIC/UnB

Prof. Dr. Ricardo von Borries
University of Texas at El Paso (UTEP)

Prof.^a Dr.^a Alba Cristina M. de Melo
Coordenadora do Mestrado em Informática

Brasília, 09 de Agosto de 2014

Agradecimentos

Primeiramente, agradeço a minha professora e orientadora Mylène, que acreditou em mim, me incentivou e convidou a fazer o mestrado. Tal incentivo foi fundamental para a minha perseverança, motivação e meu sucesso. Agradeço ao meu co-orientador Cristiano, que pulou de para-quedas no meio do meu mestrado e me ajudou muito numa área que nem eu nem Mylène éramos especialistas. Agradeço ao tempo, dedicação e paciência que eles dedicaram a mim. Foi de extrema importância para a conclusão deste trabalho.

Agradeço a Chengbo Li, pela atenção e pela disponibilidade de fornecer os códigos de minimização TV3D do TVAL3 que foram utilizados neste trabalho.

Agradeço aos meus amigos Amanda, Ariane, Gabriela, Helard, Lucas e Paula. Com eles convivi vários dias, tivemos aulas juntos, muito estudo, viagens e aventuras. Esse meu mestrado não seria a mesma coisa sem eles. Aprendi muito com eles, tanto de cunho acadêmico, como coisas da vida. Definitivamente, ocuparam um espaço no meu coração.

Agradeço aos meus pais José Alis e Clara que amo muito, por acreditar em mim e pelo apoio. Agradeço por ainda deixarem eu morar com eles, por não exigirem que eu lave louça, por deixar eu levar esse estilo de vida peculiar. Sem isso acho que eu não conseguiria concluir este trabalho.

Agradeço à secretaria do Programa de Pós-Graduação em Informática (PPGInf) (em especial a Paula e a professora Alba) por todo o apoio com as burocracias necessárias, agradeço também ao órgão de Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Resumo

Vídeos adquiridos em alta resolução temporal têm aplicações em vários segmentos como ciência, indústria, segurança, entretenimento, esportes, etc. Porém, as câmeras que adquirem vídeos em alta resolução temporal são muito caras e, portanto, há uma necessidade de desenvolvimento de novas formas de aquisição. Novas arquiteturas de câmeras computacionais têm utilizado aquisição em câmeras comuns de baixa resolução temporal e a teoria de *compressive sensing* para reconstruir vídeos de maior resolução temporal. Porém, as técnicas de aquisição presentes na literatura apresentam dependência temporal entre as medidas e descarte de informação luminosa. Neste trabalho, apresentamos um novo método para a aquisição de medidas lineares denominado *Per-pixel mirror-based measuring* (PPM). O método proposto apresenta medidas temporalmente independentes e aproveita 100% da informação luminosa. Nossa metodologia de testes consiste em simular os modos de aquisição presentes na literatura e o método proposto, adquirir amostras de um conjunto de vídeos (sintéticos e reais), reconstruir utilizando o mesmo algoritmo de reconstrução e, finalmente, avaliar os resultados utilizando uma métrica de qualidade. Nas simulações, avaliamos os resultados testando variações de diversos parâmetros, como resolução espacial do vídeo, resolução temporal, quantidade de movimento, número de dimensões da reconstrução, etc. Os resultados mostram que a técnica proposta apresenta melhor desempenho quanto a qualidade dos vídeos reconstruídos do que as outras técnicas testadas. Trabalhos futuros incluem testar novas configurações da técnica, testar novos métodos de reconstrução e avaliar a construção de um hardware que possa viabilizar a técnica em câmeras.

Palavras-chave: *compressive sensing*, amostragem de vídeo, eficiência luminosa, independência temporal

Abstract

Video acquisition in high temporal resolution has several applications in science, industry, security, entertainment, sports, etc. However, high speed cameras are expensive and, therefore, there is a great need for new acquisition methods. New computational cameras architecture have been using acquisition using common low speed cameras and compressive sensing theory for reconstructing high speed videos. But acquisition techniques present in literature show time dependence among the measures and throw away light information. This work presents a new linear acquisition method named Per-Pixel mirror-based measuring (PPM). The proposed method presents temporal independent measures and keeps 100% of light information. Our test methodology consists of simulating the acquisition methods present in literature and proposed method, acquiring measures from a set of videos (synthetic and natural), reconstructing the videos with the same algorithm and, finally, comparing the results using a quality metric. On simulations, we evaluate the results testing variations considering temporal and space resolutions of the videos, amount of video motion, number of reconstruction dimensions, etc. The results show that PPM presents better performance than other techniques with respect to the quality of the reconstructed videos. Future works include testing new configurations of the technique, testing new reconstruction methods and evaluating hardware construction for bringing the technique to video cameras.

Keywords: compressive sensing, video sampling, light efficiency, temporal independence

Sumário

1	Introdução	1
1.1	Motivação	1
1.1.1	Viabilidade de Câmeras de Alta Velocidade	3
1.2	Métodos e Problemas Relacionados à Obtenção de Vídeo de Alta Velocidade Usando Sensores Comuns	5
1.3	Hipótese e Objetivos	7
1.4	Justificativa	8
1.5	Estrutura da Dissertação	9
2	Fundamentos e Conceitos	10
2.1	Sinais em Domínio Discreto e Representações por Transformadas Lineares .	10
2.1.1	Representações por Transformadas Lineares	11
2.1.2	Notação Matricial das Transformadas	12
2.2	<i>Compressive Sensing</i>	13
2.2.1	Sinais Esparsos e Aproximadamente Esparsos	13
2.2.2	Aquisição por Medidas Lineares	14
2.2.3	O Problema da Reconstrução	15
2.2.4	Projeto da Matriz de Medidas	15
2.2.5	Projeto do Algoritmo de Reconstrução	17
2.2.6	<i>Compressive Sensing</i> para Sinais Aproximadamente Esparsos	20
2.2.7	Aquisição por Medidas Lineares Ruidosas	21
2.2.8	Reconstrução a Partir de Medidas Ruidosas	21
2.3	Aquisição de Vídeo Digital	21
2.3.1	Tempo de Integração e de Exposição	22
2.3.2	A Escolha do Tempo de Exposição	23
2.3.3	Moduladores de Luz	25
2.4	Modos de Aquisição de Medidas Lineares de Vídeo	27
2.4.1	Forma Tradicional de Aquisição de Cenas	27
2.4.2	Método de Aquisição por Modulador Global	28

2.4.3	Método de Aquisição por Modulador por Pixel	30
3	Método de Aquisição <i>Per-Pixel Mirror-Based Measuring</i>	34
3.1	<i>Per-Pixel Mirror-Based Measuring</i>	34
3.2	Comparação Entre os Métodos de Aquisição	38
3.2.1	Propriedades Gerais	38
3.2.2	Incoerência	41
3.3	Outras Variações do PPM	42
3.3.1	PPM em Blocos	42
3.3.2	PPM para Super-resolução	44
3.4	Modelo de Reconstrução dos Vídeos	46
3.5	Implementação	47
3.5.1	O algoritmo TVAL3	47
3.5.2	A função que Simula o HPDMD e a Matriz de Índices	48
3.5.3	Implementação dos Métodos de Amostragem com a HPDMD para Vídeos	52
4	Simulação e Testes	55
4.1	Testes em Imagens	55
4.2	Testes em Vídeos Sintetizados	57
4.2.1	Reconstrução 2D: Colunas dos Sub-quadros pelo Tempo	59
4.2.2	Reconstrução 3D	61
4.2.3	Robustez ao Ruído	65
4.2.4	PPM2 Dividida em Blocos	66
4.3	Vídeos Naturais	68
4.4	Análise dos Tempos de Execução	80
4.5	Discussão dos Resultados	85
5	Conclusão e Trabalhos Futuros	87
5.1	Conclusão	87
5.2	Trabalhos Futuros	89
5.2.1	Outras Variações do PPM	89
5.2.2	Reconstrução	90
5.2.3	Implementação em Hardware	91
	Referências	92

Lista de Figuras

1.1	Um conjunto de quadros obtidos a partir de um vídeo adquirido a 30.000 quadros por segundo. Neste exemplo, a captura do momento exato da quebra do bico da furadeira ilustra a importância de uma alta resolução temporal para este tipo de aplicação. Fonte: [45].	2
1.2	Diversas aplicações de vídeo de alta velocidade.	3
1.3	Exemplo de um processo computacional de formação de imagens para uma aplicação em visualização around the corner.	4
2.1	Exemplo de um sinal em representação esparsa.	14
2.2	Exemplo de parte do processo da aplicação das funções PSF e TPSF . . .	17
2.3	Minimização da ℓ_2 para um domínio de tamanho 2.	18
2.4	Minimização da ℓ_1 para um domínio de tamanho 2.	19
2.5	a) Um sensor e seus componentes e b) um exemplo de uma disposição de vários sensores em forma matricial utilizados para aquisição de imagem [21].	22
2.6	Sensores para aquisição de imagens coloridas.	23
2.7	Influência do indicador de sensibilidade ISO no ruído obtido pelos sensores.	24
2.8	Borrão de movimento (<i>Motion Blur</i>) devido ao alto tempo de exposição. .	24
2.9	Exposição por modulador global. No gráfico de cima, vemos o esquema tradicional de exposição, ligamos e desligamos a exposição apenas uma vez durante o tempo de integração (linha tracejada). Com o modulador global, podemos ligar e desligar a exposição quando desejarmos durante o tempo de integração.	25
2.10	Funcionamento do dispositivo de espelhos DMD e sua aplicação na técnica <i>Single Pixel Camera</i> (SPC).	26
2.11	Método de aquisição por modulador global	29
2.12	Sistema óptico externo que implementa o método <i>per-pixel shutter</i> (PPS), adaptado de [18].	30
2.13	Aquisição pela primeira configuração do método <i>per-pixel shutter</i> (PPS1). .	32
2.14	Aquisição utilizando a segunda configuração do modulação por pixel (PPS2).	33

3.1	Método de aquisição proposto em sua primeira (PPM1)	36
3.2	Método de aquisição PPM, na configuração PPM2	37
3.3	Comparativo de intensidade de luz para cada sub-quadro de todas os métodos	39
3.4	Resultado das funções TPSF e PSF para um sinal de domínio 1D de di- mensões 1×100 , e taxa de subamostragem de $4\times$	42
3.5	Resultado das funções TPSF e PSF para um sinal 2D de dimensões 32×32 , e taxa de subamostragem de $4\times$	43
3.6	Quadros gerados pela pré-visualização a partir das medidas PPM por blocos 5×5	44
3.7	Método de aquisição PPM para super-resolução para uma câmera com $M = 4$ pixels, e um HPDMD com $N = 20$ espelhos.	45
4.1	Resultado da reconstrução para uma imagem de 64 por 64.	56
4.2	Resultado da reconstrução para uma imagem de 256 por 256	57
4.3	Resultado de reconstrução para uma imagem de 1024 por 1024	58
4.4	Resultado em um trecho de 64 por 128 da imagem de 1024 por 1024	59
4.5	Um dos quadros da reconstrução do vídeo do fantoma de tamanho 200×200 $\times 128$, utilizando o PPM2 a uma taxa de sub-amostragem de $16x$. a) o quadro original. b) o quadro reconstruído: SER = 17,8dB.	60
4.6	Gráficos para o teste em vídeos sintetizados 2D, das linhas pelo tempo.	61
4.7	SER das reconstruções utilizando a TV3D do vídeo do vídeo sintetizado do fantoma, para taxas de 2, 4, 8 e 16. Em cada gráfico, o 1 se refere ao método FS, o 2 se refere ao PPS1, o 3 ao PPS2, o 4 se refere ao PPM1 e o 5 se refere ao PPM2.	63
4.8	Teste com os métodos FS, PPS1, PPS2, PPM1 e PPM2 a partir de um vídeo de tamanho $200 \times 200 \times 128$, a uma taxa de sub-amostragem de $16x$	64
4.9	SER das reconstruções ao adicionar ruído às medidas	67
4.10	SER das reconstruções a partir do método PPM2 por blocos.	68
4.11	O primeiro e último quadro de cada um dos 12 vídeos escolhidos para os testes.	71
4.12	SER da reconstrução dos 12 vídeos.	72
4.13	Reconstruções do segundo e terceiro quadros do vídeo 4 a partir das diversos métodos, a uma taxa de $16x$ (6,25% das amostras).	74
4.14	Reconstruções do primeiro e último quadros do vídeo 7 a partir das diversos métodos, a uma taxa de $16x$ (6,25% das amostras).	75
4.15	SER ordenados pela métrica de movimento para taxas de $4x$, $8x$ e $16x$. Em azul é o PPS2, em vermelho é o PPM2, em verde o PPM-2D, e o preto corresponde ao PPM3.	77

4.16	Resultados para os métodos PPS2, PPM2, PPM-2D e PPM3 para o vídeo 9, com taxa de sub-amostragem de 16x.	78
4.17	Resultados para os métodos PPS2, PPM2, PPM-2D e PPM3 para o vídeo 8, com taxa de sub-amostragem de 4x.	79
4.18	Tempos de reconstrução dos testes de reconstrução 2D das colunas pelo tempo.	81
4.19	Tempos de reconstrução dos testes de reconstrução 3D para vídeos sintetizados.	81
4.20	Tempos de reconstrução dos testes de reconstrução 3D com ruído nas medidas.	82
4.21	Tempos de reconstrução dos testes de reconstrução do PPM2 ao variar o tamanho do bloco.	83
4.22	Tempos de reconstrução dos testes de reconstrução 3D para vídeos naturais para os métodos PPS2, PPM2, PPM3 e PPM-2D.	84

Lista de Tabelas

3.1	Métodos de aquisição e suas vantagens e desvantagens.	38
3.2	Percentagem de informação luminosa adquirida em cada método e quantidade de ruído adquirido ao alterar o ISO para obter a mesma quantidade de luz que o método tradicional.	40
3.3	Métodos de aquisição e reconstrução compatíveis.	47
4.1	Métrica SER para cada método e cada imagem.	56
4.2	SER (dB) dos vídeos reconstruídos em relação ao vídeo original.	60
4.3	SER (dB) médios dos vídeos reconstruídos para os métodos de aquisição testados, para a reconstrução TV2D das colunas pelo tempo.	62
4.4	Os quadros de início e final e os saltos dos vídeos, note que cada um contém 16 quadros.	70
4.5	SER (dB) médio dos vídeos naturais reconstruídos.	72
4.6	Medida de movimento utilizada para medir a atividade temporal dos vídeos dada por 4.1. Quanto maior o valor obtido pela métrica, maior o movimento do vídeo.	76
4.7	Average SER (dB): reconstructed natural videos.	84

Lista de Símbolos e Abreviações

<i>SPC</i>	–	<i>Single Pixel Camera</i>
<i>FS</i>	–	<i>Flutter Shutter</i>
<i>PPS</i>	–	<i>Per-pixel Shutter</i>
<i>PPM</i>	–	<i>Per-pixel Mirror-based Measuring</i>
<i>DMD</i>	–	<i>Digital Micromirror Device</i>
<i>HPDMD</i>	–	<i>High Precision Digital Micromirror Device</i>
<i>TV</i>	–	<i>Total Variation</i>
<i>SER</i>	–	<i>Signal-to-Error Ratio</i>
<i>TPSF</i>	–	<i>Transform Point Spread Function</i>
<i>PSF</i>	–	<i>Point Spread Function</i>

Capítulo 1

Introdução

Este trabalho apresenta um novo método para aquisição de vídeo de alta resolução temporal a partir de uma câmera com sensores comuns de baixa resolução temporal denominado *Per-Pixel Mirror-Based Measuring*. O método proposto necessita de um novo dispositivo a ser introduzido em uma câmera comum para adquirir as amostras, e a partir delas, o vídeo de alta resolução temporal é obtido utilizando *compressive sensing*. Avaliamos o método proposto, comparando os seus resultados de reconstrução com os resultados obtidos a partir de outros métodos disponíveis na literatura. A avaliação é feita em termos de qualidade objetiva e subjetiva e de tempo de reconstrução.

Este capítulo apresenta a relevância das aplicações da aquisição de vídeo de alta resolução temporal e as vantagens de utilizar câmeras com sensores comuns para adquirir estes sinais. Descrevemos ainda, um conjunto de métodos presentes na literatura, detalhando suas vantagens e desvantagens; esta análise permite observar quais aspectos devem ser atacados por um novo sistema de aquisição de vídeo em alta resolução temporal, como o proposto nesta pesquisa. Apresentamos também a proposta deste trabalho, com as hipóteses e objetivos considerados no seu desenvolvimento. Finalmente, descrevemos a organização desta dissertação.

1.1 Motivação

A aquisição de vídeos a alta resolução temporal é muito requisitada em várias aplicações na indústria. Uma destas aplicações consiste em analisar colisões controladas de carros com o objetivo de verificar a sua segurança. Assim, pesquisadores podem analisar as imagens e entender melhor a interação entre as partes do carro e os ocupantes e determinar quais aspectos do carro precisam ser aperfeiçoados para uma maior segurança. A mesma ideia é aplicada em diversos tipos de teste, com o objetivo de medir a resistência de produtos a quedas, impactos, tensões, bem como inferir ponto de fusão, etc. A Figura 1.1

apresenta um conjunto de imagens que ilustra o momento em que uma broca de furadeira se quebra. Para ser possível observar tal efeito, é necessária uma resolução temporal de 30.000 quadros por segundo [45].

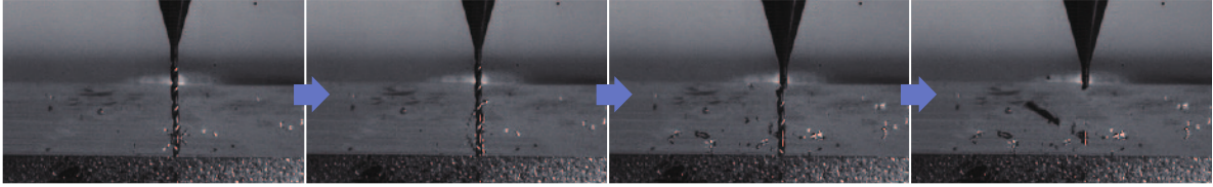


Figura 1.1. Um conjunto de quadros obtidos a partir de um vídeo adquirido a 30.000 quadros por segundo. Neste exemplo, a captura do momento exato da quebra do bico da furadeira ilustra a importância de uma alta resolução temporal para este tipo de aplicação. Fonte: [45].

Assim como existem aplicações na indústria, existem também na academia. Neste caso, as imagens captadas são utilizadas para entender o funcionamento de fenômenos específicos, representando uma importante ferramenta para pesquisas científicas. Entre os exemplos das aplicações de vídeo de alta velocidade na ciência, podemos citar a análise biomecânica, onde é possível visualizar melhor o comportamento de tecidos, ossos e músculos humanos nas condições de movimento ou estresse, conforme ilustrado na Figura 1.2(a). Existem ainda aplicações em microscopia, com o objetivo de analisar o movimento de microrganismos e suas funções celulares, como ilustrado na Figura 1.2(b). Outra aplicação é para visualização do fluxo de fluidos de densidade variável, o que pode fornecer um melhor entendimento de seu comportamento sobre diversas condições (*Schlieren photography*, um processo ilustrado na Figura 1.2(c) [13]).

Também existem aplicações militares, nas quais se pode observar como as minas e explosões afetam os materiais militares ou analisar como os projéteis atravessam o ar e os materiais em geral. Existem aplicações até mesmo na área de entretenimento. Na Figura 1.2(d) podemos observar flechas atravessando balões. Observamos também aplicações no esporte, onde a resolução temporal pode ajudar a observar com clareza o que aconteceu com os jogadores em lances duvidosos, ou observar se a bola passou ou não das linhas do campo. Todas essas aplicações exigem uma resolução temporal de até milhares de quadros por segundo, o que não é possível a olho nu ou com câmeras convencionais.

O número de aplicações para vídeos de alta resolução temporal aumenta quando a taxa de quadros por segundo atinge a ordem de trilhões. A essa velocidade de aquisição, o tempo entre quadros é tão pequeno que a luz viaja apenas décimos de milímetros e, assim, é possível observar fótons. Com isso, podemos observar a luz e como ela se espalha nos materiais, incluindo diversos efeitos óticos. Isso permite, por exemplo, a visualização de cenas e até mesmo a sua modelagem 3D por volta de uma parede. No esquema proposto por Velten *et al.* [54], um laser (emissor de fótons) emite uma luz e essa luz é refletida

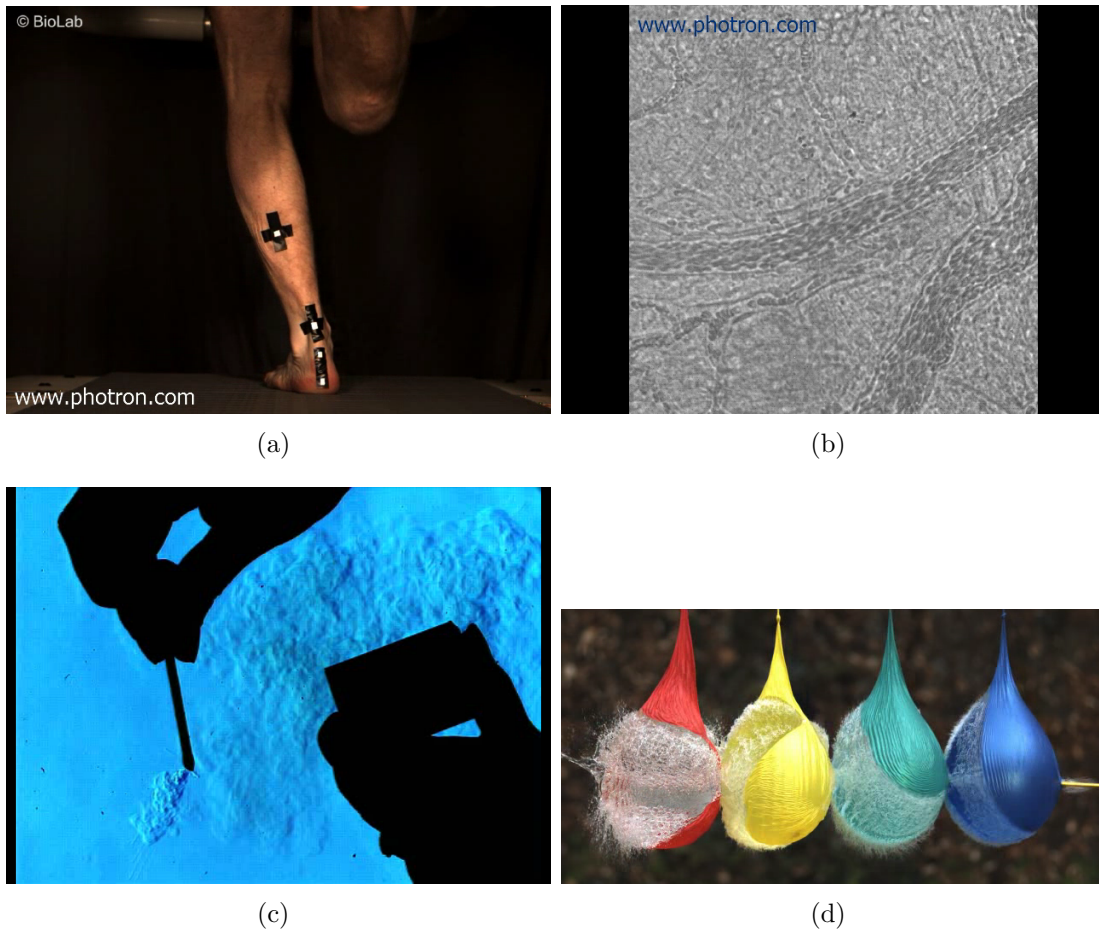


Figura 1.2. Diversas aplicações de vídeo de alta velocidade. (a) Aplicação em biomecânica, neste caso as imagens a 2000 FPS ajudam a entender o comportamento dos músculos na atividade de corrida. (b) Aplicação em microbiologia, neste caso as imagens a 500 FPS ajudam a ilustrar como é o fluxo de fluidos internos nas células. (c) Visualização do fluxo dos gases gerados no processo de se acender um fósforo. (d) As imagens a 2000 FPS ajudam a observar melhor como balões estouram, mostrando que o material dos balões perdem sua integridade estrutural muito rápido, e a água parece flutuar por algum tempo. Fonte: [45]

num anteparo e no objeto a ser observado. Alguns dos fótons voltam à câmera. Medindo o tempo entre fótons e a intensidade, é possível estimar o formato dos objetos que estão por trás de uma parede. O esquema é representado pela Figura 1.3.

1.1.1 Viabilidade de Câmeras de Alta Velocidade

Câmeras de vídeo estão demandando cada vez mais resolução espacial e temporal. Nos últimos anos, a tecnologia dos sensores nas câmeras de vídeo tem aumentado consideravelmente a resolução espacial (número de pixels) das câmeras. As câmeras de vídeo atuais chegam a resoluções superiores a 40 Megapixels (milhões de pixels) [12]. Até mesmo ce-

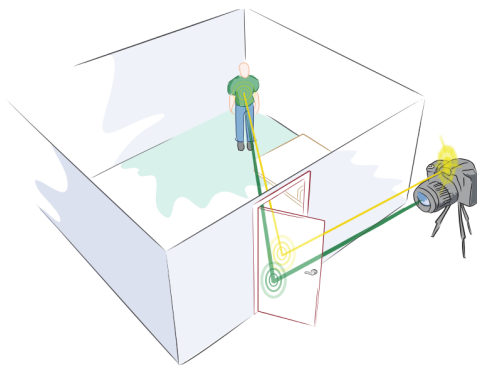


Figura 1.3. Exemplo de um processo computacional de formação de imagens para uma aplicação em visualização por trás da parede. Nesta aplicação é possível obter uma modelagem 3D dos objetos atrás de uma parede a partir da reflexão de fótons pela porta. A câmera requer uma taxa de trilhões de quadros por segundo, permitindo a análise da diferença de tempo entre as detecções de fótons para estimar as posições dos objetos correspondentes. Fonte: [54]

lulares apresentam câmeras com resolução entre 8 e 13 Megapixels [44], alguns chegando a 41 Megapixels [41]. Porém, considerando câmeras para aquisição de vídeo, a resolução temporal, ou seja, o número de quadros por segundo (*frames per second* - FPS) não tem aumentado na mesma proporção que a resolução espacial. As câmeras comuns têm cerca de 30 a 60 quadros por segundo [47]. Portanto, não é possível obter os vídeos de alta resolução temporal a partir de câmeras convencionais.

A forma mais intuitiva para adquirir um vídeo em alta resolução temporal é utilizando câmeras de vídeo de alta velocidade (*high-speed video cameras*) [45]. Estas câmeras utilizam sensores sensíveis e rápidos, podendo fazer a aquisição de vídeo em alta resolução temporal (com soluções comerciais podendo chegar a dezenas de milhares de FPS) sob certas condições de luminosidade. Infelizmente, atualmente essas câmeras ainda são muito caras.

Outra forma de adquirir vídeo de alta velocidade é utilizando um conjunto de diversas câmeras alinhadas para adquirir as imagens a vários quadros por segundo [57, 58]. Porém, esta também é uma alternativa cara, uma vez que exige dezenas de câmeras muito bem sincronizadas.

Essas alternativas convencionais se mostram inviáveis em muitas aplicações, por aspectos econômicos e de praticidade. Sendo assim, baratear o processo de aquisição de vídeo de alta velocidade é um passo importante na popularização da captura de vídeo em alta resolução temporal.

1.2 Métodos e Problemas Relacionados à Obtenção de Vídeo de Alta Velocidade Usando Sensores Comuns

Os métodos convencionais de aquisição de vídeo em alta resolução temporal são geralmente limitados pelo critério de Nyquist-Shannon, o qual estabelece uma taxa de aquisição de medidas suficiente para se obter o sinal adequadamente. Embora este critério estabeleça um número suficiente de amostras a serem adquiridas para se obter o sinal, geralmente é possível obter o sinal mesmo com um número menor de medidas. Porém, os métodos convencionais de aquisição de vídeo em alta resolução temporal utilizam o número máximo de medidas dado pelo critério de Nyquist-Shannon, seja por câmeras de alta velocidade (em que os sensores se atualizam muito rápido e muitas medidas são tomadas), seja por vetor de câmeras (em que o número de medidas é multiplicado pelo número de câmeras).

Novos métodos, baseados em avanços nas áreas de câmeras computacionais e *compressive sensing*, têm conseguido reconstruir vídeos em alta resolução temporal a partir de um número de medidas menor que a exigida pela taxa de Nyquist [47, 23, 26, 25]. Esses métodos utilizam uma aquisição de um número bem menor de medidas (lineares) do que a resolução do sinal que se deseja obter, sendo o vídeo completo reconstruído posteriormente por um processo de otimização numérica. Este tipo de método de reconstrução se baseia no fato de que as cenas naturais são aproximadamente esparsas em algum domínio conhecido e, a partir de algoritmos de otimização que exploram esta esparsidade, é possível obter um sinal muito aproximado do sinal desejado. Desta forma, a resolução temporal não fica limitada à taxa de quadros da câmera.

Vários trabalhos baseados em *compressive sensing* em vídeos e câmeras computacionais já foram realizados para obter vídeos de alta resolução temporal sem utilizar sensores de alta velocidade. Alguns desses trabalhos incluem aquisição por um único pixel, ou aquisição por um modelador de luz global, ou aquisição por um modelador de luz por pixel. Entre esses trabalhos, alguns se limitam a tipos específicos de cenas, tais como cenas com tipos específicos de movimento (movimento linear [24, 36, 1, 20], ou periódico [51]), enquanto outros trabalhos aceitam cenas arbitrárias, com qualquer tipo de movimento [16, 55, 47, 25, 26, 23].

Takar *et al.* propuseram um esquema de aquisição e recuperação de imagens conhecido como *single pixel camera* [52, 16], que utiliza com apenas um sensor para a aquisição e tem sua reconstrução baseada em métodos de *compressive sensing*. Nesse esquema, a luz de um padrão aleatório de pixels é escolhida para atingir um único sensor. A cada atualização do sensor, o padrão é alterado para outro padrão aleatório. A partir de poucas medidas

fornecidas pelo sensor, em uma quantidade muito menor que o número de amostras da imagem, uma nova imagem é reconstruída utilizando técnicas de otimização.

Para vídeos, o processo de aquisição é semelhante. Entretanto, a cena é exposta pelo tempo desejado, alterando-se o padrão continuamente. Então, o vídeo é reconstruído usando uma otimização que explora a redundância espaço-temporal do vídeo [16, 55]. Como existem sensores únicos que operam em frequências muito altas, a aquisição de vídeo a alta velocidade é possível. Chan *et al.* utilizaram um transmissor e receptor de luz operando em frequências na ordem de Terahertz, e puderam recuperar radiações em cenas a frequência de 100 Gigahertz [9]. Sankaranarayanan *et al.* se basearam na *single pixel camera* para adquirir e reconstruir vídeos [51]. Apesar de conseguir uma reconstrução com poucas medidas, a técnica é limitada a vídeos que podem ser modelados como sistemas dinâmicos lineares.

Apesar da técnica *single pixel camera* mostrar resultados razoáveis com apenas um pixel, é possível utilizar uma câmera de vídeo com uma grande quantidade de sensores para adquirir cenas e, assim, obter uma qualidade mais satisfatória. Um modelador de luz global seleciona o tempo em que os sensores da câmera irão receber a luz. Veeraraghavan *et al.* [53] utilizam um modulador de luz global em sua técnica para a aquisição de vídeos periódicos de alta velocidade. Porém, apesar de obter ótimos resultados, o uso de um modelo para o movimento da cena limita a variedade de cenas que podem ser capturadas. Ramesh *et al.* [1] utilizaram a aquisição por modulador de luz global para eliminar o efeito de borrão (*motion blur*) de cenas com movimento rápido, porém com a limitação de que a cena deve ser representada como um sistema dinâmico linear. Também utilizando um sistema de aquisição por modelador de luz global, Holloway *et al.* propuseram uma técnica de câmera computacional sem restrição com relação ao tipo de movimento da cena [26]. Eles reconstróem a cena tanto por otimização da esparsidade de diferenças finitas dos vídeos reais, como por treinamento de dicionário. Embora não tenham obtido resultados melhores que outras técnicas disponíveis na literatura, a sua viabilidade é superior pois os modeladores de luz globais já estão disponíveis em certas câmeras comerciais.

Uma outra forma de aquisição possível consiste em utilizar um modelador de luz por pixel. Esse modulador seleciona o tempo de exposição à luz isoladamente para cada sensor. Tal método ainda não está presente em câmeras comerciais, sendo que as implementações existentes utilizam um dispositivo externo acoplado ao sistema ótico das câmeras. Recentemente, diversos trabalhos têm utilizado a aquisição dessa maneira em seus algoritmos para resolver o problema de reconstrução por *compressive sensing* em vídeos. Gupta *et al.* utilizaram esse tipo de aquisição em sua técnica para recuperação do volume espaço-temporal de um vídeo [23]. Nessa técnica, é possível escolher a resolução espaço-temporal após a captura apenas recombinação das medidas. O restante da cena é preenchida por

interpolação. Com um algoritmo que detecta movimento é possível reconstruir as partes do vídeo que têm maior movimento com melhor resolução temporal, e as partes mais contantes com uma maior resolução espacial. A técnica, porém, se baseia na forma do modelador por pixel que conta com uma entrada de luz bastante reduzida, o que limita a captação a cenários iluminados intensamente (que gerariam vídeos bastante ruidosos).

Também adquirindo a cena utilizando um modulador por pixel, porém com um fluxo maior de luz, Reddy *et al.* utiliza um algoritmo de reconstrução que se baseia em minimizar o fluxo ótico [47]. Ele obtém bons resultados para cenas que poderiam ser facilmente comprimidas. Porém, os resultados obtidos não são muito bons em cenas com muitas transições (como por exemplo, um balão estourando). A partir de um processo de aquisição similar, Hitomi *et al.* propuseram algoritmo de reconstrução baseado na minimização da esparsidade da cena em uma base criada a partir de inúmeros pequenos trechos de um conjunto de vídeos pré-escolhidos, representando um dicionário pré-treinado [25]. A técnica obtém resultados similares aos da técnica de Reddy.

Como mostrado, muitas das técnicas de câmeras computacionais que utilizam os métodos atuais de aquisição apresentam problemas quanto à eficiência de absorção luminosa. Outras funcionam melhor em tipos específicos de cenas. Outro problema de alguns dos métodos atuais é a dependência temporal entre as amostras, isto causa dificuldades na separação apropriada do conteúdo de quadros subsequentes. Neste trabalho tentamos atacar esses problemas através da proposta de um novo método de aquisição.

1.3 Hipótese e Objetivos

Como vimos, diversas técnicas utilizam alguns métodos de aquisição linear de vídeos e reconstroem os vídeos em alta resolução temporal. Porém, os métodos não têm boa eficiência de luz e apresentam dependência temporal entre as medidas. Nós admitimos a hipótese de que um método de aquisição que tenha maior eficiência luminosa e apresente independência temporal entre as medidas pode levar a melhores resultados do que os métodos tradicionais. Considerando isso, nós traçamos os seguintes objetivos de pesquisa:

- Propor um método de aquisição viável fisicamente, temporalmente independente e com eficiência ótica igual ou superior a observada em outros métodos presentes na literatura, executar testes simulados para comparar o desempenho do método proposto com os outros métodos da literatura e avaliar os resultados.

Mais especificamente, vamos:

- Avaliar os métodos de amostragem de vídeo na literatura e observar o nível de eficiência de luz de cada um dos métodos, e o nível de dependência das amostras,

além de classificar outras vantagens e desvantagens dos métodos de amostragem na literatura e em comparação com o proposto;

- Implementar e avaliar o método proposto para reconstrução de vídeos de referência, para comparação;
- Verificar a viabilidade de implementar o método proposto em *hardware*;
- Na simulação, variar apenas o método de amostragem, sendo assim, os demais parâmetros devem se manter inalterados em cada teste. Parâmetros como o vídeo de teste, o tamanho dele, a taxa de ganho amostragem, e principalmente o mesmo método de reconstrução devem ser empregados nos testes;
- Testar parâmetros e observar a influência deles no resultado. A resolução espacial do vídeo, a resolução temporal, vídeos sintéticos e reais, o nível de movimentação do vídeo, robustez da amostragem ao ruído, reconstruções 2D ou 3D, entre outros;
- Verificar as vantagens e desvantagens do método, decidir se este método é viável ou não em relação aos outros métodos, e estabelecer as restrições de uso;
- Avaliar os recursos computacionais empregados nos testes, tais como tempo de execução e memória.

1.4 Justificativa

Se for possível propor um método de aquisição de vídeo que leve a melhores resultados de reconstrução do que os métodos na literatura, e se for possível implementar o hardware ou pelo menos avaliar a sua viabilidade, podemos cogitar incluir o dispositivo nas câmeras de vídeo. Desta forma, mesmo as câmeras normais poderão adquirir medidas para reconstrução de vídeo de alta velocidade. Com isso, os vídeos de alta velocidade estariam acessíveis para uma gama bem maior de usuários.

Mesmo que a implementação do dispositivo que implementa o método ainda não seja viável comercialmente para os padrões atuais de resoluções das câmeras de vídeo convencionais, esta pesquisa pode responder qual é a importância da independência temporal para a reconstrução de vídeo de alta velocidade e a influência da quantidade de informação luminosa captada. Esta resposta pode levar a outros métodos de aquisição, que podem vir a ser melhores do que o proposto. Sendo assim, o método proposto pode contribuir para novas formas de aquisição de medidas em *compressive sensing* em outras aplicações.

1.5 Estrutura da Dissertação

No Capítulo 2, apresentamos os conceitos necessários para o entendimento do trabalho. Entre eles os conceitos de sinais discretos, transformadas, *compressive sensing*, aquisição de vídeo digital e os métodos de amostragem de medidas lineares para reconstrução de vídeo de alta velocidade com *compressive sensing*. Para efeito de comparação, apresentamos o modelador de luz global e modeladores por pixel.

No Capítulo 3, apresentamos um novo esquema de aquisição de vídeo, denominado *Per-Pixel Mirror-Based Measuring*, bem como a comparação deste método com outros métodos considerando aspectos como a propriedade de incoerência, tipo de dependência entre as medidas, e eficiência luminosa, entre outros. Mostramos as variações possíveis do método. Apresentamos também o modelo de reconstrução para testes comparativos entre os métodos de aquisição, bem como os algoritmos de reconstrução utilizados e detalhes de implementação.

No Capítulo 4, apresentamos os resultados numéricos de nossas simulações dos diversos métodos de aquisição, com parâmetros referentes ao tipo de sinal testado (imagens, vídeos sintetizados, vídeos reais, resolução temporal e espacial), robustez ao ruído, dimensão do algoritmo de reconstrução (2D ou 3D). Avaliamos também os recursos computacionais e discutimos os resultados.

No Capítulo 5, analisamos quais os objetivos foram alcançados neste trabalho. Além disso, apresentamos os próximos passos do trabalho, gerados a partir de novas ideias que surgiram no decorrer do desenvolvimento.

Capítulo 2

Fundamentos e Conceitos

Neste capítulo, são apresentados conceitos básicos relacionados ao tema da pesquisa. Apresentamos os conceitos de sinais em domínios discretos e representações por transformadas lineares. Em seguida, apresentamos a teoria de *compressive sensing* que é considerada de grande interesse para a amostragem e reconstrução de vídeo em alta resolução temporal. Finalmente, apresentamos os conceitos básicos de aquisição de vídeo digital, para mostrar os principais métodos de aquisição de medidas lineares para reconstruir vídeos de alta velocidade presentes na literatura.

2.1 Sinais em Domínio Discreto e Representações por Transformadas Lineares

Um sinal em domínio discreto é uma sequência de números indexados por números inteiros. Por exemplo a sequência real

$$x[n], n \in \mathbb{Z}$$

é um sinal em domínio discreto, onde em cada posição n o sinal assume um valor real $x[n]$. Um sinal em domínio discreto pode ser de duração finita ou infinita. Ele tem duração finita se está definido em um intervalo finito

$$N_1 \leq n \leq N_2,$$

com $-\infty < N_1, N_2 < \infty$ e $N_2 > N_1$. O comprimento desse sinal é $N = N_2 - N_1 + 1$ [38].

Podemos ver um sinal real em domínio discreto iniciando em $N_1 = 0$ e de tamanho N como um vetor de \mathbb{R}^N , em que cada valor $x[n]$ representa um componentes desse vetor. Assim cada ponto de \mathbb{R}^N representa um único sinal.

Um sinal representado dessa forma é um sinal de domínio unidimensional, pois apenas um índice é necessário para a indexação. O sinal é de domínio bidimensional se sua intensidade varia em função de duas dimensões (*e.g.* $x[n, m]$ representa uma imagem). Um vídeo, por sua vez, é uma sequência de imagens, que pode ser vista como um sinal de domínio tridimensional $x[n, m, t]$.

2.1.1 Representações por Transformadas Lineares

Uma vez definido um sinal $x[n]$, $0 \leq n \leq N - 1$, de comprimento N , temos uma representação dele na forma temporal. Podemos representar esse sinal como uma combinação linear de impulsos deslocados $\delta[n - k]$ (sinal que tem valor igual a 1 na posição $n = k$, e zero nas demais posições), onde os coeficientes são as próprias amostras do sinal $x[n]$:

$$x[n] = \sum_{k=0}^{N-1} x[k] \delta[n - k].$$

Porém, esta é apenas uma forma de representar este sinal. Podemos também o representar como combinação linear de outros tipos de sinais diferentes de impulsos deslocados.

Uma forma geral de representar sinais em domínio discreto é utilizando transformadas discretas. Uma representação desta forma é vista a seguir:

$$X[k] = \sum_{n=0}^{N-1} x[n] \psi^*[k, n], \quad 0 \leq k \leq N - 1 \quad (2.1)$$

e

$$x[n] = \sum_{k=0}^{N-1} X[k] \psi[k, n], \quad 0 \leq n \leq N - 1, \quad (2.2)$$

onde $X[k]$ são os coeficientes da transformada de N pontos do sinal $x[n]$. (2.1) é conhecida como equação de análise e (2.2) é conhecida como equação de síntese. Nessas equações, $\psi[k, n]$ são sequências conhecidas como funções de base da transformada direta, de tamanho N cada. Assim, o sinal $x[n]$ pode ser representado por uma combinação linear de $\psi[n, k]$, ponderadas pelos coeficientes $X[k]$. $\psi^*[k, n]$ são as funções de base da transformada inversa, que são as versões conjugadas complexas de $\psi[k, n]$.

A transformada é definida como sendo ortonormal se as funções $\psi[k, n]$ e $\psi^*[k, n]$ satisfazem a condição de ortonormalidade, dada pela seguinte equação:

$$\sum_{n=0}^{N-1} \psi[k, n] \psi^*[l, n] = \begin{cases} 1, & l = k, \\ 0, & l \neq k. \end{cases}$$

Em outras palavras, as bases $\psi[k, n]$ e $\psi^*[k, n]$ são ortonormais entre si.

Por exemplo, se utilizarmos as exponenciais complexas $\frac{1}{\sqrt{N}}e^{j2\pi kn/N}$, com $n, k \in [0, \dots, N-1]$, podemos representar o sinal $x[n]$ como

$$x[n] = \sum_{k=0}^{N-1} X[k] \frac{1}{\sqrt{N}} e^{j2\pi kn/N}. \quad (2.3)$$

Assim, temos uma representação do sinal $x[n]$ como um somatório de bases exponenciais complexas ponderadas pelos coeficientes $X[k]$. Pode-se demonstrar que se $\frac{1}{\sqrt{N}}e^{j2\pi kn/N}$ e $\frac{1}{\sqrt{N}}e^{-j2\pi kn/N}$ são ortonormais entre si. Logo, podemos obter os coeficientes da transformada utilizando a seguinte equação:

$$X[k] = \sum_{n=0}^{N-1} x[n] \frac{1}{\sqrt{N}} e^{-j2\pi kn/N}.$$

Após o ajuste de colocar os fatores escalares $\frac{1}{\sqrt{N}}$ na equação de síntese 2.3, obtemos o par de equações

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j2\pi kn/N}$$

e

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N}.$$

Esse par de equações é conhecido como transformada discreta de Fourier (*discrete Fourier transform* - DFT). Da mesma forma, outras transformadas são definidas a partir da seleção de funções de base diferentes.

2.1.2 Notação Matricial das Transformadas

Considere um sinal em domínio discreto finito unidimensional x de tamanho N , que pode ser visto como um vetor \vec{x} coluna $N \times 1$ no espaço \mathbb{R}^N , com elementos $x[n]$. Representaremos as funções de base como N vetores $N \times 1$ $\psi_i, i = [1, \dots, N]$. Utilizando a matriz $\Psi = [\psi_1 | \psi_2 | \dots | \psi_N]$, podemos representar um sinal $x[n]$ como

$$x = \sum_{i=0}^{N-1} s_i \psi_i \quad (2.4)$$

ou

$$x = \Psi s, \quad (2.5)$$

em que s é o vetor de coeficientes $N \times 1$.

Para o exemplo dado pela Equação 2.1.1, no qual a base é formada por impulsos deslocados, Ψ é a matriz identidade. A DFT de $x[n]$ é obtida ao multiplicar a matriz $A_{i,j} = e^{-j2\pi il/N}$, com $0 \leq i, l \leq N - 1$, pelo sinal $x[n]$.

2.2 *Compressive Sensing*

O teorema da amostragem de Nyquist-Shannon diz que uma taxa de amostragem de um sinal igual a duas vezes a maior frequência do sinal, é suficiente para o sinal ser reconstruído sem perda de informação [38]. Chamamos essa taxa de captura de **taxa de Nyquist**, a qual é geralmente utilizada nos processos de aquisição tradicionais de sinais. Em muitas aplicações, o sinal contém frequências muito altas e, para a sua aquisição, a taxa de Nyquist é muito alta. Tão alta que os sensores necessários para capturar a estas taxas podem ser muito caros. Como muitos dados são gerados, o armazenamento e transmissão do sinal se torna difícil. Particularmente, para sinais de vídeo que têm uma grande quantidade de dados, é necessário realizar uma compressão dos dados no domínio do espaço e do tempo. Além disso, os sensores temporais de alta frequência são muito caros.

Compressive sensing [6, 8, 14, 7, 3] é um novo método de captura e representação de sinais aproximadamente esparsos a uma taxa significativamente inferior a taxa de Nyquist. Esse método adquire o sinal por projeções lineares em um número menor do que a taxa de Nyquist, em uma base em que o sinal seja aproximadamente esparsos. Além disso, o método reconstrói o sinal por otimização em medidas de esparsidade do sinal, considerando a restrição das projeções lineares.

Na seção seguinte, apresentamos o conceito de esparsidade de um sinal, característica necessária para aplicar o método *compressive sensing*. Também apresentaremos o processo de aquisição e recuperação do sinal.

2.2.1 Sinais Esparsos e Aproximadamente Esparsos

Um sinal é dito esparsos na base Ψ se os coeficientes desse sinal representado nessa base são esparsos, ou seja, apenas poucos coeficientes deste sinal são não nulos nesta base. Denotaremos um sinal com K coeficientes não nulos de sinal K -esparsos, ou seja, ele é uma combinação linear de apenas K colunas de Ψ . Quanto menor o valor de K , mais esparsos é o sinal. A Figura 2.1 ilustra um sinal 6-esparsos de comprimento 64.

Um sinal é dito aproximadamente esparsos se a representação do sinal na base Ψ tem poucos coeficientes altos e muitos coeficientes com valores desprezíveis. Os sinais naturais, tais como áudios, imagens e vídeos tirados da natureza, são geralmente aproximadamente esparsos.

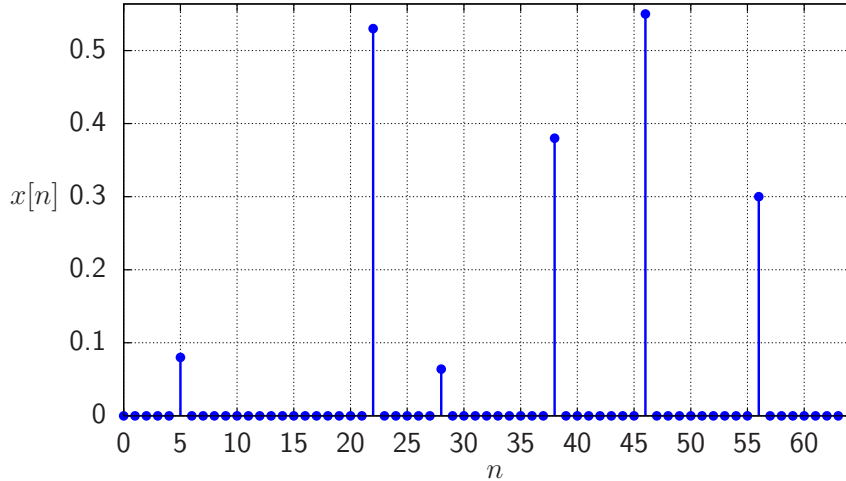


Figura 2.1. Exemplo de um sinal em representação esparsa (domínio transformado). Apenas 6 dos 64 coeficientes são não nulos.

Em compressão de sinais, as propriedades de sinais esparsos ou aproximadamente esparsos são bastante utilizadas. Se um sinal é esparsos na base Ψ , podemos representar/enviar apenas os poucos coeficientes não nulos, juntamente com suas posições. Na hora de reconstruir o sinal completo, utilizamos apenas esses poucos coeficientes para recuperá-lo. Desta forma obtemos uma reconstrução perfeita. Para sinais aproximadamente esparsos, em um processo de compressão com perdas, os coeficientes com pouca energia são desprezados (geralmente na fase de quantização) e a reconstrução não é perfeita. A qualidade do sinal reconstruído dependerá da quantidade de coeficientes desprezados e da importância perceptual desses coeficientes [40].

2.2.2 Aquisição por Medidas Lineares

Seja N o tamanho do sinal x a ser adquirido. Suponha que este sinal é K -esparsos em uma base Ψ (ou seja, $x = \Psi s$, em que s tem apenas K coeficientes não nulos), com $K \ll N$. Seja y a projeção linear de x em um conjunto de $M < N$ vetores $\Phi_i, i = [1, \dots, M]$ de tamanho N . Considerando o conjunto de vetores como a matriz Φ de tamanho $M \times N$, podemos representar y por

$$y = \Phi x = \Phi \Psi s = \Theta s, \quad (2.6)$$

em que $\Theta = \Phi \Psi$ é uma matriz $M \times N$. Chamaremos Φ de matriz de medidas. Temos, então, o processo de aquisição de y , no qual y tem apenas M medidas lineares de x .

De fato, na maioria das aplicações utilizando *compressive sensing*, o sinal x não é conhecido a priori (até porque, se x já fosse conhecido, não precisaríamos o reconstruir). Os hardwares na verdade obtêm o sinal y em seu processo de aquisição, a partir de

um processo que simula a relação $y = \Theta x$, sem que essa operação matricial tenha sido realizada.

2.2.3 O Problema da Reconstrução

O problema da reconstrução consiste em recuperar o sinal x de tamanho N ou a sua representação esparsa s na base Ψ , a partir das M medidas lineares y e da matriz de medidas Φ . Para sinais K -esparsos existem infinitos s' que satisfazem $\Theta s' = y$, pois Θ é uma matriz sub-determinada (Θ tem tamanho $M \times N$, com $M < N$, ou seja, tem menos linhas que colunas). Devemos encontrar o s' que seja exatamente K -esparso. Temos, portanto, que resolver um sistema linear inverso subdeterminado.

2.2.4 Projeto da Matriz de Medidas

A matriz de medidas Φ deve permitir que o sinal x de tamanho N possa ser reconstruído a partir de apenas M amostras. Esse, aparentemente, é um problema mal posto¹, pois $M < N$. Se x é K -esparso e se as posições dos coeficientes diferentes de zero são conhecidas, então o problema pode ser resolvido para $M \geq K$. Uma condição necessária e suficiente para que o problema seja bem posto é que todo vetor v que tenha K coeficientes não nulos deve obedecer a seguinte equação [3]:

$$1 - \epsilon \leq \frac{\|\Theta v\|_2}{\|v\|_2} \leq 1 + \epsilon \quad (2.7)$$

para algum $\epsilon > 0$ pequeno. Ou seja, a matriz Θ deve preservar os comprimentos de todos os vetores K -esparsos. Em geral, as posições dos coeficientes não-nulos não são conhecidas. Entretanto, uma condição suficiente para a solução estável para sinais exatamente K -esparsos ou aproximadamente esparsos é que Θ satisfaça 2.7 para qualquer vetor v $3K$ -esparso. Essa condição é a propriedade de isometria restrita (*restricted isometry propriety* - RIP) [3, 6] e, quanto mais o valor de ϵ se aproxima de zero, melhor essa condição é satisfeita. Outra condição relacionada, conhecida como *incoerência*, requer que as linhas de Φ não possam representar esparsamente as colunas de Ψ e vice-versa.

Para construirmos uma matriz Φ de tal forma que $\Theta = \Phi\Psi$ satisfaça a RIP, teríamos que verificar todas as $\binom{N}{K}$ possíveis combinações de sinais K -esparsos de tamanho N . Porém, se escolhermos Φ tomando M linhas aleatórias da matriz identidade, com alta probabilidade ela satisfaz a RIP [14]. Para satisfazer a incoerência, basta escolher Ψ de tal forma que a representação das linhas de Φ não sejam esparsas. Para a Φ escolhida,

¹Um problema bem posto é um problema que tem uma única solução e há pouca variação em sua solução quando há pouca variação nas condições iniciais deste problema. Um problema mal posto admite mais de uma solução se a sua solução varia muito com pequenas variações nas condições iniciais [46]

podemos escolher Ψ como, por exemplo, a matriz de coeficientes da DFT, pois os coeficientes da DFT aplicadas a sinais do tipo $\delta[n - k]$ (que correspondem às linhas de Φ) têm módulo 1 em qualquer posição, ou seja, não são esparsas.

Uma forma de medir a incoerência de um sistema de *compressive sensing*, juntamente com suas relações de transformada esparsificante, matriz de medidas, número de medidas, é através da função *Transform Point Spread Function* (TPSF) proposta por Lustig *et al.* [34]. Essa função é definida da seguinte forma:

$$TPSF(i, j) = e_j^* \Psi \Phi^* \Phi \Psi^* e_i, \quad (2.8)$$

em que e_j é o vetor-coluna de tamanho N com todos os componentes iguais a zero, exceto na posição j . Idealmente, é desejado que a $TPSF(i, j)$ tenha energia muito baixa e um comportamento ruidoso para $i \neq j$. Isto implica que artefatos de *aliasing* introduzidos por sub-amostragem produzam uma incerteza relativamente pequena ao resolver o i -ésimo coeficiente da transformada no j -ésimo coeficiente [43].

Para o caso em que o sinal já é esparsa, e não aplicamos nenhuma transformada esparsificante (ou seja, Ψ é igual a identidade), a TPSF se reduz à *Point Spread Function* (PSF):

$$PSF(i, j) = e_j^* \Phi^* \Phi e_i. \quad (2.9)$$

Uma forma simples de se avaliar a incoerência de um sistema de *compressive sensing*, é medir o máximo lóbulo lateral (*sidelobe-to-peak* - SPR), dado por:

$$\max_{i \neq j} \left| \frac{PSF(i, j)}{PSF(i, i)} \right|. \quad (2.10)$$

A Figura 2.2 ilustra alguns resultados intermediários do cálculo das funções PSF e a TPSF para um Ψ igual à transformada wavelet e Φ igual a uma função de sub-amostragem de coeficientes aleatórios da transformada discreta de Fourier, aplicados a um impulso.

O resultado final dos processos ilustrados pela Figura 2.2(a) e 2.2(b), mostra um sinal parecido com o sinal de entrada: um impulso na mesma posição que o sinal de entrada, porém, existem pequenas perturbações de pouca intensidade nas outras posições. Ao se multiplicar o sinal resultante por um impulso na mesma posição (que é a parte do processo mostrado em 2.9 mas não ilustrado na Figura 2.2), o resultado contém alta energia. Se multiplicarmos o sinal resultante pelo impulso em alguma outra posição, o resultado tem baixa energia. Este é o comportamento esperado para um sistema de *compressive sensing* incoerente.

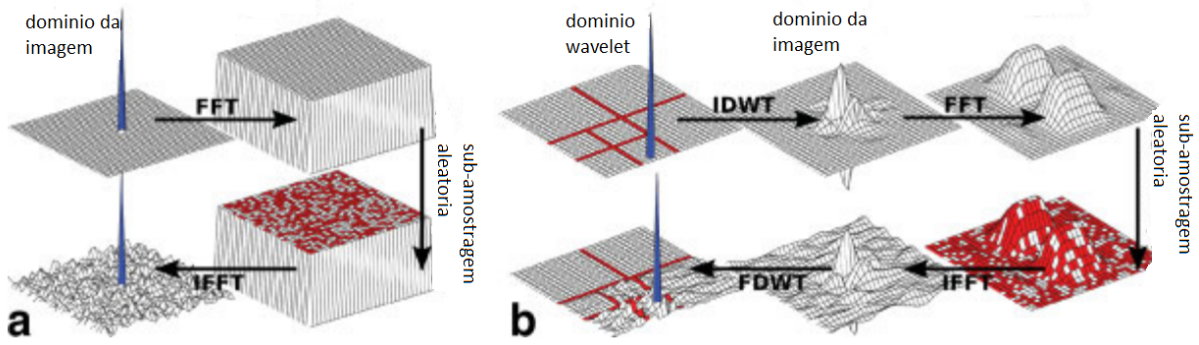


Figura 2.2. Exemplo de parte do processo da aplicação das funções PSF e TPSF, adaptado de [34]. a) Ilustra parte do processo do PSF para uma sub-amostragem dos coeficientes da FFT. Há um pico grande e um ruído de baixa energia fora do ponto escolhido como sinal impulso de entrada, isso mostra que a sub-amostragem da FFT apresenta baixa incoerência. b) Ilustra parte do processo da TPSF para a mesma ϕ , porém no domínio wavelet. Esta configuração apresenta o mesmo comportamento, então o sistema de *compressive sensing* desta forma é incoerente.

2.2.5 Projeto do Algoritmo de Reconstrução

Para obter as possíveis soluções, precisamos primeiro definir o tipo de métrica que utilizaremos: a métrica ℓ_r . Definimos a ℓ_r [19] de um vetor s como

$$\|s\|_r = \left(\sum_{i=1}^N |s_i|^r \right)^{\frac{1}{r}} \quad (2.11)$$

para qualquer $r \in \mathbb{R}_*^+$. Essa classe de métrica é a mais genérica das métricas utilizadas nos processos de reconstrução por *compressive sensing*. Como por exemplo as métricas ℓ_0 , ℓ_1 , ℓ_2 , entre outras. Definimos como bola ℓ_r todos os pontos do espaço \mathbb{R}^N que têm a mesma ℓ_r .

Minimização da ℓ_2

Quando escolhermos o valor de $r = 2$, obtemos a métrica ℓ_2 . Ela é bastante conhecida como norma de um vetor, pois mede a distância euclidiana do vetor até a origem [19]. Uma abordagem clássica para a solução de problemas lineares inversos é encontrar o vetor com a menor ℓ_2 , resolvendo o problema dado pela equação:

$$\hat{s} = \operatorname{argmin}_{s'} (\|s'\|_2), \text{ sujeito a } \Theta s' = y. \quad (2.12)$$

Essa otimização tem a solução fechada da forma $\hat{s} = \Theta^T(\Theta\Theta^T)^{-1}y$. Porém, a minimização da ℓ_2 quase nunca leva a uma solução esparsa, pois ela mede a energia do sinal, que nem sempre está relacionada com a sua esparsidade.

Na Figura 2.3, podemos ver geometricamente porque a solução gerada não é esparsa, em um exemplo para $N=2$. A bola ℓ_2 é formada por todos os pontos em \mathbb{R}^2 que têm a mesma ℓ_2 , nesse caso representada pelo círculo. Ao minimizar a ℓ_2 mantendo a restrição $\Theta s' = y$ (representada pelo segmento de reta), estamos procurando o círculo de menor raio que intercepte a reta. A solução final é o sinal representado pelo ponto vermelho em 2.3(b). Podemos ver que esse sinal não é esparsa, pois o ponto que representa a solução final não se encontra nem no eixo x , nem no eixo y , ou seja, não contém nenhum valor zero em suas coordenadas.

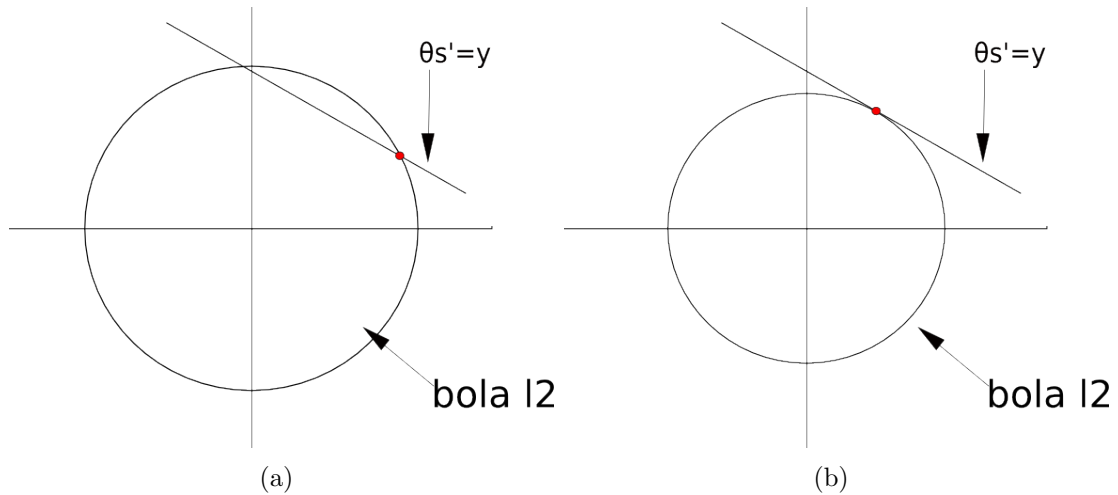


Figura 2.3. Minimização da ℓ_2 para um domínio de tamanho 2. a) A solução inicial, e com as iterações é escolhido o ponto que tem a menor ℓ_2 , como mostrado em b)

Minimização da ℓ_0

Considerando que

$$r^0 = \begin{cases} 0 & , \text{se } r = 0 \\ 1 & , \text{se } r \neq 0, \end{cases}$$

então, por 2.11, a métrica ℓ_0 de um vetor soma 1 para cada elemento do vetor que é diferente de zero, e soma zero caso contrário. Ou seja, essa métrica conta o número de elementos não nulos deste vetor. Então, a otimização dada pela equação:

$$\hat{s} = \operatorname{argmin}_{s'} (\|s'\|_0), \text{ sujeito a } \Theta s' = y \quad (2.13)$$

recupera um sinal K -esparsa fielmente. Porém, o problema de minimização 2.13 é um problema NP-completo, uma vez que sua solução geralmente exige que todas as $\binom{N}{K}$

possíveis posições de entradas não nulas de s' sejam testadas. Desta forma, a utilização da ℓ_0 não é viável em aplicações reais, uma vez que, mesmo para sinais pequenos, o tempo de execução se torna demasiadamente grande.

Minimização da ℓ_1

A otimização baseada na métrica ℓ_1 é definida pela seguinte equação:

$$\hat{s} = \operatorname{argmin}_{s'} (\|s'\|_1), \text{ sujeito a } \Theta s' = y. \quad (2.14)$$

Resolvendo essa equação podemos recuperar sinais K -esparsos, pois o hiperplano $y = \Theta s'$ corta a menor bola ℓ_1 onde aparece a maior quantidade de zeros. Observe que a ℓ_1 de um vetor é linear em relação aos termos deste vetor. Então, 2.14 é um problema de programação linear.

Vamos utilizar o mesmo exemplo apresentado na Figura 2.3 para mostrar geometricamente porque a ℓ_1 é mais adequada para sinais esparsos. A bola ℓ_1 é representada pelo quadrado, o qual inclui todos os pontos em \mathbb{R}^2 que têm a mesma ℓ_1 . Ao minimizar a ℓ_1 mantendo a restrição $\Theta s' = y$ (representada pelo segmento de reta), estamos procurando o menor quadrado que intercepta a reta. A solução final é o ponto vermelho na Figura 2.4(b). Note que a solução final é mais esparsa, pois tem uma coordenada igual a zero. Podemos ver que a menor bola ℓ_1 vai interceptar a reta (no geral, o hiperplano) sempre no ponto com maior quantidade de zeros.

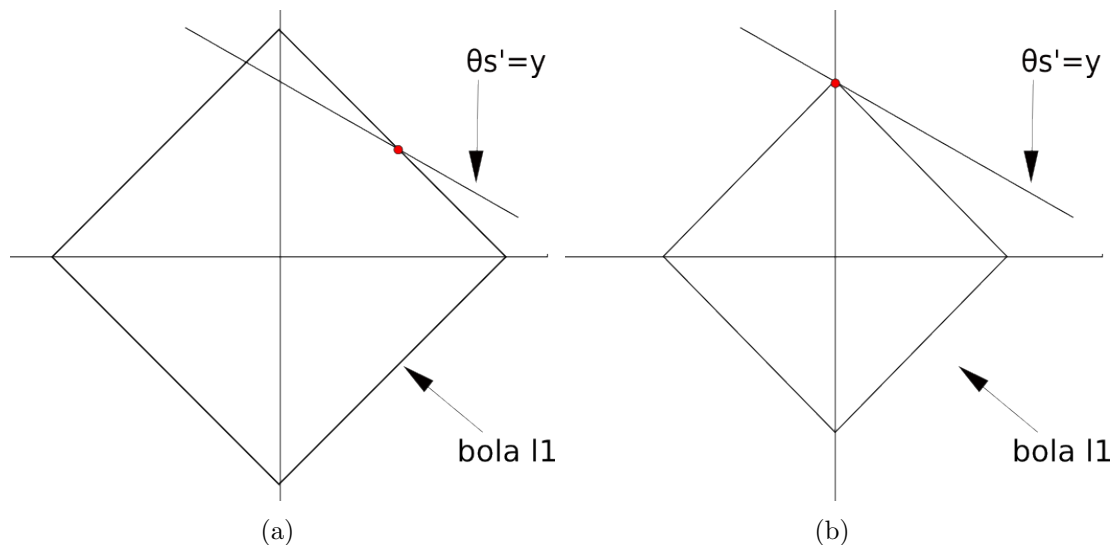


Figura 2.4. Minimização da ℓ_1 para um domínio de tamanho 2. a) A solução inicial, e com as iterações é escolhido o ponto que tem a menor ℓ_1 , como mostrado em b), neste caso a solução escolhida é mais esparsa.

Minimização da Variação Total (*Total Variation*)

A minimização ℓ_1 tem sido amplamente utilizada desde a descoberta que a minimização ℓ_1 leva a soluções próximas em relação a minimização ℓ_0 . Entretanto, para reconstrução de imagens, pesquisas mostram que o uso da minimização da variação total (*total variation* - TV) tem levado a melhores resultados. As imagens reconstruídas pela minimização TV apresentam bordas mais definidas, o que caracteriza melhor a imagem e leva a melhores resultados do que a ℓ_1 [30]. Seja $D_i s$ o gradiente discreto do vetor s . O problema de otimização utilizando a minimização TV é um pouco diferente do problema descrito pela Equação 2.14. A partir das mesmas medidas, resolvemos o seguinte problema

$$\hat{s} = \operatorname{argmin}_{s'} (\|D_i s'\|_r) \text{ sujeito a } y = \Theta s', \quad (2.15)$$

para $r = 1$ ou $r = 2$. O modelo de TV é denominado isotrópico para $r = 2$, e denominado anisotrópico para $r = 1$ [32].

Ao resolver o problema descrito por 2.15 no lugar do problema descrito por 2.14, estamos alterando o modelo para melhor representar imagens. Ao invés de procurar a imagem mais esparsa na base escolhida, estamos procurando pela imagem com o menor ℓ_1 ou ℓ_2 do gradiente, isto diminui os elementos de alta frequência que causavam oscilações na imagem [7]. Neste trabalho, utilizamos o modelo anisotrópico, e denotaremos a métrica TV desta forma como $\|\cdot\|_{TV}$.

2.2.6 *Compressive Sensing* para Sinais Aproximadamente Esparsos

Normalmente os sinais de imagem e vídeo não são exatamente esparsos, mas eles são aproximadamente esparsos em alguma base Ψ . Em outras palavras, poucos dos seus coeficientes têm muita energia, enquanto que a maioria têm baixíssima energia. Geralmente na fase de quantização de um processo de compressão, esses coeficientes de pouca energia são desconsiderados. O sinal reconstruído sem esses coeficientes fica indistinguível do sinal original, segundo a percepção humana [40].

Na próxima seção, consideramos que esses coeficientes de pouca energia atuam no sinal da mesma forma que medidas ruidosas atuariam, uma vez que não é necessário saber exatamente o valor desse ruído em cada coeficiente. Assim, ao modelar a aquisição considerando medidas ruidosas, teremos uma boa aproximação para a reconstrução de um sinal não exatamente K-esparso, mas apenas aproximadamente esparso. Neste caso, o valor dos coeficientes de baixa energia não é importante.

2.2.7 Aquisição por Medidas Lineares Ruidosas

Seja Θ a mesma matriz de 2.6, o processo de aquisição é semelhante. A diferença é apenas que assumimos que o dado adquirido é ruidoso, ou seja, adicionamos um erro e à 2.6 [7]:

$$y = \Theta s + e, \quad (2.16)$$

no qual e é uma pequena perturbação com amplitude limitada $\|e\|_2 < \delta$, para $\delta > 0$ pequeno.

2.2.8 Reconstrução a Partir de Medidas Ruidosas

Após adquirir as amostras, temos que recuperar o sinal. Baseando-se no fato de que Θ satisfaz a propriedade da isometria restrita, obtemos uma boa aproximação da solução resolvendo o problema de otimização dado pela equação abaixo:

$$\hat{s} = \operatorname{argmin}_{s'} (\|s'\|_1) \text{ sujeito a } \|y - \Theta s'\|_2 < \delta. \quad (2.17)$$

Os problemas de otimização dessa forma são conhecidos como problemas de otimização com restrição quadrática. A solução não é exatamente a solução exata, mas uma aproximação suficientemente boa. Esta solução difere do sinal original apenas por um nível de ruído [7] que obedece a seguinte desigualdade

$$\|\hat{s} - s\|_2 < C \cdot \delta, \quad (2.18)$$

em que C é uma constante. Assim, o erro da reconstrução é proporcional ao ruído do sinal original.

2.3 Aquisição de Vídeo Digital

Em uma cena, uma fonte de luz ilumina os objetos, que absorvem parte da luz e refletem outra parte. A luz que foi refletida passa pelo obturador da câmera e é capturada pelos sensores da câmera durante um certo período de tempo. Assim, obtemos um quadro do vídeo e o processo se repete a cada instante de tempo (tempo de integração da câmera). A velocidade em que cada quadro é obtido é relacionado com a resolução temporal da câmera, a qual é medida em taxa de quadros por segundo (*frames per second* - FPS), que é igual ao inverso do tempo de integração da câmera.

A Figura 2.5 ilustra o processo de captura de uma imagem em uma câmera digital. Como cada sensor tem um formato quadrado, a imagem recebida é amostrada de acordo

com o tamanho dos pixels. Após esse processo, a imagem é quantizada por um conversor analógico-digital (ADC). Ao final, obtemos uma sinal em domínio discreto quantizado e em duas dimensões, ou seja, uma imagem em níveis de cinza [21].

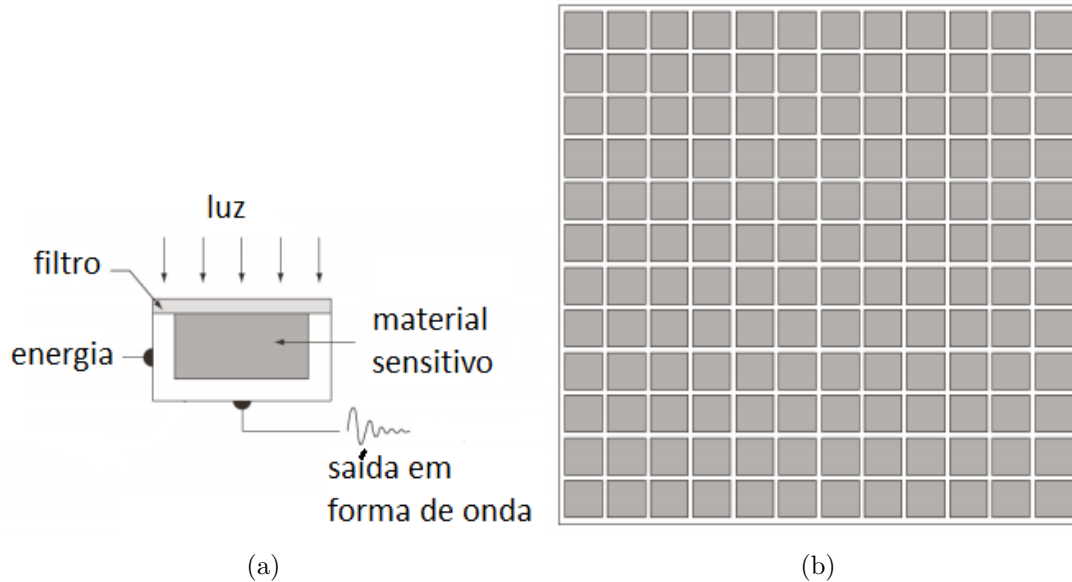


Figura 2.5. a) Um sensor e seus componentes e b) um exemplo de uma disposição de vários sensores em forma matricial utilizados para aquisição de imagem [21].

Para câmeras coloridas, filtros por cores são utilizados. Esses filtros deixam passar as ondas com luz em frequências específicas (vermelho, verde e azul), como vemos na Figura 2.6(a). Na Figura 2.6(b) vemos que há mais sensores capturando o verde do que sensores capturando as outras cores, isso se deve ao fato do sistema visual humano ser mais sensível ao verde do que às outras cores.

2.3.1 Tempo de Integração e de Exposição

O tempo de integração da câmera é o tempo total do intervalo entre dois quadros consecutivos. Ao final desse tempo, a amplitude do sinal elétrico de saída de cada sensor é proporcional a quantidade de fótons que o atingiram durante esse tempo. Esse tempo é fixo e depende da capacidade de renovação dos sensores.

O tempo de exposição é o tempo em que os sensores irão efetivamente receber luz. Em vídeos, esse tempo é menor ou igual ao tempo total de integração, pois é limitado a taxa FPS da câmera. Nos sensores atuais, esse tempo é ajustável. É possível escolher o início e o final do tempo de exposição, mas esta escolha é limitada pelo tempo de integração da câmera [25].

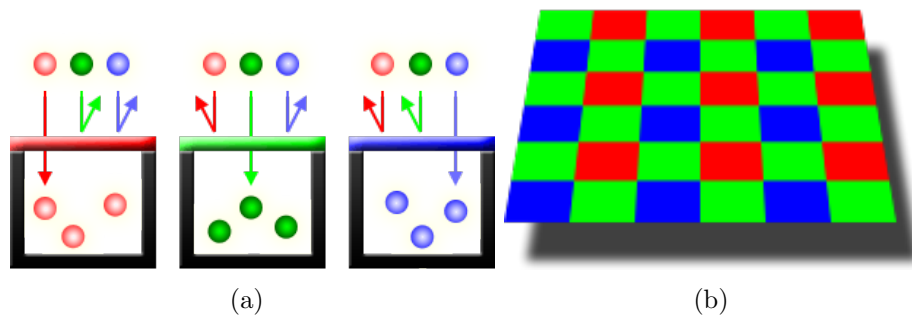


Figura 2.6. Sensores para aquisição de imagens coloridas. Filtros são utilizados para obter os 3 canais de cores. a) Sensores coloridos deixam passar apenas a luz correspondente à cor do filtro e refletem as outras cores. b) A disposição dos filtro de cores. Contém mais filtros verdes para tomar proveito das características do sistema visual humano, que é mais sensível ao verde.

2.3.2 A Escolha do Tempo de Exposição

Câmeras digitais contém um padrão chamado ISO ², que indica o nível de sensibilidade à luz dos sensores da câmera [27]. O ISO é controlado aumentando ou diminuindo o ganho da saída numérica do nível de luz dos sensores. Os padrões ISO são 100, 200, 400, 800, 1600, etc. Os valores padrões variam exponencialmente, sendo que próximo valor é igual dobro do valor anterior.

Se o tempo de exposição for muito pequeno, o sensor recebe pouca luz. Uma forma de equilibrar a quantidade de luz é aumentar o ISO dos sensores. Porém, isso aumenta a relação entre o ruído de medida e a informação de luz lida, aumentando o ruído de leitura dos sensores e gerando uma imagem mais ruidosa. A Figura 2.7 ilustra uma comparação de duas imagens capturadas com a mesma câmera com dois tempos de exposição diferentes, mas com a quantidade de luz equilibrada pelo ISO. Pode-se observar a influência do ISO na quantidade de ruído.

Se escolhermos um tempo de exposição muito grande, teremos uma única imagem que corresponde a todo o tempo de integração da câmera. Para cenas com pouco ou nenhum movimento não há problema. Mas em cenas com movimento rápido, aparece um efeito 'borrado' devido ao movimento (*motion blur*), como ilustrado na Figura 2.8. Nesse efeito, observamos um borrão causado pela trajetória do movimento durante o tempo de exposição. Logo, acabamos perdendo informação da configuração da cena em tempos

²ISO é uma abreviação de *International Organization of Standardization*, um corpo governamental baseado na europa que mantém padrões para uma vasta variedade de assuntos. Entre os padrões ditados por este corpo, os valores padrões de sensibilidade dos sensores de câmeras digitais também estão incluídos, e receberam o acrônimo desta instituição.

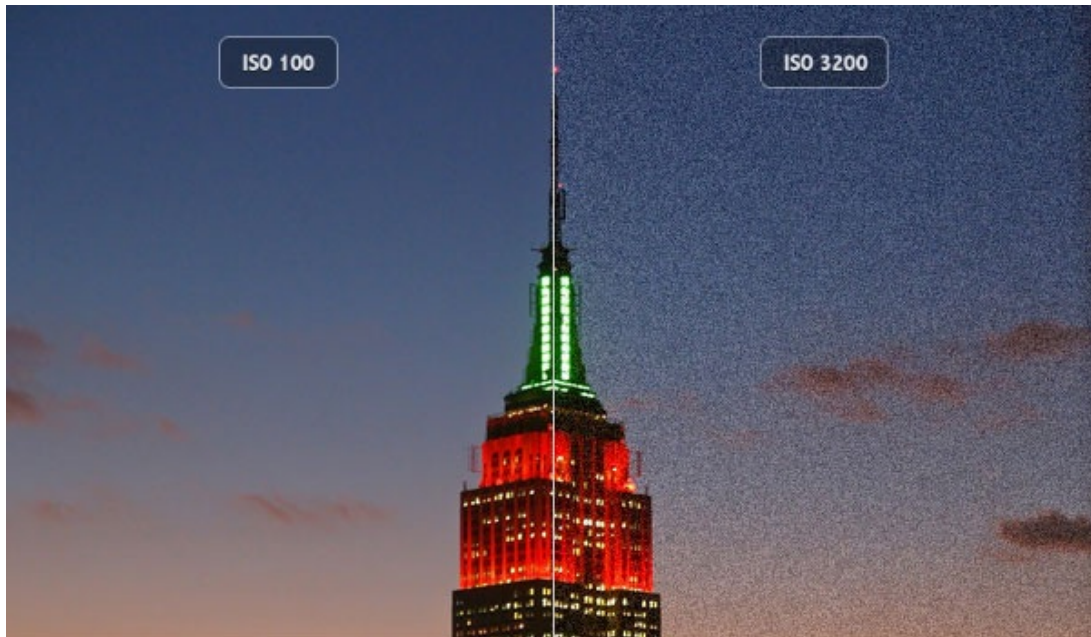


Figura 2.7. Influência do indicador de sensibilidade ISO no ruído obtido pelos sensores, adaptado de [22]. Na parte da esquerda da imagem, o tempo de exposição é muito maior, o que permite um valor pequeno de ISO. Na parte da direita da imagem, o tempo de exposição é muito menor, o que requer um valor do ISO bem maior para obter o mesmo nível de intensidade. Em consequência disso, o nível de ruído é muito maior.

específicos da imagem. Obtemos então uma integração (soma infinitesimal dos quadros pelo tempo) de todas as configurações da cena durante todo o tempo de exposição.



Figura 2.8. Borrão de movimento (*Motion Blur*) devido ao alto tempo de exposição. Em (a) a câmera acompanha o carro, percebe-se o efeito *motion blur* nas letras no fundo. Em (b) o guarda-chuva rotaciona rapidamente e, pelo movimento ser circular, a parte interna dele se move mais lentamente do que a parte externa. Podemos observar que o efeito *motion blur* está mais evidente na parte externa do guarda-chuva, onde a quantidade de movimento da cena é maior.

2.3.3 Moduladores de Luz

Moduladores de Luz Globais

Muitas vezes deseja-se um controle maior sobre a exposição da câmera. Em algumas aplicações [26, 1, 53] é necessário que a exposição não seja controlada por apenas um instante de tempo que se inicia e termina durante o tempo de integração. Nessas aplicações, a exposição se inicia e termina várias vezes durante o tempo de integração, em instantes de tempo arbitrários, como ilustrado na Figura 2.9. Chamaremos os dispositivos e arquiteturas de sensores que aplicam este tipo de exposição de moduladores de luz globais. Atualmente, algumas câmeras comerciais já possuem esse modo de exposição implementado.

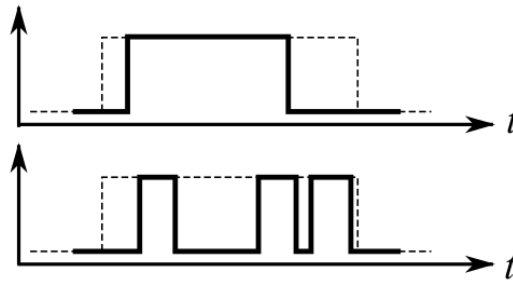


Figura 2.9. Exposição por modulador global. No gráfico de cima, vemos o esquema tradicional de exposição, ligamos e desligamos a exposição apenas uma vez durante o tempo de integração (linha tracejada). Com o modulador global, podemos ligar e desligar a exposição quando desejarmos durante o tempo de integração.

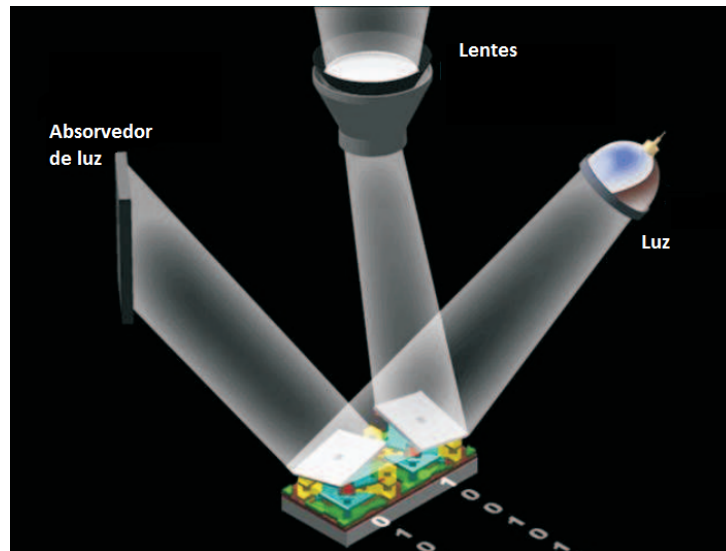
Moduladores de Luz Espaciais

Em algumas aplicações [23, 47, 25] é desejado uma forma ainda maior de controle sobre a exposição. Nessas aplicações, além do controle temporal da exposição como feita por moduladores de luz globais, deseja-se que o controle seja independente para cada pixel. Chamamos este tipo de moduladores de moduladores de luz espaciais (*spatial light modulators* - SLM). Ao controlar a luz adquirida em cada posição espacial, inclui-se também a possibilidade do comportamento dos espelhos ser o mesmo em todos os instantes. Neste caso, eles também funcionam como moduladores de luz globais.

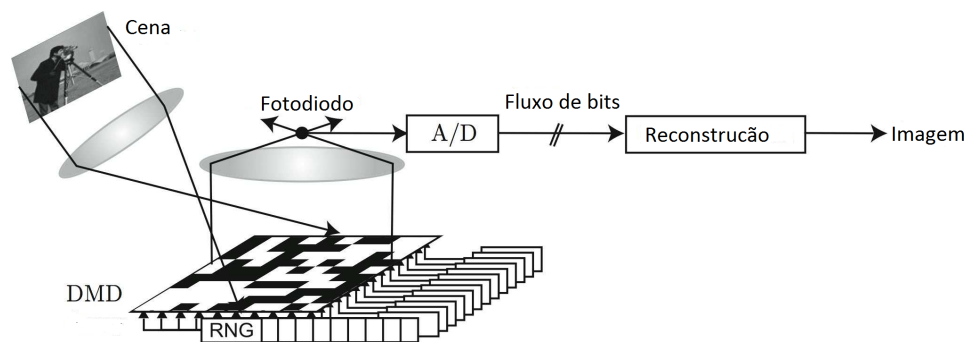
Existem, atualmente, alguns dispositivos que implementam a modulação de luz espacial. Um deles é conhecido como *digital micromirror device* (DMD) [15, 39, 5]. O DMD é implementado por um conjunto de microespelhos, um para cada pixel, que mudam a angulação de acordo com um controle. Há dois tipos possíveis de orientações. A primeira angulação aponta para a lente que leva a luz ao conjunto de sensores. A segunda

aponta para fora da lente. Dessa forma, a luz recebida por cada sensor pode ser ligada ou desligada. Na Figura 2.10(a) vemos uma ilustração de como funciona o DMD.

A técnica conhecida como *single pixel camera* (SPC) [52] utiliza um modulador por pixel, implementada com um DMD. Neste caso, a luz de todos os pixels ligados vai até um único sensor e a luz desligada é perdida. Ilustramos o processo na Figura 2.10(b).



(a)



(b)

Figura 2.10. Funcionamento do dispositivo de espelhos DMD e sua aplicação na técnica *Single Pixel Camera* (SPC). a) Ilustração do dispositivo DMD. Os espelhos que recebem ‘0’ são inclinados para algo que absorve a luz e os espelhos que recebem ‘1’ são inclinados para a lente. b) Uso do DMD para a técnica SPC [52].

Vale citar que outros dispositivos também podem implementar o tipo de exposição controlada individualmente para cada pixel. Porém, estes dispositivos não modulam o pixel em ‘ligado’(‘1’) e ‘desligado’(‘0’) inclinando espelhos. Eles controlam a absorção de luz e reflexão de luz alterando a característica do físico-química dos espelhos correspondentes a cada pixel durante o tempo. Por exemplo, dispositivos utilizando cristal líquido (*liquid crystal on silicon* - LCoS) [28, 37, 4] ou efeitos magnético-óticos [48, 49], entre outras

alternativas que oferecem magnitudes de velocidade de atualização na ordem de terahertz [10, 11].

Além de aplicações em modelamento de luz por pixel para vídeos de alta resolução temporal (o foco deste trabalho), o dispositivo DMD também apresenta aplicações em litografia, onde é utilizado para impressão de chips de DNA para descoberta de drogas, exibição volumétrica de imagens tridimensionais, telecomunicações, holografia, codificação de informação, entre muitas outras aplicações [17].

2.4 Modos de Aquisição de Medidas Lineares de Vídeo

Como mencionado na Seção 2.2, a teoria de *compressive sensing* nos permite recuperar um sinal esparso em uma determinada base, a partir de poucas medidas lineares. Geralmente, a aquisição do sinal é feita realizando medidas aleatórias em algum domínio transformado, como por exemplo a DFT. Porém, para a aquisição de vídeos, as amostras só podem ser obtidas no domínio espaço-temporal, pois os sensores lêem apenas a quantidade de luz em um único espaço por um certo tempo. Desta forma, obtém-se um valor de intensidade luminosa para cada pixel a cada quadro.

Nesta sessão, vamos mostrar os métodos de aquisição de medidas lineares para *compressive sensing* em vídeo propostas na literatura. Seja uma câmera com resolução espacial N pixels e resolução temporal de q quadros por segundo, o tempo de integração da câmera é $\frac{1}{q}$. Os métodos de aquisição de medidas lineares são descritos considerando essas constantes.

2.4.1 Forma Tradicional de Aquisição de Cenas

A forma tradicional de aquisição das cenas em uma câmera de vídeo digital consiste em captar a luz de cada quadro utilizando N sensores (pixels) por um certo intervalo de tempo. O tempo de início e fim da exposição da câmera pode ser escolhido, desde que esteja dentro do intervalo de tempo de integração.

Nesse método, a única variável é o tempo de exposição desejado. Se escolhermos um tempo de exposição muito longo, as imagens com movimento aparecerão borradas, o chamado efeito *motion blur*. Contudo, se escolhermos um tempo de exposição muito curto, a imagem pode ficar muito escura ou ruidosa. Isso acontece porque a quantidade de luz que atinge cada sensor diminui, pois cada sensor recebe luz por um intervalo de tempo menor. Além disso, perdemos informação luminosa pois o tempo de exposição à luz é menor que o intervalo de tempo total de integração da câmera.

Ao final do tempo de integração, obtemos uma imagem de tamanho N na qual cada pixel contém toda a luz que atingiu o sensor durante o tempo de exposição.

2.4.2 Método de Aquisição por Modulador Global

Em algumas câmeras atuais (como por exemplo a Point Grey Dragonfly 2) é possível obter um controle ainda maior sobre a exposição. Nelas, é possível ligar e desligar o fluxo de luz nos sensores a uma taxa maior que a taxa de quadros da câmera. Chamaremos esse método de aquisição de modulador global (*flutter shutter* - FS), pois este tipo de aquisição utiliza um modulador global.

Suponha que queremos dividir o tempo de exposição em k unidades temporais de mesma duração de tal forma que o tempo de exposição $1/q$ é dividido em k unidades de $1/(q \cdot k)$. Neste trabalho, chamaremos cada unidade de tempo de sub-quadro. No método de aquisição FS, o tempo de exposição é dividido em sub-quadros, de tal forma que cada sub-quadro pode ter sua exposição ligada ou desligada para todos os pixels.

O esquema de aquisição por modulador global é apresentado na Figura 2.11. Ao final do tempo de integração, obtemos uma imagem de tamanho N , na qual cada pixel contém toda a luz correspondente aos quadros que estavam ‘ligados’, contendo assim a informação de vários sub-quadros de tempos distintos. A menos que os sub-quadros estejam todos ‘ligados’ durante todo tempo de integração (o que corresponde ao modo tradicional de aquisição), cada pixel recebe uma quantidade menor de luz, ou seja, parte da informação luminosa é perdida. Esse modo de aquisição é utilizado por Holoway *et al.* em sua técnica *Flutter Shutter Video Camera* [26], com cerca de 66% de eficiência luminosa.

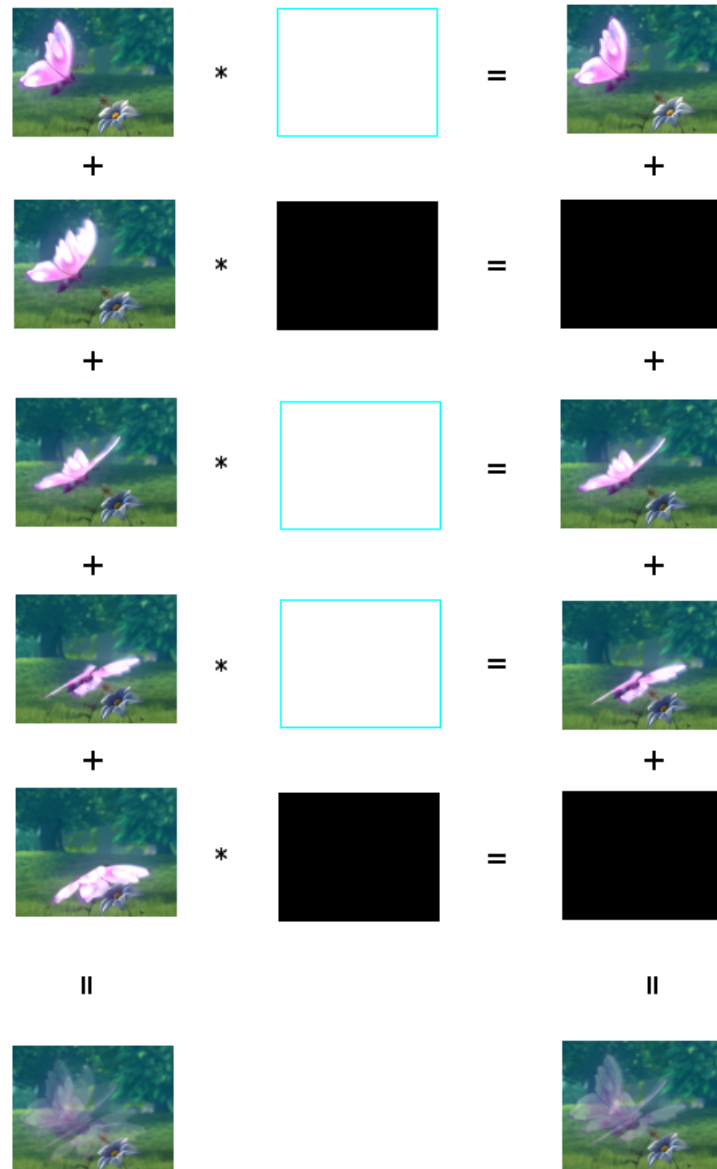


Figura 2.11. Método de aquisição por modulador global. A primeira coluna corresponde a aquisição tradicional, onde todos os sub-quadros são considerados para formar a imagem. A segunda coluna ilustra o padrão dos sub-quadros ‘ligados’ ou ‘desligados’. A terceira coluna mostra a aplicação do padrão à luz da cena, no qual ao final do tempo de integração obtemos uma combinação linear dos sub-quadros da cena original.

2.4.3 Método de Aquisição por Modulador por Pixel

Com moduladores por pixel (*per-pixel shutter* - PPS), temos um controle de exposição ainda maior do que com os moduladores globais, uma vez que podemos ‘ligar’ e ‘desligar’ qualquer um dos N pixels durante o intervalo de tempo de exposição. Para tal, utilizamos *digital micromirror device* (DMD) [15], um dispositivo que ainda não está implementado nas câmeras atuais, exigindo que seja utilizado em um sistema óptico externo à câmera.

A Figura 2.12 mostra como é o sistema óptico que implementa o PPS. A luz da imagem que queremos adquirir passa por uma das lentes e chega até o DMD. A luz que atinge os espelhos ‘desligados’ é direcionada a uma superfície escura, para a menor reflexão de luz, enquanto que a luz que atinge os espelhos ‘ligados’ é direcionada até outra lente e, então, a luz segue até os sensores da câmera.

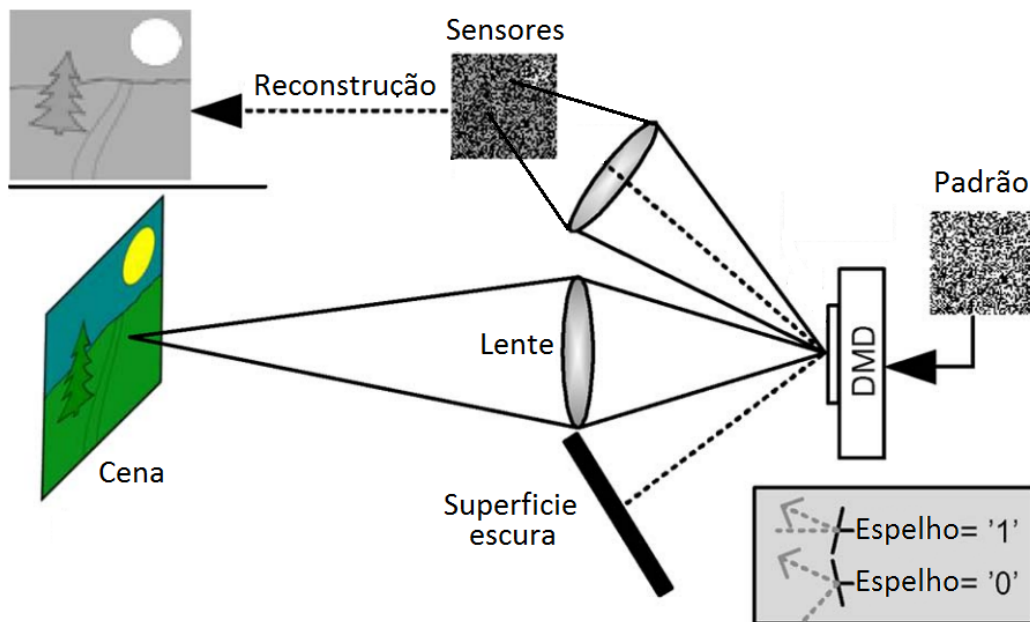


Figura 2.12. Sistema óptico externo que implementa o método *per-pixel shutter* (PPS), adaptado de [18].

De forma semelhante ao que acontece no FS, para utilizar esse método de aquisição para adquirir amostras de vídeo, escolhemos o tempo de cada sub-quadro como uma fração do tempo total de exposição. Desta forma, para qualquer sub-quadro, podemos escolher livremente se um pixel recebe ou não a luz que chegaria ao pixel.

Esse método de aquisição é utilizado de duas formas. Na primeira, cada pixel é ‘ligado’ em um único sub-quadro. Assim, ao final do tempo de integração, todos os N pixels recebem a luz apenas pelo período de tempo de um único sub-quadro, ou seja, cada pixel não contém informações luminosas de sub-quadros distintos. A Figura 2.13 ilustra o funcionamento da método, mostrando a forma em que se adquire as medidas lineares nos

sensores da câmera. Nesta Figura, o nível de intensidade luminosa adquirida foi adaptada para melhor visualização do processo de aquisição.

Essa forma de aquisição é utilizada por Gupta *et al.* em sua técnica conhecida como *flexible voxels* [23], Hitomi *et al.* utiliza esta forma de exposição em sua técnica que utiliza treinamento de dicionário [25], entre outras. Um dos problemas desse método de aquisição é que se perde a grande maioria da informação luminosa no vídeo obtido. Neste trabalho, utilizaremos a abreviação PPS1 para denominar esse método.

Uma segunda forma de aquisição é possível, na qual, em quaisquer sub-quadros, os pixels são ‘ligados’ e ‘desligados’ independentemente. Ao final do tempo de integração, obtemos uma imagem que, para cada pixel, é somada a luz que o sensibilizou durante os sub-quadros em que ele estava ‘ligado’. Ou seja, cada pixel contém a informação temporal de vários quadros simultâneos. Na Figura 2.14 é ilustrado o esquema de aquisição de medidas lineares utilizando o método PPS2. Reddy *et al.* utilizam esta forma de aquisição em sua técnica *programmable pixel compressive camera*(P2C2) [47]. Entretanto, a implementação deste método de aquisição nessas técnicas gera uma perda de cerca de 50% da informação luminosa. Nesse trabalho, utilizaremos a abreviação PPS2 para esse método.

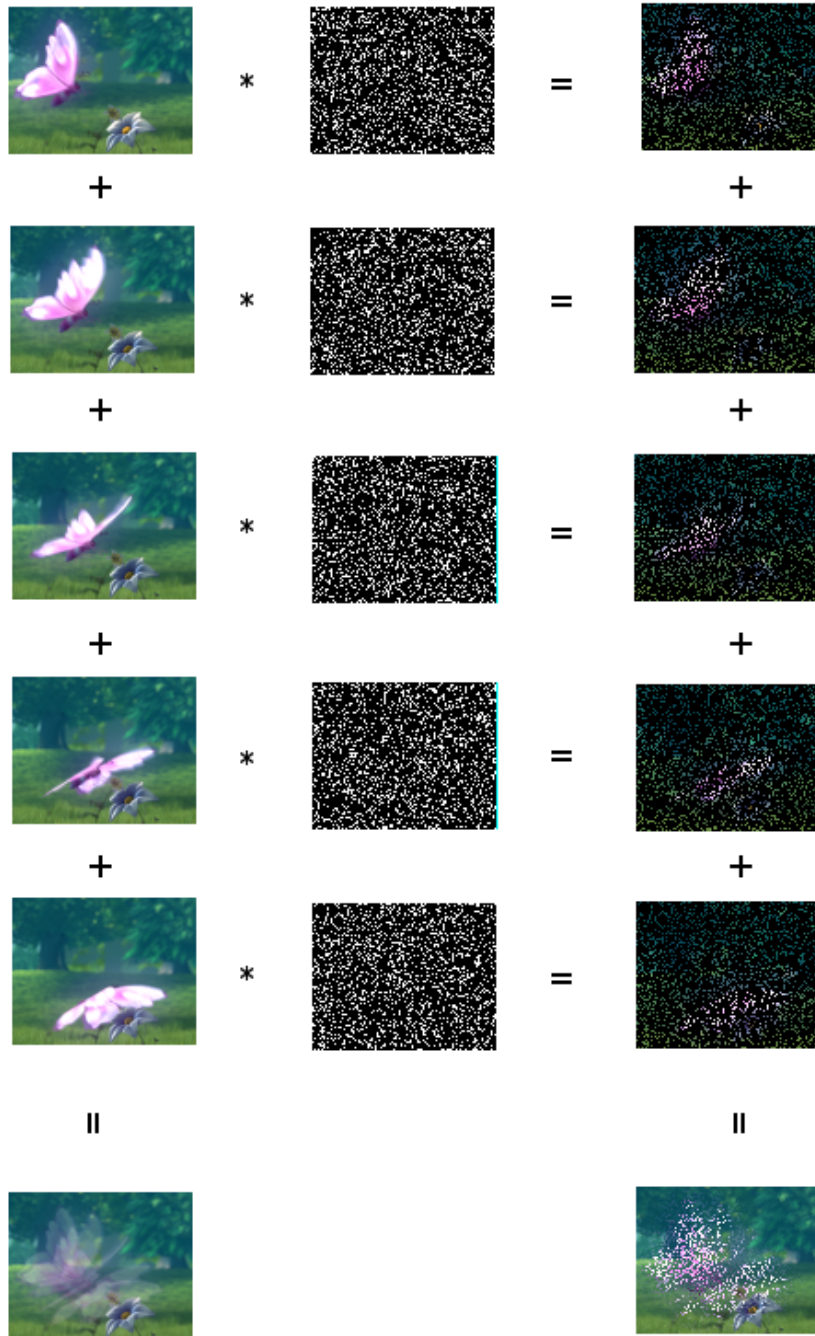


Figura 2.13. Aquisição pela primeira configuração do método *per-pixel shutter* (PPS1). A primeira coluna corresponde a aquisição tradicional, onde todos os sub-quadros são considerados para formar a imagem ao final do tempo de integração. A segunda coluna ilustra, para cada sub-quadro, o padrão de pixels ‘ligados’, que são não-intercalados entre sub-quadros distintos. A terceira coluna ilustra a aplicação do padrão à luz da cena. Ao final do tempo de integração obtemos uma combinação linear dos sub-quadros da cena original operados (operação matemática equivalente a multiplicação de cada sub-quadro pelo padrão) com os padrões correspondentes.

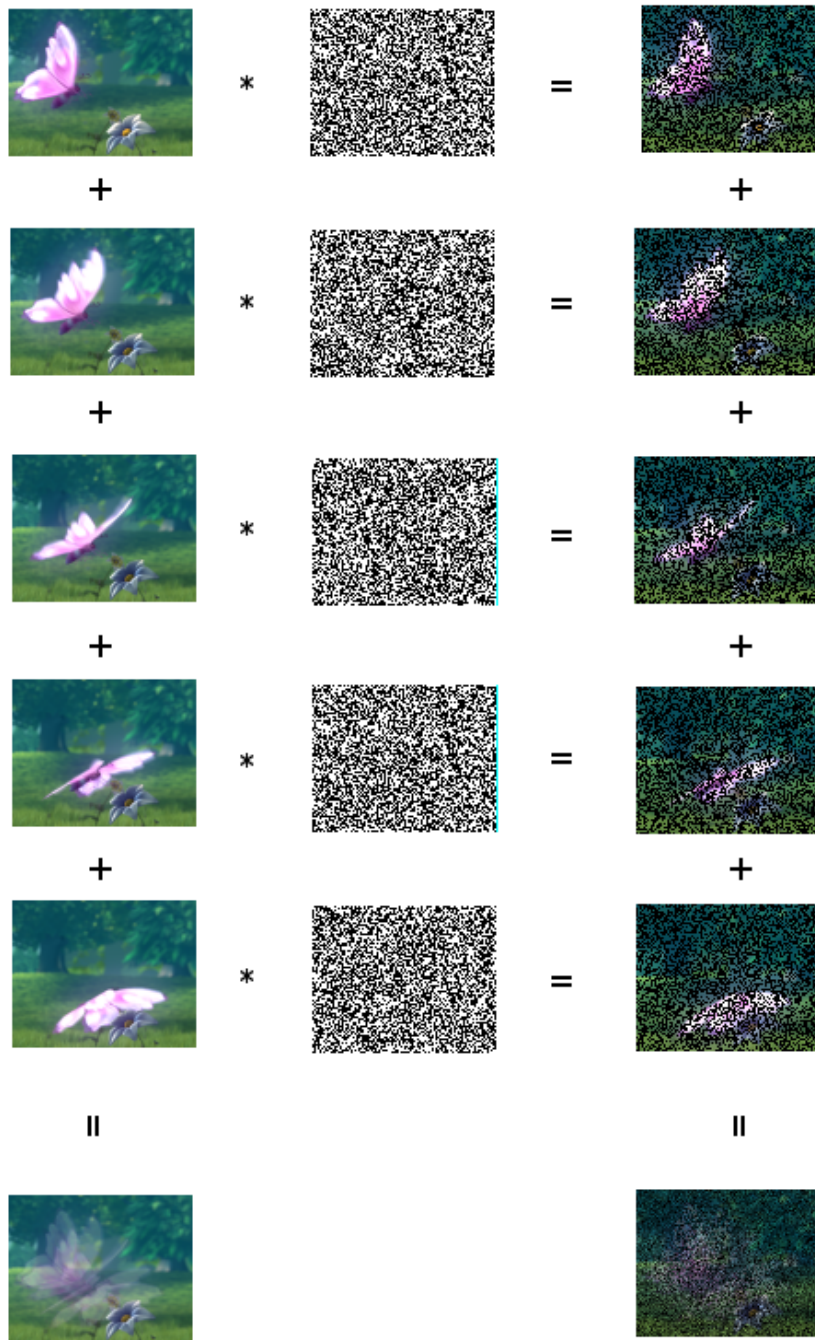


Figura 2.14. Aquisição pelo método PPS2. A primeira coluna corresponde à aquisição tradicional, onde todos os sub-quadros são considerados para formar a imagem. A segunda coluna ilustra o padrão de pixels ‘ligados’ para cada sub-quadro, onde cada pixel é intercalado com os pixels de mesma posição em alguns sub-quadros distintos. Na terceira coluna, ilustramos a aplicação do padrão à luz da cena. Ao final do tempo de integração, obtemos uma combinação linear dos sub-quadros da cena original operados com os padrões correspondentes.

Capítulo 3

Método de Aquisição *Per-Pixel* *Mirror-Based Measuring*

No capítulo anterior, apresentamos os principais métodos de aquisição de vídeos, detalhando suas vantagens e desvantagens. Em particular, discutimos os desafios no que se diz respeito a quantidade de informação luminosa adquirida e à informação conflituosa entre sub-quadros que pode comprometer a reconstrução da cena.

Neste capítulo, apresentamos uma nova proposta para aquisição de medidas lineares de vídeos. Em seguida, comparamos as características do método proposto com os métodos apresentados no capítulo anterior. Apresentamos também um modelo de reconstrução para os métodos de aquisição, com o objetivo de realizar uma comparação justa. E, por fim, são apresentados detalhes das implementações dos testes.

3.1 *Per-Pixel Mirror-Based Measuring*

Os modos de aquisição apresentados no capítulo anterior sofrem de dois problemas. O primeiro é que muita da informação luminosa é perdida. O segundo é que, exceto na primeira forma do modulador por pixel, ao final do tempo de exposição os pixels contém informação visual de vários sub-quadros.

Por causa disso, precisamos de uma nova forma de adquirir medidas lineares no domínio espaço-temporal que não desperdice informação luminosa e que não combine a informação temporal entre sub-quadros distintos. Apresentamos, neste trabalho, um modelo de aquisição denominado *per-pixel mirror-based measuring* (PPM).

A primeira etapa do nosso trabalho é a proposta de um novo dispositivo chamado *high-precision digital micromirror device* (HP-DMD). Este dispositivo seria uma versão com melhor precisão do DMD, em que cada micro espelho teria a precisão angular suficiente para redirecionar a luz recebida para qualquer outro pixel. Além disso, o HP-DMD poderia

mudar o seu ângulo mais rápido que o FPS da câmera, assim como o DMD o faz. Note que tal dispositivo, além de oferecer muitas outras opções de comportamento do fluxo da luz entre os pixels, ele também pode simular o comportamento do DMD e, assim, pode funcionar para todos os métodos de aquisição apresentados.

Seja N o número de pixels da câmera e k o número de sub-quadros. No método PPM, em cada sub-quadro, escolhemos N/k pixels para receber a luz dos N pixels correspondentes ao sub-quadro atual. Para cada um dos N/k pixels, k espelhos aleatórios são direcionados para esses pixels. No sub-quadro seguinte, outros N/k pixels são escolhidos e os espelhos são redirecionados para estes pixels. No fim dos k sub-quadros, cada um dos N pixels receberão luz de k espelhos. O método proposto é ilustrado na Figura 3.1. Em termos da operação matemática correspondente, esse processo implica que cada um dos pixels receberá a soma do valor de intensidade luminosa de k pixels aleatórios.

Ao final do tempo de integração desse processo de aquisição, nenhuma luz é descartada, pois toda a luz é redirecionada para outro pixel. Nenhuma luz é redirecionada para fora dos pixels e, assim, toda luz é absorvida pelos pixels. Além disso, toda a luz de um único sub-quadro é refletida em um único conjunto de N/k pixels. Isso significa que a luz de cada sub-quadro não se mistura à luz de outro sub-quadro, ou seja, não há dependência temporal entre as medidas, embora ocorra dependência espacial, diferentemente das outras formas de aquisição. Observe que essa forma de aquisição é semelhante com a forma de aquisição apresentada na SPC, na qual jogamos a luz de 50% dos pixels para um único sensor e descartamos o restante da luz proveniente. Se a luz descartada for direcionada para um outro sensor, obteríamos um resultado equivalente a configuração de um sub-quadro da proposta de aquisição proposta, considerando $k = N/2$.

O PPM pode receber algumas variações em sua configuração. Cada pixel de cada quadro pode receber a luz de k espelhos. Uma primeira é uma configuração tal que, para cada sub-quadro, o mesmo pixel recebe a luz dos mesmos espelhos aleatórios. Chamaremos essa configuração de PPM1, conforme ilustrado na Figura 3.1.

Em uma segunda configuração, podemos escolher pixels aleatórios diferentes para cada pixel em cada sub-quadro diferente. Chamaremos esta configuração de PPM2, ilustrada na Figura 3.2. Note que o método do PPM1 (ver a primeira coluna da Figura 3.1) apresenta o mesmo padrão de espelhos em todos os sub-quadros. No método PPM2, cada sub-quadro apresenta o próprio padrão aleatório de espelhos (ver a primeira coluna da Figura 3.2). Desta forma, tal configuração pode ser mais interessante que o PPM1 para, pois pode satisfazer melhor as propriedades de incoerência.

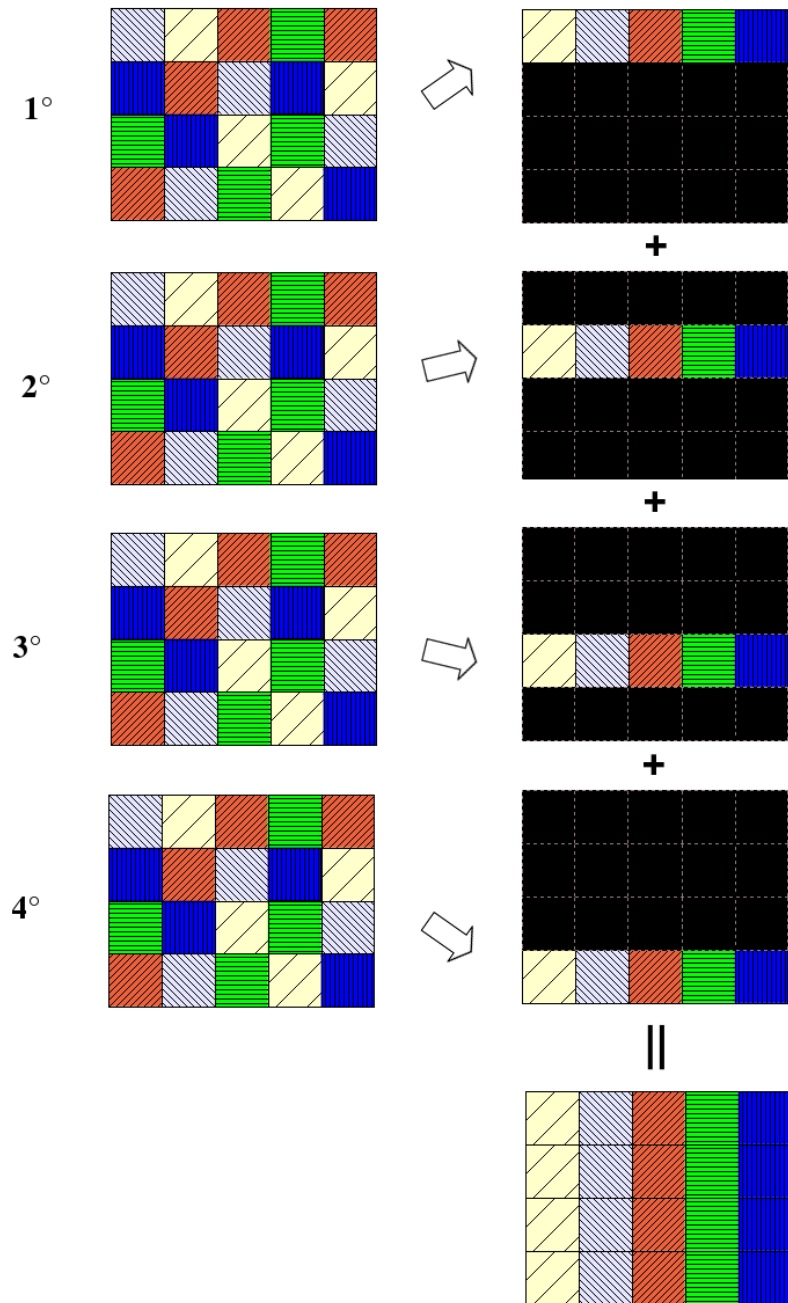


Figura 3.1. Método de aquisição proposto em sua primeira (PPM1) para $N = 80$, e $M = 20$ e $k = 4$. Para cada sub-quadro, apenas uma pequena faixa de pixels ($M/k = 5$ pixels) recebe luz. O padrão de espelhos é o mesmo para todos os sub-quadros.

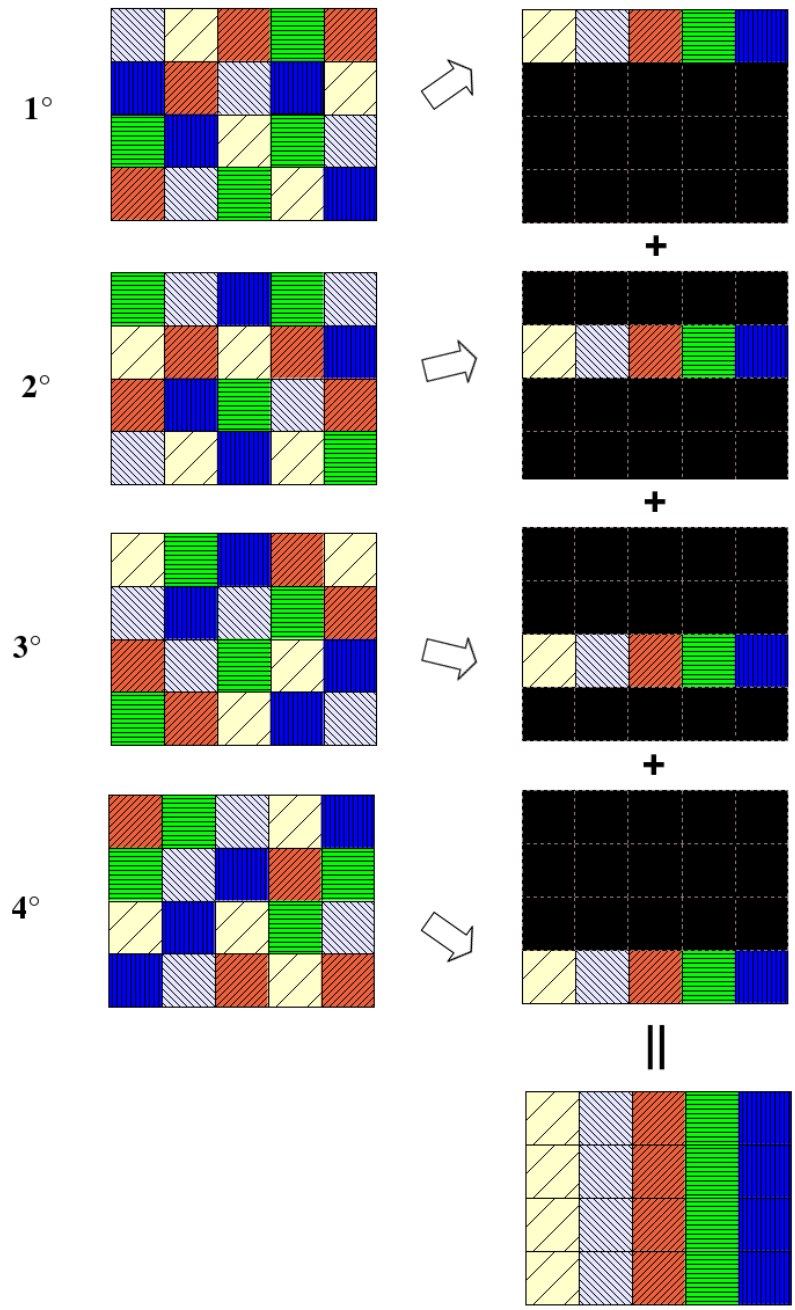


Figura 3.2. Método de aquisição PPM, na configuração PPM2 para $N = 80$, e $M = 20$ e $k = 4$. Para cada sub-quadro, apenas uma pequena faixa de pixels ($M/k = 5$ pixels) recebe luz. O padrão de espelhos é diferente para cada sub-quadro.

3.2 Comparação Entre os Métodos de Aquisição

3.2.1 Propriedades Gerais

Podemos classificar os métodos de aquisição apresentadas nesse trabalho como viáveis ou inviáveis de serem implementadas em uma câmera digital. Dentre os métodos viáveis, algumas já são implementadas naturalmente nas câmeras, enquanto outras precisam de um dispositivo extra externo já disponível. A nossa proposta (PPM) necessita de um dispositivo que ainda não existe. Na Tabela 3.1 são apresentadas informações sobre viabilidade e o tipo de dependência entre as medidas.

Tabela 3.1. Métodos de aquisição e suas vantagens e desvantagens.

Aquisição	Viabilidade em câmera de vídeo	Dependência
Tradicional	Viável em todas as câmeras	temporal
FS	Viável em algumas câmeras	temporal
PPS1	Requer dispositivo DMD	nenhuma
PPS2	Requer dispositivo DMD	temporal
PPM1	Requer dispositivo HPDMD	espacial
PPM2	Requer dispositivo HPDMD	espacial

Nível de Luz e Pré-visualização

As Figuras 2.11, 2.13 e 2.14, que representam respectivamente os métodos FS, PPS1 e PPS2, não correspondem à real quantidade de luz adquirida. O nível de intensidade de luz recebida em cada figura foi alterado para melhor visualização. A Figura 3.3 ilustra melhor o nível de intensidade de luz adquirida em cada sub-quadro de todas os métodos. Ela também mostra, na última coluna, as medidas obtidas e uma possível pré-visualização dos quadros a partir delas.

Podemos observar na primeira linha a aquisição de uma câmera de alta velocidade, a qual contém sensores rápidos e sensíveis. Cada um dos sub-quadros é adquirido (note que assim requer maior largura de banda). As figuras apresentam intensidade de luz significativa e representam bem o movimento. Nas demais linhas, os sensores são comuns. A sexta coluna indica o quadro resultante ao final do tempo de integração, o que é realmente obtido pelos sensores, e compõe as medidas dos métodos de aquisição. A segunda linha (Figura 3.3(b)) representa o modo de aquisição natural nas câmeras. Como o tempo de exposição de cada sub-quadro é pequeno e os sensores são comuns, então pouca luz é adquirida em cada um deles. Porém, ao final do tempo de integração, a imagem obtida contém uma boa intensidade luminosa pois contém a soma de todas os cinco sub-quadros. Também podemos observar na segunda linha o efeito *motion blur* ao final do tempo de

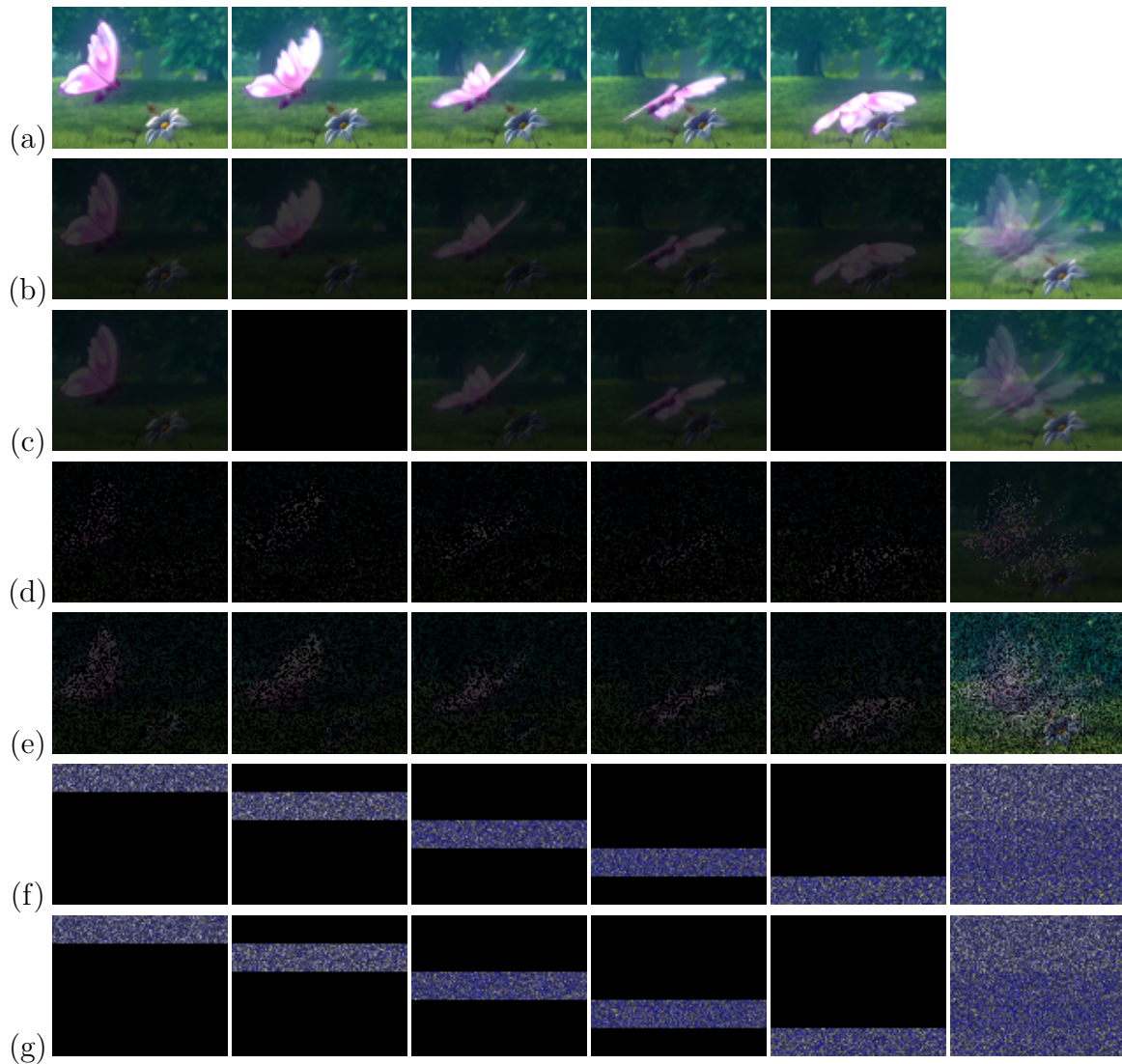


Figura 3.3. Comparativo de intensidade de luz para cada sub-quadro de todas os métodos. (a) representa a aquisição por sensores sensíveis e rápidos em uma câmera de alta velocidade; (b) representa aquisição em uma câmera normal; (c) mostra o método FS; (d) ilustra o PPS1; (e) representa o PPS2, (f) representa o PPM1 e (g) ilustra o PPM2. A última coluna representa a imagem obtida ao final do tempo de integração, que equivale às medidas para cada método e pode ser vista como uma pré-visualização do conjunto de quadros obtidos ao final da reconstrução.

integração nas partes em movimento. Na terceira linha (Figura 3.3(c)) é ilustrado o modo de aquisição FS. A quantidade de luz obtida é proporcional ao número de sub-quadros escolhidos. Neste caso, 60%, pois 3 dos 5 sub-quadros foram escolhidos. A quarta linha (Figura 3.3(d)) ilustra o PPS1. No capítulo anterior, a Figura 2.13 foi ajustada para melhor visualização do processo de aquisição, porém não ilustra bem a quantidade de luz que é adquirida. A real quantidade de luz obtida é bem pequena (nesse caso, 20%). A pré-visualização apresenta *motion blur* com padrões aleatórios nas partes que contém mo-

vimento, porém, as partes imóveis estão muito bem visíveis. A quinta linha (Figura 3.3(e)) ilustra o PPS2. Novamente, a Figura 2.14 não ilustra muito bem a quantidade de luz adquirida para o método PPS2. De cada sub-quadro, apenas metade da luz é adquirida, razão que se mantém no quadro final obtido pelos sensores. Logo, a pré-visualização fica muito comprometida, mesmo nas partes sem nenhum movimento. A sexta linha (Figura 3.3(f)) ilustra o PPM1, enquanto que a sétima (Figura 3.3(g)) ilustra o PPM2. Nesses casos, toda a luz é concentrada em apenas uma porção dos pixels em cada sub-quadro. Como isso acontece para todos os sub-quadros e esses pixels não se interceptam, toda luz é obtida. Porém a pré-visualização é muito ruidosa (por causa da dependência espacial das medidas) e não fornece nenhuma informação visual.

Compensação do Ruído

Uma forma de corrigir a intensidade de luz adquirida pelos sensores é ajustando o ISO da câmera, como descrito na Seção 2.3.2. Para a aquisição de cenas com pouca luz, podemos aumentar o ISO da câmera para obter uma leitura de luz maior nos sensores. Como a luz é de fato captada apenas ao final do tempo de aquisição, o ruído de medida dos sensores é somado apenas no quadro formado pelas medidas, assim, quanto maior o ISO, maior a quantidade de ruído. A Tabela 3.2 mostra uma estimativa da quantidade de luz e a proporção de ruído gerado ao aumentar o ISO de tal forma que a quantidade de luz seja equivalente entre os métodos, que é dada pelo inverso da quantidade de luz adquirida pelos métodos de aquisição. Observe que a eficiência luminosa do método proposto propicia um nível de ruído menor do que a das outros métodos.

Tabela 3.2. Percentagem de informação luminosa adquirida em cada método e quantidade de ruído adquirido ao alterar o ISO para obter a mesma quantidade de luz que o método tradicional.

Método de Aquisição	Informação luminosa	Ruído para equivalência de luz
Tradicional	100%	1×
FS	50% a 67%	1,5× a 2×
PPS1	100/ k %	k ×
PPS2	aprox. 50%	2×
PPM1	100%	1×
PPM2	100%	1×

3.2.2 Incoerência

Outro parâmetro que deve ser considerado quando comparamos os diversos métodos é suas matrizes de medidas satisfazerem as propriedades de incoerência. Na seção 2.2.4 do capítulo anterior, vimos que é possível medir a incoerência de uma matriz de medidas utilizando a função TPSF. Para isso vamos comparar a TPSF do método PPM com a TPSF do método PPS1, que é equivalente, em termos de operações matemáticas, a fazer sub-amostragem aleatória, técnica que é amplamente utilizada nas aplicações de *compressive sensing*. Obter a TPSF de uma matriz de medidas é custoso computacionalmente (mesmo com as otimizações descritas na Seção 3.5). Por isso, fazemos comparações para sinais em domínio 1D de tamanho 100, e para sinais de domínio 2D de tamanho 32×32 . Por limitações computacionais, não realizamos os testes com sinais de domínio 3D. Logo, os métodos de amostragem exclusivos para reconstrução 3D (PPS2 e FS) não puderam ser implementados. Os métodos PPS2 e FS não podem ser implementados para um só quadro pois extrapolam o número de medidas dada pela taxa de sub-amostragem. No caso PPS2, se obedecesse o número de medidas, ele seria igual ao PPS1, e também sofreria da baixa eficiência luminosa. No caso FS, se o quadro estivesse ligado, teríamos 100% das medidas, e se estivesse desligado, seriam zero medidas, ou 100% de medidas com energia igual a zero, ou seja, impossível de obter a quantidade de medidas diferente disso.

Fizemos a comparação do TPSF para a matriz de medidas (Φ) correspondente aos métodos PPM e PPS1, e para os domínios transformados (Ψ) como sendo a DCT (1D e 2D), o operador de diferenças finitas (TV), e a identidade. O fator de sub-amostragem é de $4\times$. Os resultados podem ser visualizados nas Figuras 3.4 e 3.5.

O comportamento que melhor satisfaz a propriedade de incoerência é aquele em que a diagonal principal, a energia da TPSF é alta se comparada a energia da TPSF nos outros pontos. Além disso, o comportamento do sinal fora da diagonal principal é ruidoso e de baixa energia. Sendo assim, podemos observar que para o PPM com ψ igual a DCT, a TPSF satisfaz esse comportamento. Para o PPS1 com ψ igual a DCT, o comportamento é parecido, apresentando grande energia na diagonal principal e baixa fora dela. Porém, o comportamento da fora da diagonal sugere algo mais previsível que um ruído.

As imagens centrais (Figuras 3.4 e 3.5, (c) e (d)) apresentam o Ψ igual ao operador de diferenças finitas. Apesar de observarmos o comportamento em que a diagonal principal contém um sinal com energia, fora dela a energia ainda é muito alta. Por este comportamento, conclui-se que esse sistema de *compressive sensing* não contém uma alta incoerência. Acreditamos que ao aumentar o tamanho do sinal, a incoerência também deve aumentar. Sustentamos esse argumento nos resultados para reconstrução com vídeos de tamanho maior, que ficaram melhores, como pode ser visto no próximo capítulo.

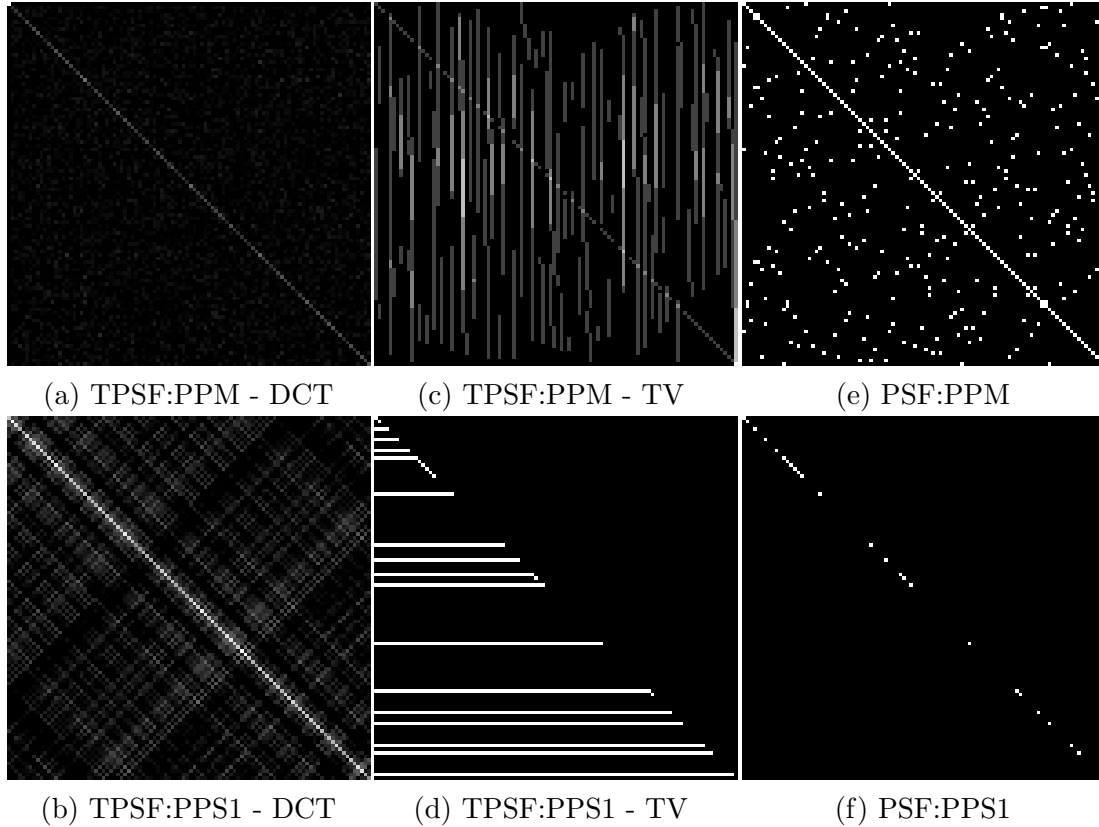


Figura 3.4. Resultado das funções TPSF e PSF para um sinal de domínio 1D de dimensões 1×100 , e taxa de subamostragem de $4\times$.

3.3 Outras Variações do PPM

3.3.1 PPM em Blocos

Embora usamos apenas as configurações PPM1 e PPM2 como modo de aquisição neste trabalho, é possível aplicar o método proposto de formas diferentes. Uma delas é escolher os espelhos a partir de posições específicas. Por exemplo, escolher apenas os espelhos na mesma linha do pixel que está recebendo a luz. Outra abordagem seria escolher apenas os espelhos de uma região próxima (um bloco quadrado por exemplo) ao pixel que está o recebendo a luz. Tal abordagem poderia diminuir a dependência espacial entre as medidas para pixels que estão mais próximos ou pixels em uma mesma linha. Outra vantagem seria na pré-visualização, na qual poderíamos visualizar uma versão em baixa resolução espacial do vídeo a ser reconstruído utilizando apenas as medidas capturadas. Uma das desvantagem destas alterações diz respeito a incoerência. Sem a limitação de espelhar os pixels por toda a imagem, a incoerência pode ser comprometida.

Um exemplo de uma possível pré-visualização a partir das medidas do PPM por blocos é ilustrado na Figura 3.6. Neste caso, as medidas foram tomadas em blocos de tamanho

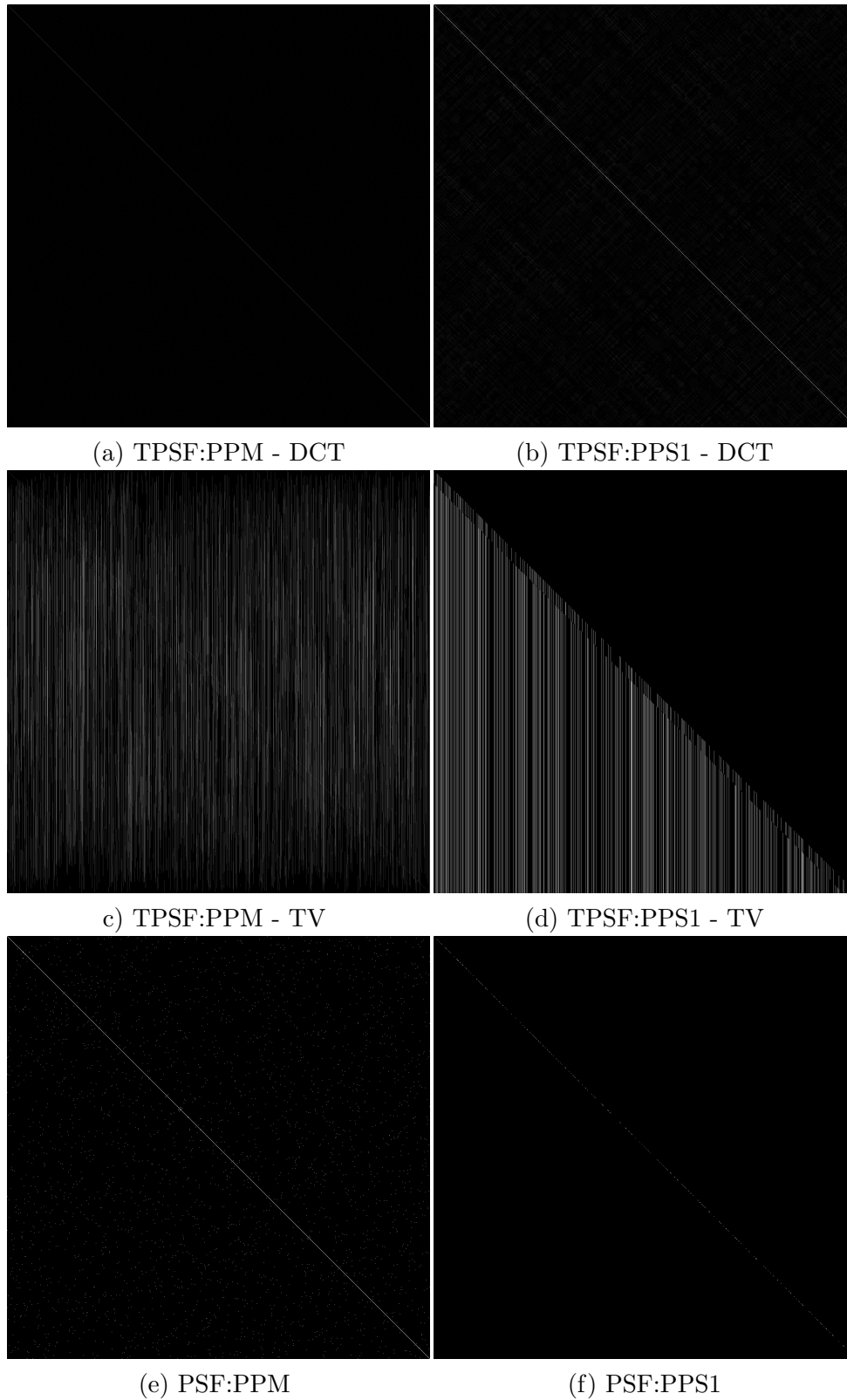


Figura 3.5. Resultado das funções TPSF e PSF para um sinal 2D de dimensões 32×32 , e taxa de subamostragem de $4\times$.

25 (5×5), ou seja, apenas os pixels em uma vizinhança de 5×5 são escolhidos para serem integrados conjuntamente para formar as medidas. A partir das medidas, cada pixel na Figura 3.6 foi formado por 5 pixels consecutivos. Obtemos uma pré-visualização com 5 quadros de alta velocidade, mas com a resolução espacial 25 vezes menor.



Figura 3.6. Quadros gerados pela pré-visualização a partir das medidas PPM por blocos 5×5 .

Neste trabalho, focamos na análise do método nas configurações PPM1 e PPM2. Porém, para avaliar possíveis ganhos na pré-visualização, mostramos também alguns testes com o PPM por blocos. Avaliamos também se esta forma compromete a qualidade de reconstrução.

3.3.2 PPM para Super-resolução

Super-resolução é o processo em que, a partir de uma ou mais imagens de menor resolução, obtemos uma imagem de resolução maior. No contexto de aquisição e reconstrução de imagens em *compressive sensing*, o processo de super-resolução consiste em adquirir uma quantidade menor de pixels, e reconstruir a imagem de alta resolução [42]. Neste sentido, o método de aquisição PPM também pode ser utilizado para super-resolução.

Suponha que temos uma câmera com resolução igual a M pixels e, temos como objetivo, tirar uma única foto. Suponha que introduzíssemos o dispositivo HPDMD na câmera e aplicássemos o método de amostragem PPM para um quadro, utilizando uma taxa de sub-amostragem de fator k . Ao aplicar o método PPM, apenas M/k amostras seriam adquiridas. O processo de reconstrução recuperaria uma imagem de resolução M a partir de M/k medidas. O que ainda não seria super-resolução, pois a imagem resultante teria a mesma resolução da imagem obtida pela câmera.

Agora, suponha que o dispositivo HPDMD tenha um número muito maior de espelhos, por exemplo N , tal que $N = Mk$. Ao realizarmos o mesmo processo de aquisição com esses novos parâmetros, obtemos M amostras, ou seja, o número correspondente à resolução da câmera. Mas, a partir do processo de reconstrução, obtemos uma imagem de tamanho N , k vezes o tamanho da resolução da câmera. O processo está ilustrado na Figura 3.7.

Como esse método não recupera vídeos de alta velocidade, não é preciso dividir o tempo de integração em sub-quadros. O conjunto de pixels que recebem a luz é sempre o mesmo (que são todos os sensores da câmera), pois não há a necessidade de dividir as

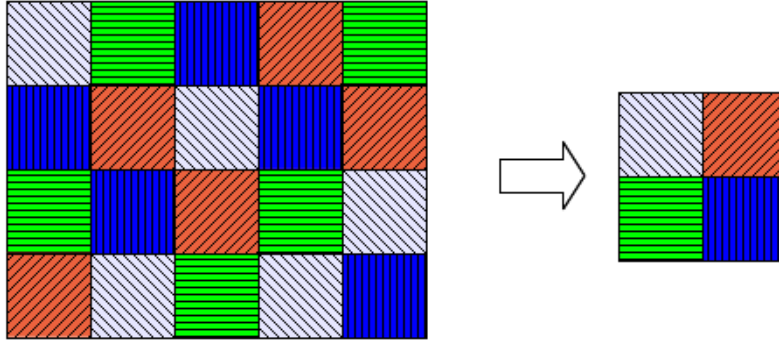


Figura 3.7. Método de aquisição PPM para super-resolução para uma câmera com $M = 4$ pixels, e um HPDMD com $N = 20$ espelhos.

medidas temporalmente. O padrão de espelhos pode ser único também, pois o objetivo é reconstruir imagens. Nesse caso, não precisamos mover o conjunto de espelhos para cada sub-quadro. Assim eles ficam numa posição fixa e isso pode trazer diversas vantagens na construção do dispositivo.

No caso de reconstrução de vídeo de alta velocidade, temos a liberdade de escolher a taxa de sub-amostragem arbitrária no processo de aquisição. A única consequência dessa escolha é que se a taxa for muito grande, a qualidade da reconstrução não fica satisfatória, pois há poucas medidas para o grande número de quadros a serem reconstruídos. Para super-resolução existe uma limitação quanto a taxa de sub-amostragem. O número de espelhos no HPDMD determina o tamanho da imagem reconstruída. Se considerarmos que o número de espelhos no HPDMD e a resolução espacial da câmera são fixos, a taxa de sub-amostragem também será fixa. Ou seja, perdemos toda a liberdade de escolher a taxa de sub-amostragem livremente. É importante salientar que o método PPM oferece um compromisso entre as resoluções espacial e temporal. Podemos escolher aumentar a resolução temporal de um vídeo, ou aumentar a resolução espacial na mesma taxa, mas não os dois ao mesmo tempo. Inicialmente, o foco deste trabalho não foi utilizar o PPM para super-resolução, e sim para captura de vídeo para reconstrução em alta resolução temporal. Por isso, não comparamos os outros métodos de super-resolução a partir de *compressive sensing* [59, 18, 47, 1] com o método de super-resolução pelo PPM aqui proposto. Mas de fato, em uma das simulações que realizamos neste trabalho (Seção 4.1), parte do processo de reconstrução de vídeos de alta resolução temporal inclui as mesmas simulações que poderiam ser feitas para reconstrução de imagens em super-resolução. Então, este trabalho também mostra um potencial para super-resolução.

3.4 Modelo de Reconstrução dos Vídeos

Após adquirir as medidas, o próximo passo é reconstruir o vídeo de alta velocidade. Atualmente, existem algumas formas de reconstruir as cenas apenas com as medidas lineares no domínio do tempo. Nessa sessão, apresentamos os detalhes de uma técnica de reconstrução de imagens e vídeos que utilizamos nos nossos testes, relacionando-as com as formas de aquisições citadas anteriormente neste trabalho.

Na Sessão 2.2.5 apresentamos a ideia de algoritmo de reconstrução baseado na minimização da Variação total (*Total Variation* - TV [32, 50]), utilizando o problema de otimização dado pela equação:

$$\hat{s} = \operatorname{argmin}_{s'} (\|s'\|_{TV}) \text{ sujeito a } y = \Phi\Psi s'. \quad (3.1)$$

Nesta equação, Φ é a matriz que representa o método de aquisição e Ψ é a base transformada. Nesse trabalho, tomamos as medidas no domínio espaço-temporal, ou seja, o mesmo domínio dos pixels. Se considerarmos isto, Ψ é a identidade. Considerando que $\|s\|_{TV} = \|D_i s\|_1$ e fazendo a substituição $s'' = D_i s'$, 3.2 é equivalente ao seguinte problema de otimização:

$$\hat{s} = \operatorname{argmin}_{s''} (\|s''\|_1) \text{ sujeito a } y = \Phi D_i^{-1} s''. \quad (3.2)$$

Então, nesse caso, esse método é equivalente à minimização ℓ_1 , onde a transformada esparsificante é o operador de diferenças finitas invertido.

Φ é a operação linear correspondente ao modo de aquisição. Para os testes desse trabalho, utilizaremos os diversos modos de aquisição mostrados no Capítulo 2, e o modo de aquisição proposto neste capítulo. Com o modo de reconstrução fixo, e variando apenas o modo de aquisição, podemos avaliar corretamente a eficácia do modo de aquisição proposto.

Sobre a aquisição de um vídeo, se as medidas de um dado modo de aquisição forem independentes temporalmente, é possível reconstruir cada sub-quadro separadamente. E também aplicar o método para imagens. Os métodos que entram nessa categoria são o PPS1 e o PPM. Então, primeiramente realizamos os testes com imagens, comparando os métodos para os quais a reconstrução é a minimização TV2D. Note que desse modo não é possível tirar proveito das redundâncias temporais.

Em um outro caso, quando as medidas forem temporalmente dependentes (como em FS, PPS2), também podemos utilizar esses métodos para reconstrução TV2D dos conjuntos de sub-quadros de teste. Em vídeos, podemos testar todos os métodos de aquisição apresentados neste trabalho. Podemos aplicar a reconstrução TV2D em cada linha ou

coluna através do tempo, tratando independentemente cada linha ou coluna, mas levando em consideração as redundâncias temporais [26]. A outra opção é a reconstrução 3D do sinal, que leva em consideração as redundâncias espaciais e temporais dos sub-quadros do vídeo como um todo. Entre os métodos de aquisição/reconstrução, com fins de comparação, adicionamos também, a aquisição sub-amostrada com medidas igualmente espaçadas e a reconstrução por interpolação.

A Tabela 3.3 mostra a compatibilidade entre os diversos modos de aquisição e de reconstrução.

Tabela 3.3. Métodos de aquisição e reconstrução compatíveis.

Aquisição/Reconstrução	Interpolação	TV2D: espacial	TV2D: linhas por tempo	TV3D
Sub-amostragem	✓	✓	✓	✓
FS	✓	×	✓	✓
PPS1	✓	✓	✓	✓
PPS2	✓	×	✓	✓
PPM	✓	✓	✓	✓

Para avaliar os resultados, comparamos a qualidade da reconstrução de cada modo de aquisição utilizando a reconstrução compatível. Como medida de desempenho, utilizamos a medida de relação sinal-erro (*signal-to-error ratio* - SER) entre a imagem original e cada uma das imagens obtidas. O SER é uma medida de similaridade entre dois sinais dada por um número. Quanto maior o SER, mais similar são os dois sinais considerados. Se representarmos o sinal original por sig_1 e o sinal reconstruído por sig_2 , a fórmula para o cálculo do SER é

$$SER = 20 \log_{10} \left(\frac{\|sig_1\|_2}{\|sig_2 - sig_1\|_2} \right). \quad (3.3)$$

Note que essa relação difere da relação sinal-ruído (*signal-noise ratio* - SNR) apenas na nomenclatura, pois ao invés de medir a energia no ruído, estamos medindo o erro da reconstrução do sinal em relação a um sinal original.

3.5 Implementação

3.5.1 O algoritmo TVAL3

Para nossas simulações, utilizamos o algoritmo de reconstrução por minimização da *total variation* desenvolvido por Chenbo Li, conhecido como *TV minimization by Augmented Lagrangian and ALternating direction ALgorithms*(TVAL3) [32, 30]. Sua implementação para vetores 1D e 2D está disponível online [29].

Em nossos testes, fazemos também a reconstrução 3D. Para tal, utilizamos a implementação 3D do TVAL3 fornecida pelo mesmo autor [31]. O algoritmo tem como entradas as medidas, o tamanho do sinal a ser reconstruído, parâmetros de opções (tais como, número máximo de interações, tolerância, mais restrições, a solução inicial, etc) e o operador linear que toma as medidas. A solução inicial utilizada pelo algoritmo não é necessariamente a de menor energia. O algoritmo utiliza por padrão a solução inicial dada pela matriz de medidas transposta vezes as medidas, mas ele também dá a opção para o usuário fornecer uma outra solução inicial. O operador linear que simula o processo de aquisição (Φ) pode ser tanto a matriz de medidas, quanto um *function handle* para uma função que opera um vetor da mesma forma que a matriz de medidas. O *function handle* esperado recebe o vetor de entrada e um inteiro que indica o modo. O modo 1 aplica o operador no vetor de entrada, e o modo 2 aplica o operador na forma transposta.

3.5.2 A função que Simula o HPDMD e a Matriz de Índices

Para os modos de aquisição, foi criada uma função chamada HPDMD, que simula o comportamento do dispositivo HPDMD. Assim, com ela é possível implementar todos os modos de aquisição mostrados neste trabalho, apenas mudando a matriz de medidas. O algoritmo 1 implementa a função correspondente ao HPDMD.

A matriz de medidas que simula o comportamento do HPDMD contém ‘1’ nas posições dos pixels escolhidos para receberem a luz do espelho correspondente aquela linha, e ‘0’ nas demais posições. Como cada espelho só pode estar apontado para um único pixel, só existe um valor ‘1’ por coluna.

A matriz mtx define o modo de aquisição que deve ser equivalente à multiplicação da matriz de medidas pelo vetor x . Cada elemento da matriz é um índice correspondente a uma posição do vetor, ou seja, os elementos da matriz mtx são números naturais limitados de 1 até tamanho de x .

No modo 1, para cada linha de mtx , a função seleciona a posição do vetor x correspondente aos índices dessa linha e soma, gerando uma medida. E isso acontece para todas as linhas de mtx . No modo 2 é aplicada a correspondente à matriz de medidas transposta. Para cada linha da matriz de medidas transposta, escolhemos a amostra correspondente ao índice no vetor de entrada para compor cada elemento da saída.

Por definição, os espelhos do HPDMD não podem mirar sua luz para dois pixels ao mesmo tempo, então, por isso, é adicionada a limitação de que para cada conjunto de medidas no mesmo tempo (por exemplo, as primeiras M/k medidas correspondentes ao primeiro sub-quadro), não se tenham índices repetidos. Na prática, durante um conjunto de k sub-quadros, o mesmo pixel em sub-quadros diferentes são indexados por números diferentes (o i -ésimo pixel no j -ésimo sub-quadro é indexado por $i + jN$). Então o teste

```

Data: x, mtx, mode, N, M
Result: y
início;
[h,w]=size(mtx);
switch mode do
    case 1
        q=size(mtx());
        for i=1:M do
            for j=1:w do
                | y(i)=y(i)+x(mtx(i,j));
            end
        end
    end
    case 2
        for i=1:h do
            for i=1:w do
                | y(mtx(i,j))=x(i);
            end
        end
    end
end

```

Algorithm 1: Função que implementa o HPDMD. x é o vetor a ser operado, mtx é a matriz de índices em que as linhas e colunas contém números naturais não repetidos. $mode$ é o modo de operação, 1 é o modo direto, 2 é o modo transposto, N é o tamanho do sinal a ser reconstruído, M é o número de medidas. No modo 1, y recebe as medidas a partir da soma das linhas de mtx . No modo 2, y recebe os valores de x nas posições dadas por mtx .

de validação necessita que a matriz mtx não contenha nenhum valor repetido. E tal verificação foi adicionada ao código.

Vamos exemplificar como a implementação simula dois modos de aquisição, PPS1 e PPM, para um vetor de uma dimensão. No caso que queremos obter 3 medidas de um vetor de 6 posições ($N = 6$ e $M = 3$). Por exemplo $v = [2, 3, 5, 7, 11, 13]$. Pelo modo PPS1 (equivalente a escolher amostras aleatórias), se queremos escolher as amostras 3, 6 e 1 (que é o caso em que apenas os espelhos 3, 6 e 1 estão ligados), a matriz de medidas

e o processo de aquisição é dado por:

$$y = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \\ 11 \\ 13 \end{bmatrix} = \begin{bmatrix} 5 \\ 13 \\ 2 \end{bmatrix}. \quad (3.4)$$

Note que a matriz de medidas tem um tamanho de 3×6 . Porém, através da função HPDMD, não precisamos armazenar todo esse espaço. Basta apenas passar a mtx como a matriz coluna

$$m1 = \begin{bmatrix} 3 \\ 6 \\ 1 \end{bmatrix}, \quad (3.5)$$

PMue é bem menor que a matriz de medidas. Essa operação é realizada chamando $y = HPDMD(v, m1, 1, N, M)$. Desta forma, economizamos bastante espaço e tempo.

A operação transposta, tendo como entrada a saída da operação anterior, seria:

$$\hat{v} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 5 \\ 13 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 5 \\ 0 \\ 0 \\ 13 \end{bmatrix}. \quad (3.6)$$

Esta operação é equivalente ao comando $\hat{v} = HPDMD(y, m1, 2, N, M)$. A mtx ainda é a mesma, ou seja, ainda há economia de espaço e tempo de execução. O vetor \hat{v} gerado por esse processo mantém as amostras escolhidas, porém perde os valores não escolhidos.

Para o método PPM, se quisermos tomar 3 medidas deste mesmo vetor v , em que cada par de espelhos (por exemplo 4 e 6; 5 e 2; e 1 e 3) está apontado para um pixel diferente,

gerando 3 medidas, o processo de aquisição é expresso pelas seguintes equações:

$$y = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \\ 11 \\ 13 \end{bmatrix} = \begin{bmatrix} 7 + 13 \\ 3 + 11 \\ 2 + 5 \end{bmatrix} = \begin{bmatrix} 20 \\ 14 \\ 7 \end{bmatrix}. \quad (3.7)$$

Porém, para simular esse processo de aquisição, basta apenas que mtx seja igual a:

$$m2 = \begin{bmatrix} 4 & 6 \\ 5 & 2 \\ 1 & 3 \end{bmatrix}. \quad (3.8)$$

Desta forma, o tamanho da matriz passada também diminui muito em relação a matriz de medidas. O processo é chamado pelo comando $y = HPDMD(v, m2, 1, N, M)$.

No caso transposto, com y como entrada, a operação linear é igual a:

$$\hat{v} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 20 \\ 14 \\ 7 \end{bmatrix} = \begin{bmatrix} 7 \\ 14 \\ 7 \\ 20 \\ 14 \\ 20 \end{bmatrix}. \quad (3.9)$$

Esta expressão é equivalente a $\hat{v} = HPDMD(y, m2, 2, N, M)$. O resultado nesse caso preserva a soma do sinal, porém perde os valores individuais.

Em geral, considerando que N é o tamanho do sinal original, M o número de medidas, e $k = N/M$ a taxa de sub-amostragem, a matriz de índices mtx é de tamanho no máximo $M \times k$ (o número de colunas é no máximo k por causa da limitação de que cada espelho só pode estar apontado para um pixel de cada vez), ou seja, muito menor do que a matriz de medidas original, que seria de tamanho $M \times N$. Desta forma, temos uma redução de $\Theta(N)^2$ para $\Theta(N)$ em espaço de armazenamento. O tempo de execução também cai na mesma proporção. Isso é de suma importância, pois ao se trabalhar com vídeos grandes, a implementação pode ficar impraticável nos computadores atuais. Por exemplo, um vídeo HD de 16 quadros tem $N = 1280 \times 720 \times 16 = 14.745.600$ pixels. No processo de aquisição e reconstrução desse vídeo com taxa de sub-amostragem de 16, ou seja, com um $M=N/16=921.600$, a matriz de medidas teria tamanho $M \times N$, ou seja, a matriz teria 13.589.544.960.000 posições. Se cada entrada corresponder a um byte, precisaríamos de

aproximadamente 12,96 terabytes para armazenar a matriz de medidas. Se armazenar já é difícil, operar com ela é mais difícil ainda. Nesse caso, a matriz de índices (mtx) teria tamanho entre 900 kilobytes e 14 megabytes, um ganho considerável. No caso de 1 segundo de um vídeo de resolução 4k a 60 FPS, precisaríamos de aproximadamente 13,75 petabytes para armazenar a matriz de medidas, enquanto a matriz de índices teria tamanho entre 30 e 475 megabytes. São valores ainda praticáveis, mesmo para resolução dos vídeos perto do máximo da tecnologia atual.

3.5.3 Implementação dos Métodos de Amostragem com a HPDMD para Vídeos

Para a implementação da amostragem utilizando a função HPDMD para imagens e vídeos, consideramos a entrada 2D ou 3D como um vetor coluna. É necessário adaptar tanto a matriz de medidas quanto o modo de amostragem a isso.

Suponha, por exemplo, um vídeo de tamanho $2 \times 3 \times 3$. Os seus índices seriam representados de 1 a 18, onde os índices de 1 a 6 correspondem ao primeiro sub-quadro, os índices de 7 a 12 correspondem ao segundo, e os índices de 13 a 18 ao terceiro sub-quadro. Primeiramente, representamos na matriz primeiro na forma de soma, onde a soma com 0 representa o primeiro quadro, a soma com 6 representa o segundo quadro, e a soma com 12 representa o terceiro quadro. Mostramos como deve ser a matriz de índices para o processo de aquisição de um vídeo de tamanho $2 \times 3 \times 3$, com $N=18$ e $M=6$, ou seja, uma taxa de sub-amostragem de $3x$.

Para uma aquisição FS onde são escolhidos o primeiro e terceiro sub-quadro, o vídeo poderia ser adquirido pela função HPDMD com a matriz mtx igual a:

$$\begin{bmatrix} 0 + 1 & 12 + 1 \\ 0 + 2 & 12 + 2 \\ 0 + 3 & 12 + 3 \\ 0 + 4 & 12 + 4 \\ 0 + 5 & 12 + 5 \\ 0 + 6 & 12 + 6 \end{bmatrix} = \begin{bmatrix} 1 & 13 \\ 2 & 14 \\ 3 & 15 \\ 4 & 16 \\ 5 & 17 \\ 6 & 18 \end{bmatrix}. \quad (3.10)$$

Desta forma, o primeiro e terceiro sub-quadros inteiro são somados, exatamente como a forma de aquisição FS. O segundo sub-quadro é descartado, o que leva a uma eficiência luminosa de $2/3$.

Para o método de aquisição PPS1, a matriz de índices é:

$$\begin{bmatrix} 6 + 1 \\ 12 + 2 \\ 0 + 3 \\ 6 + 4 \\ 0 + 5 \\ 12 + 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 14 \\ 3 \\ 10 \\ 5 \\ 18 \end{bmatrix}. \quad (3.11)$$

O zero na soma significa que a amostra é tomada do primeiro quadro, o 6 significa que ela é tomada do segundo quadro, e o 12 significa que ela é tomada do terceiro. Assim, após o processo de aquisição, obtemos o terceiro e quinto pixel do primeiro quadro, o primeiro e quarto pixel do segundo quadro, e o segundo e sexto pixel do terceiro quadro. Desta forma, na soma final, temos apenas um pixel de cada sub-quadro, sendo que cada um destes pixels está em sua posição original dentro do quadro. Como no PPS1, 2/3 dos pixels originais são descartados.

Utilizando o método de aquisição PPS2 para aquisição de um vídeo deste tamanho, a matriz de índices é:

$$\begin{bmatrix} 0 + 1 & 12 + 1 \\ 6 + 2 & \\ 0 + 3 & 6 + 3 \\ 6 + 4 & 12 + 4 \\ 0 + 5 & \\ 12 + 6 & \end{bmatrix} = \begin{bmatrix} 1 & 13 \\ 8 & \\ 3 & 9 \\ 10 & 16 \\ 5 & \\ 18 & \end{bmatrix}. \quad (3.12)$$

Após o processo de aquisição, o primeiro pixel do resultado contém a soma do primeiro pixel do primeiro e terceiro sub-quadros, o segundo pixel do resultado contém o valor do segundo pixel do segundo quadro, o terceiro pixel contém a soma do terceiro pixel do primeiro e segundo sub-quadros, o quarto pixel contém a soma do último pixel do segundo e terceiro sub-quadros, o quinto pixel contém o quinto pixel do primeiro sub-quadro, e o sexto contém o pixel correspondente do último sub-quadro. Desta forma, simulamos o modo de aquisição PPS2 assim como ele é. Note que metade dos pixels originais são descartados.

O método PPM1 poderia ser representado pela matriz:

$$\begin{bmatrix} 0+2 & 0+6 & 0+5 \\ 0+3 & 0+1 & 0+4 \\ 6+2 & 6+6 & 6+5 \\ 6+3 & 6+1 & 6+4 \\ 12+2 & 12+6 & 12+5 \\ 12+3 & 12+1 & 12+4 \end{bmatrix} = \begin{bmatrix} 2 & 6 & 5 \\ 3 & 1 & 4 \\ 8 & 12 & 17 \\ 9 & 7 & 10 \\ 14 & 18 & 17 \\ 15 & 13 & 16 \end{bmatrix}. \quad (3.13)$$

Os dois primeiros pixels recebem 3 pixels distintos do primeiro sub-quadro, o terceiro e quarto pixels recebem as medidas dos pixels de mesma posição, porém do segundo sub-quadro. O mesmo acontece com os dois últimos pixels, porém recebem as medidas do último sub-quadro. Note que as medidas de cada um dos sub-quadros estão bem separadas das medidas dos outros sub-quadros. É possível obter as medidas correspondentes a apenas o primeiro sub-quadro separadamente das outras medidas. Porém, as medidas estão embaralhadas espacialmente.

Por fim, o PPM2 poderia ser representado por:

$$\begin{bmatrix} 0+2 & 0+6 & 0+5 \\ 0+3 & 0+1 & 0+4 \\ 6+3 & 6+5 & 6+2 \\ 6+6 & 6+4 & 6+1 \\ 12+4 & 12+1 & 12+2 \\ 12+6 & 12+3 & 12+5 \end{bmatrix} = \begin{bmatrix} 2 & 6 & 5 \\ 3 & 1 & 4 \\ 9 & 11 & 8 \\ 12 & 10 & 7 \\ 16 & 13 & 14 \\ 18 & 15 & 17 \end{bmatrix}. \quad (3.14)$$

A diferença entre o PPM2 e o PPM1 é que os pixels escolhidos de cada sub-quadro são diferentes. Vemos que tanto no PPM1 como no PPM2, todos os pixels são representados na matriz, então a eficiência luminosa é de 100%.

Após realizar o processo de aquisição, chamamos a função que reconstrói o sinal. Para isso, utilizamos a função TVAL3, para duas dimensões, e a função TVAL3_CVS_D3 para 3 dimensões.

Capítulo 4

Simulação e Testes

No capítulo anterior foi apresentado o método de aquisição PPM, assim como um modelo para reconstrução dos vídeos. Neste capítulo, apresentamos os testes do modelo de reconstrução para o método proposto e para os métodos disponíveis na literatura, conforme apresentadas na Seção 1.1. Os testes foram feitos para imagens, vídeos sintetizados e vídeos naturais. Testamos as variações de dimensões, robustez ao ruído, e comparamos os tempos de reconstrução.

4.1 Testes em Imagens

Inicialmente, escolhemos uma imagem de tamanho N em níveis de cinza e taxa de sub-amostragem k . Tomamos apenas N/k medidas a partir de cada um dos métodos de aquisição selecionados. Isso simula a aquisição de apenas um sub-quadro em um tempo de exposição de $1/k$ do tempo total de integração de cada quadro da câmera. A partir das aquisições PPS1 e PPM, reconstruímos por minimização da TV2D. Também utilizamos uma reconstrução por interpolação a partir de uma sub-amostragem igualmente espaçada.

Neste primeiro teste, utilizamos três imagens: a primeira de tamanho 64×64 ($N = 4.096$), a segunda de tamanho 256×256 ($N = 65.536$) e a terceira de tamanho 1024×1024 ($N = 1.048.576$). Utilizaremos $k = 4$ para as 3 imagens. Apenas variando o tamanho da imagem, podemos testar os resultados para diferentes resoluções de imagem.

Nas Figuras 4.1, 4.2, 4.3 e 4.4 são apresentadas imagens reconstruídas utilizando os diferentes métodos. Na Tabela 4.1 são mostrados os valores de SER para cada um dos testes. Podemos observar que o SER correspondente à aquisição utilizando a interpolação tem um resultado semelhante ao obtido com PPS1 por minimização TV. Entretanto, a aquisição pelo PPM obteve resultados melhores.

Observe também que quanto maior a imagem, maior a diferença entre a qualidade do resultado entre os métodos de aquisição, como pode ser visto nas imagens apresentadas

Tabela 4.1. Métrica SER para cada método e cada imagem.

SER	Subamostragem e interpolação	PPS1 e TV2D	PPM e TV2D
64×64	19,3	18,4	22,2
256×256	16,5	16,0	23,2
1024×1024	21,2	21,5	29,3

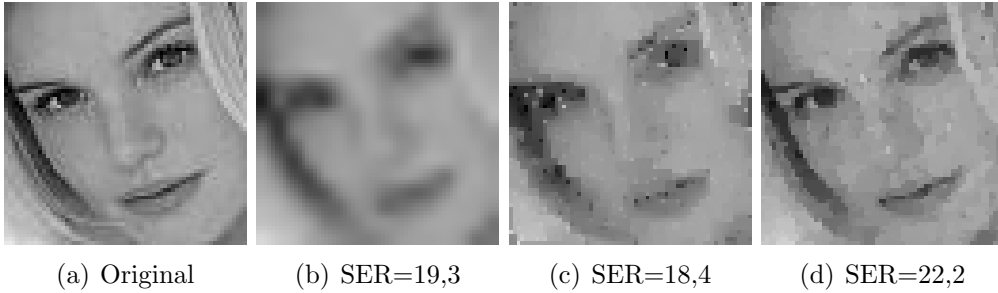


Figura 4.1. Resultado da reconstrução para uma imagem de 64 por 64. (a) Imagem original. (b) Sub-amostragem igualmente espaçada e reconstrução por interpolação, SER=19,3. (c) Amostragem pelo método PPS1 e reconstrução por TV2D, SER=18,4. (d) Amostragem pelo método PPM e reconstrução por TV2D, SER=22,2.

nas Figuras 4.2, 4.3 e 4.4. Em particular, a imagem apresentada nas Figuras 4.3 e 4.4 ficou com uma qualidade muito boa considerando que temos apenas 1/4 das medidas. Mesmo pequenos detalhes foram recuperados muito bem.

O resultado desse teste mostra o potencial do método para vídeo de alta velocidade. Se dividirmos os quadros de uma câmera em 4 e tomarmos medidas lineares de cada sub-quadro, podemos recuperar 4 imagens de qualidade tão boa quanto as mostradas neste teste, em que cada uma corresponde temporalmente a cada sub-quadro da cena. Desta forma, conseguimos aumentar a resolução temporal em 4 vezes, mantendo a resolução espacial constante. Alcançamos isso independente da redundância temporal da cena, ou seja, os 4 sub-quadros podem ser completamente diferentes uns dos outros. Além disso, podemos observar o potencial para super-resolução nesse teste. As medidas poderiam ser tomadas por uma câmera com 1/4 da resolução da imagem reconstruída, o que configura um aumento de resolução.

As cenas em geral têm muita redundância temporal. Logo, acreditamos que se as explorarmos, podemos ter resultados ainda melhores. Desta forma, na próxima seção, apresentamos os resultados da aplicação dos métodos de aquisição para sinais de vídeo, utilizando-se das redundâncias temporais.



Figura 4.2. Resultado de reconstrução para uma imagem de 256 por 256. (a) Imagem original. (b) Sub-amostragem igualmente espaçada e reconstrução por interpolação, SER=16,5. (c) Amostragem pelo método PPS1 e reconstrução por TV2D, SER=16,0. (d) Amostragem pelo método PPM e reconstrução por TV2D, SER=23,2.

4.2 Testes em Vídeos Sintetizados

Nesta seção apresentamos os testes em sinais de vídeos sintetizados que simulam os aspectos gerais de um vídeo natural, como movimento, oclusão, deformação e mudança de intensidade luminosa de objetos. Os objetos escolhidos foram elipses (fantomas) e retângulos. Eles foram escolhidos por serem largamente utilizados na área de *compressive sensing*, para reconstrução de imagens médicas, e funcionam melhor com a reconstrução TV.

A Figura 4.5(a) ilustra um quadro do vídeo de teste sintetizado. A partir desse vídeo,



(a) Original



(b) Interpolada, SER=21,2



(c) MP1, SER=21,5



(d) NI, SER=29,3

Figura 4.3. Resultado de reconstrução para uma imagem de 1024 por 1024. (a) Imagem original. (b) Sub-amostragem igualmente espaçada e reconstrução por interpolação, SER=21,2. (c) Amostragem pelo método PPS1 e reconstrução por TV2D, SER=21,5. (d) Amostragem pelo método PPM e reconstrução por TV2D, SER=29,3.

montamos 2 conjuntos de testes, variando o método de reconstrução. O primeiro conjunto de teste utiliza reconstrução 2D das colunas pelo tempo, enquanto que o segundo conjunto utiliza reconstrução 3D. Os resultados são apresentados nas subseções a seguir.

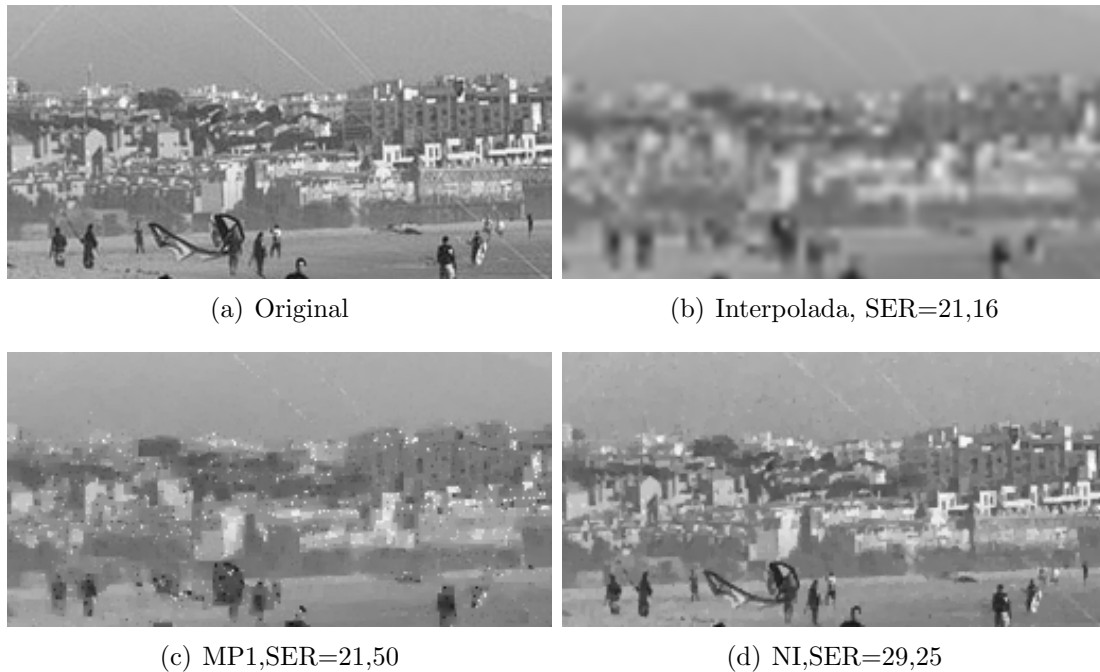


Figura 4.4. Resultado em um trecho de 64 por 128 da imagem de 1024 por 1024. (a) Imagem original. (b) Sub-amostragem igualmente espaçada e reconstrução por interpolação, SER=21,2. (c) Amostragem pelo método PPS1 e reconstrução por TV2D, SER=21,5. (d) Amostragem pelo método PPM e reconstrução por TV2D, SER=29,3.

4.2.1 Reconstrução 2D: Colunas dos Sub-quadros pelo Tempo

Para os 5 métodos de aquisição (no domínio temporal), utilizamos os vídeos sintetizados nos tamanhos 100×100 com 128 quadros, 100×100 com 256 quadros, e 200×200 com 128 quadros e 200×200 com 256 quadros. Foram utilizadas 4 taxas de sub-amostragem: 50%, 25%, 12.5% e 6.25%. Na Tabela 4.2 são apresentados os resultados das simulações. Para melhor visualização, os resultados também estão ilustrados no gráfico da Figura 4.6. Para ilustração da qualidade visual da reconstrução, um quadro do vídeo de teste ($200 \times 200 \times 128$) é apresentado na Figura 4.5(b), para o qual foi utilizado uma taxa de subamostragem de $16 \times$.

Percebe-se que o PPM2 apresentou o melhor resultado entre os métodos de aquisição para a maioria das taxas de sub-amostragem. O PPM2 só não se mostrou o melhor no caso em que a taxa de amostragem é de $2 \times$, e chegou a ter desempenho inferior ao FS para os menores tamanhos de vídeo. Por outro lado, note que essa taxa de sub-amostragem é a menos interessante. Na maior parte dos casos, o PPS2 mostra o segundo melhor desempenho. Podemos observar uma grande diferença de desempenho entre PPM1 e PPM2, o que mostra que escolhendo um padrão de espelhos diferente a cada sub-quadro leva a uma grande diferença no valor do SER da reconstrução.

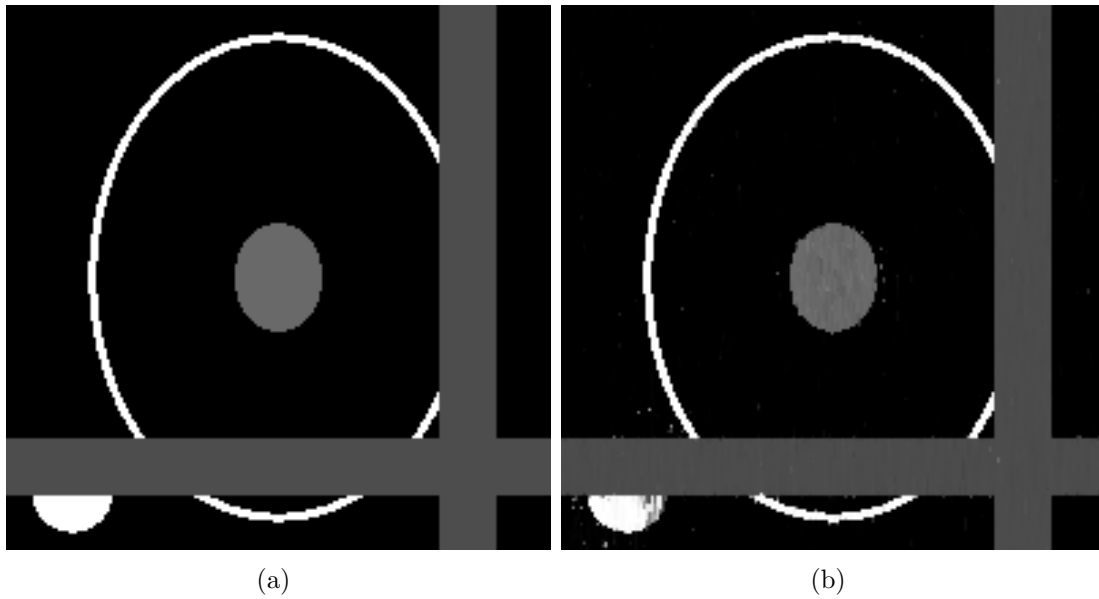


Figura 4.5. Um dos quadros da reconstrução do vídeo do fantoma de tamanho $200 \times 200 \times 128$, utilizando o PPM2 a uma taxa de sub-amostragem de 16x. a) o quadro original. b) o quadro reconstruído: SER = 17,8dB.

Tabela 4.2. SER (dB) dos vídeos reconstruídos em relação ao vídeo original.

Tamanho do vídeo	Método de aquisição	Taxa de sub-amostragem			
		$k = 2$	$k = 4$	$k = 8$	$k = 16$
100×	FS	40.3	6.5	3.8	3.2
	PPS1	10.7	3.6	0.4	-2.0
100×	PPS2	10.8	8.2	6.1	4.6
128	PPM1	15.3	10.9	3.7	-0.2
	PPM2	36.5	29.9	22.4	7.5
100×	FS	37.8	5.9	3.6	2.5
	PPS1	10.6	3.6	0.3	-2.1
100×	PPS2	10.7	7.8	6.1	4.6
256	PPM1	15.3	11.2	3.8	0.2
	PPM2	37.7	29.6	23.0	7.7
200×	FS	37.4	9.0	5.4	4.1
	PPS1	13.4	7.6	1.1	-1.2
200×	PPS2	13.4	9.9	8.0	6.6
128	PPM1	25.7	21.4	10.1	1.9
	PPM2	41.2	35.9	28.9	17.8
200×	FS	36.3	7.5	5.4	4.0
	PPS1	13.4	7.5	1.1	-1.2
200×	PPS2	13.4	10.0	8.0	6.7
256	PPM1	25.1	21.5	9.9	2.3
	PPM2	41.2	35.6	28.9	18.3

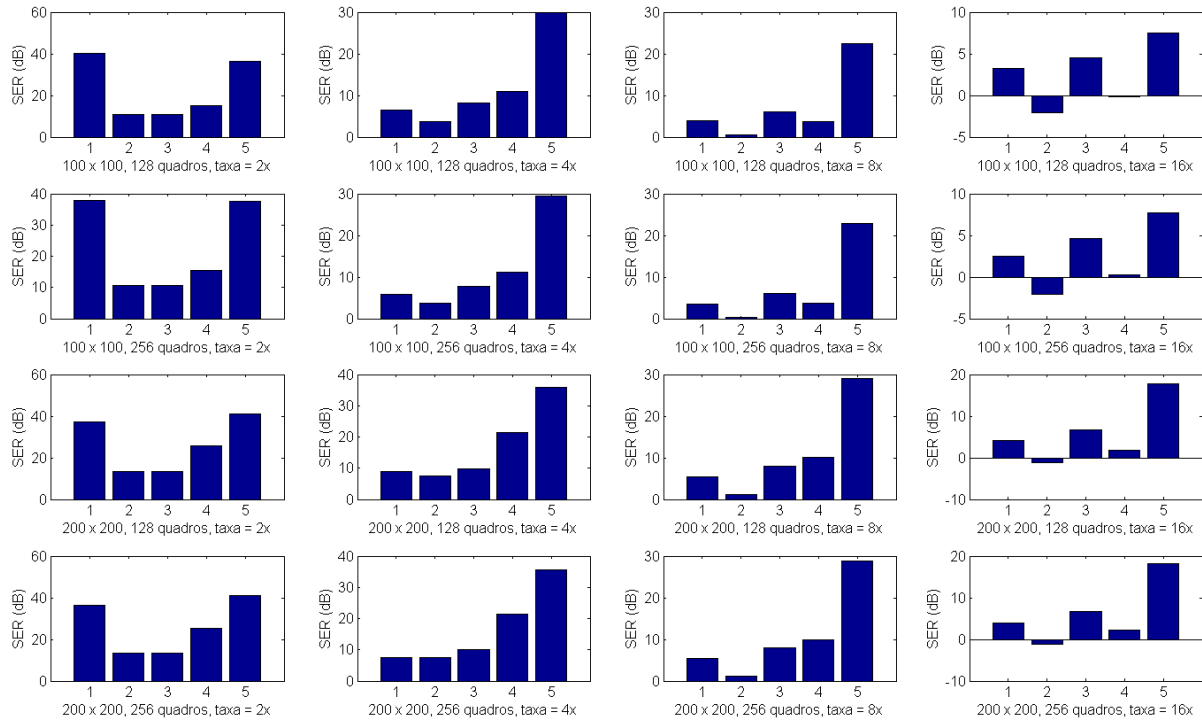


Figura 4.6. Gráficos para o teste em vídeos sintetizados 2D, das linhas pelo tempo. No eixo x são observados os diversos métodos de aquisição. 1 para o FS, 2 para o PPS1, 3 para o PPS2, 4 para o PPM1 e 5 para o PPM2. O eixo y apresenta a qualidade das reconstruções, dadas pelo SER em dBs.

Observe que, para vídeos maiores, os resultados são sempre melhores, resultado que está em concordância com os resultados para imagens apresentados na seção anterior. A expectativa é obter resultados ainda melhores quando utilizamos a reconstrução 3D, pois ela leva em consideração as redundâncias temporais e espaciais.

4.2.2 Reconstrução 3D

Repetimos os testes anteriores com reconstrução por TV3D. Os vídeos de testes são os mesmos do teste anterior. O que muda neste teste é que a reconstrução é realizada utilizando a TV3D. Os resultados são apresentados na Tabela 4.3 e na Figura 4.7.

Observe que, a reconstrução pela TV3D a partir do PPM (tanto PPM1 quanto PPM2) melhorou significativamente na maioria dos casos, com ganhos de até duas dezenas de dB em SER. Além disso, podemos observar um padrão quanto à qualidade de reconstrução de acordo com o número de quadros e a resolução espacial do vídeo. Com a resolução espacial fixa e variando apenas o número de quadros, os resultados não se alteraram muito. Porém, quando aumentamos o tamanho dos quadros, o resultado melhorou bastante. Isso aconteceu em todos os métodos de aquisição e todas as taxas de sub-amostragem, sendo mais evidentes para o método proposto e para taxas de sub-amostragem maiores. Poderia

Tabela 4.3. SER (dB) médios dos vídeos reconstruídos para os métodos de aquisição testados, para a reconstrução TV2D das colunas pelo tempo.

Tamanho do vídeo	Método de aquisição	Taxa de sub-amostragem			
		$k = 2$	$k = 4$	$k = 8$	$k = 16$
100×	FS	42.9	6.3	2.4	1.7
	PPS1	9.2	3.0	-0.3	-1.7
100×	PPS2	9.3	11.1	6.7	5.3
128	PPM1	27.2	20.1	12.2	1.2
	PPM2	67.2	55.6	33.7	8.7
100×	FS	42.4	8.5	1.2	0.74
	PPS1	9.2	2.7	-0.3	-1.7
100×	PPS2	9.3	11.6	7.0	5.3
256	PPM1	31.2	18.6	11.5	1.9
	PPM2	76.8	55.7	32.9	8.8
200×	FS	49.5	9.0	4.9	4.8
	PPS1	11.8	12.0	0.8	-1.4
200×	PPS2	11.7	13.8	10.4	8.4
128	PPM1	33.7	27.4	24.9	8.2
	PPM2	71.3	55.2	44.3	22.6
200×	FS	36.8	9.0	6.8	1.9
	PPS1	11.8	12.0	0.8	-1.4
200×	PPS2	11.7	14.1	10.4	8.5
256	PPM1	48.3	33.2	25.9	6.4
	PPM2	73.1	58.8	44.9	25.6

ser argumentado que isso é devido ao número de pixels, pois ao aumentar a resolução ou o número de quadros, estamos aumentando o número de pixels. Entretanto, o primeiro conjunto de vídeos tem tamanho $100 \times 100 \times 128 = 1.280.000$ pixels, o segundo tem tamanho $100 \times 100 \times 256 = 2.560.000$ pixels, o terceiro tem $200 \times 200 \times 128 = 5.120.000$ pixels e o quarto, $200 \times 200 \times 256 = 10.240.000$ pixels. Embora o segundo conjunto de vídeos tenha o dobro de pixels do primeiro, e o quarto conjunto de vídeos tenha o dobro do terceiro, a diferença entre as qualidades de reconstrução é pequena. Por exemplo, para o PPM2 a uma taxa de subamostragem de fator 16, o SER pulou de 8,7 para 8,8 dBs com o dobro do tamanho no caso do primeiro para o segundo conjunto de vídeos; e pulou de 22,6 para 25,6 para o dobro de pixels do terceiro para o quarto conjunto de vídeos. A diferença de qualidade de reconstrução entre o segundo e o terceiro conjunto de vídeos, o qual também tem o dobro de pixels, é muito maior (de 8,8 para 22,6). Então, a resolução espacial do vídeo é mais importante para um bom desempenho do que o número total de pixels no vídeo.

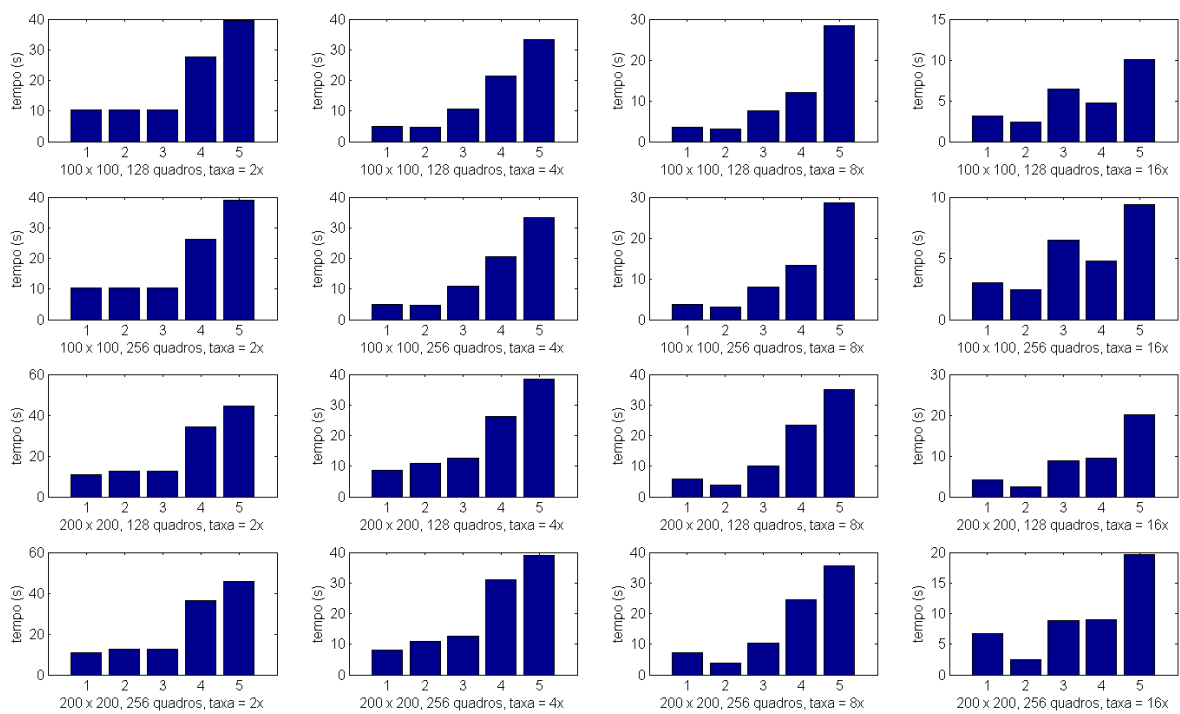
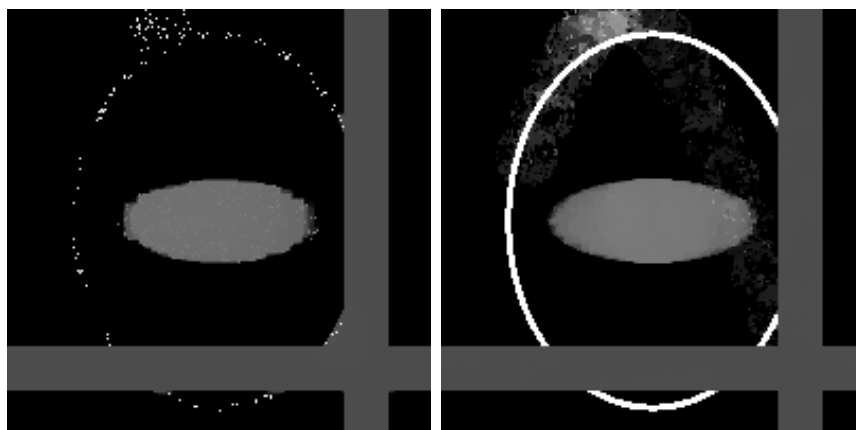


Figura 4.7. SER das reconstruções utilizando a TV3D do vídeo do vídeo sintetizado do fantoma, para taxas de 2, 4, 8 e 16. Em cada gráfico, o 1 se refere ao método FS, o 2 se refere ao PPS1, o 3 ao PPS2, o 4 se refere ao PPM1 e o 5 se refere ao PPM2.



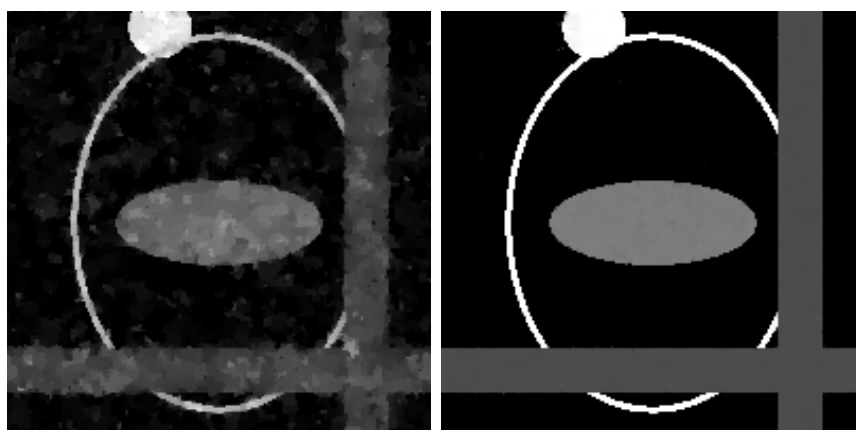
(a) Sub-quadro 85 do vídeo original

(b) FS



(c) PPS1

(d) PPS2



(e) PPM1

(f) PPM2

Figura 4.8. Teste com os métodos FS, PPS1, PPS2, PPM1 e PPM2 a partir de um vídeo de tamanho $200 \times 200 \times 128$, a uma taxa de sub-amostragem de 16x.

Na Figura 4.8 é apresentado um sub-quadro do resultado das reconstruções utilizando as 5 métodos de aquisição a partir do mesmo vídeo de teste, com um fator de sub-amostragem de 16x. Este é o sub-quadro de número 85 para o vídeo do fantoma sintetizado de tamanho $200 \times 200 \times 128$. A Figura 4.8(a) mostra o sub-quadro original. Ele é o quinto quadro de um conjunto de 16 quadros utilizados para o mesmo conjunto de medidas, em todas os métodos de aquisição testadas. Observe o círculo branco na parte superior desta imagem. Nos quadros anteriores, esse círculo está numa posição diferente. Nos quadros iniciais deste conjunto de sub-quadros, esse círculo vai se aproximando da borda de cima, até chegar nela, e então é rebatido para outra direção. Ao final do conjunto de 16 sub-quadros, o círculo se encontra numa posição completamente diferente. A qualidade da reconstrução utilizando o método FS, como ilustrado na Figura 4.8(b) fica muito distante da qualidade do quadro original. Isto mostra a influência dos quadros adjacentes.

A reconstrução a partir do PPS1 (Figura 4.8(c)) não apresenta resquícios de quadros anteriores ou posteriores, uma vez que este método não apresenta dependências temporais entre as medidas de sub-quadros distintos. Porém, muita informação da cena ficou faltando, pois a quantidade de informação luminosa adquirida é muito baixa para uma reconstrução fiel da cena.

A reconstrução a partir do método PPS2, ilustrado na Figura 4.8(d), mostra resquícios dos sub-quadros anteriores. Isso mostra que a dependência temporal das medidas nesse modo de amostragem pode influenciar no resultado da reconstrução, especialmente para áreas em que há movimento maior. Observamos que quanto maior o fator de subamostragem, maior será essa dependência e os resquícios de mais quadros comprometerão a qualidade da imagem.

O método PPM mostra como esses dois problemas podem ser abordados ao mesmo tempo. Não vemos resquícios dos sub-quadros adjacentes utilizados para o mesmo conjunto de medidas. Além disso, a cena está mais completa do que no resultado obtido com o método PPS1 porque 100% de informação luminosa foi adquirida. O resultado do PPM2 (Figura 4.8(f)) ficou melhor que o resultado do PPM1 (Figura 4.8(e)), mostrando que o padrão de espelhos aleatório por sub-quadro leva a melhores resultados. Podemos ver que os métodos com independência temporal conseguem separar melhor as informações temporais.

4.2.3 Robustez ao Ruído

É importante testar o quão robusta ao ruído é o nosso método e como essa robustez afeta a qualidade da reconstrução. Uma vez que os sensores das câmeras são sensíveis a ruído, as medidas também são afetadas pelo ruído durante a aquisição dos vídeos, para qualquer um dos modos de aquisição.

Fizemos testes com o mesmo tipo de vídeo utilizado nos testes anteriores, apenas adicionando ruído às medidas. É importante frisar que para simular o processo de aquisição, o ruído deve ser adicionado às medidas e não ao vídeo do qual estamos pegando as amostras, pois o ruído é gerado nos sensores. No caso do FS e PPS2, amostras de sub-quadros distintos são somadas e, em seguida, é adicionado o ruído. Se o ruído fosse primeiramente somado ao vídeo, teríamos nas medidas a soma de vários valores de ruídos Gaussianos de uma só vez, o que é equivalente a somar um único valor nas medidas. Esse valor também é gerado por um ruído Gaussiano, mas com uma variância diferente. O número de valores somados é cerca de $k/2$, uma vez que cerca de metade do vídeo é jogado fora. No caso do PPM (tanto PPM1 quanto PPM2), as amostras são somadas entre pixels distintos do mesmo sub-quadro. Porém, é equivalente a somar um único valor, mas com variância diferente. A variância é afetada de forma diferente em relação aos PPS2 e FS, pois são somados k valores. Apenas no caso PPS1 não há diferença entre somar os valores ao vídeo ou nas medidas, uma vez que as medidas são formadas por amostras aleatórias do vídeo, e não pela soma delas.

Em suma, adicionar o ruído às medidas torna o processo justo, pois cada método de amostragem tem um efeito diferente na variância. O mais justo para realizarmos uma avaliação do efeito do ruído em todos os métodos é adicionar ruído às medidas, pois, dependendo do modo de amostragem, o ruído adicionado ao vídeo tem pesos diferentes sobre as medidas.

Em nosso teste, nós somamos a cada amostra um valor aleatório com função de distribuição de probabilidade normal com média zero. A variância foi ajustada em valores igualmente espaçados dentro de uma escala exponencial. Os testes da seção anterior sugeriram que o número de quadros influencia pouco nos resultados, então decidimos fazer este teste com um número pequeno de quadros. Utilizamos vídeos sintéticos de tamanho $100 \times 100 \times 16$. Os resultados são apresentados na Figura 4.9.

Podemos ver que, embora a reconstrução sem ruído do PPM2 tenha ficado muito boa, ela decai rapidamente ao adicionarmos ruído às medidas. Então, nosso método de aquisição não é tão robusta ao ruído quando reconstruída com a TV3D. O PPS2 se mostra muito boa em relação a robustez ao ruído. Para taxas de sub-amostragem maiores, a qualidade da reconstrução ficou praticamente estável.

4.2.4 PPM2 Dividida em Blocos

Na Seção 3.3.1 vimos que o método PPM, se dividida em blocos, pode trazer vantagens em relação à visualização, dependendo do tamanho do bloco. Quanto menor o tamanho do bloco, melhor a pré-visualização, pois um tamanho de bloco menor pouco diminui a resolução espacial dela. Porém, o tamanho do bloco pode alterar a qualidade da

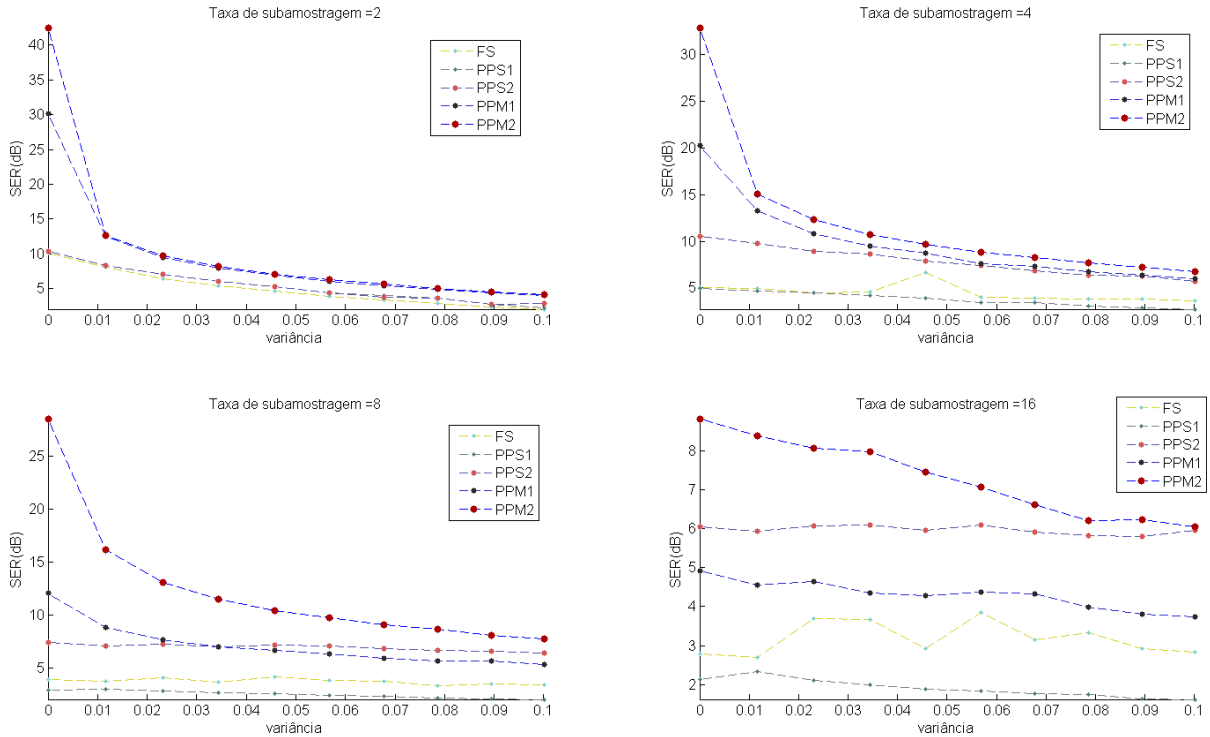


Figura 4.9. SER das reconstruções ao adicionar ruído às medidas

reconstrução. Com blocos muito pequenos, apenas os pixels mais próximos são somados. Desta forma, a informação não se espalha e a incoerência pode ser afetada. Nesta seção, medimos a influência do tamanho do bloco na qualidade da reconstrução.

A partir do fantoma de tamanho 512 por 512 de 16 quadros, tomamos medidas utilizando o método PPM2 divididos em blocos. O tamanho do bloco é a variável de controle do teste, de tal forma que ele varia em quadrados de laterais de 16 até 512, com passos de 16 em 16. Os resultados são apresentados na Figura 4.10.

Podemos observar nos gráficos ilustrados na Figura 4.10 que para blocos pequenos, os resultados da reconstrução não são muito bons. Mas, a medida que o tamanho do bloco aumenta, a qualidade da reconstrução aumenta até se estabilizar. Quanto maior a taxa de sub-amostragem, maior é o ponto em que a qualidade se estabiliza. O tamanho do bloco a partir do qual se atinge estabilidade é de 48×48 para taxa de sub-amostragem é $2 \times$, enquanto que, para uma taxa de sub-amostragem de $16 \times$ o tamanho do bloco mínimo é 144×144 . Porém, nestes casos, os quadros de pré-visualização ficam com resolução muito pequena.

Como o ganho para a pré-visualização não é muito grande e a qualidade não varia muito depois do tamanho mínimo do bloco de estabilização, os demais testes serão feitos sem utilizar o PPM2 por blocos.

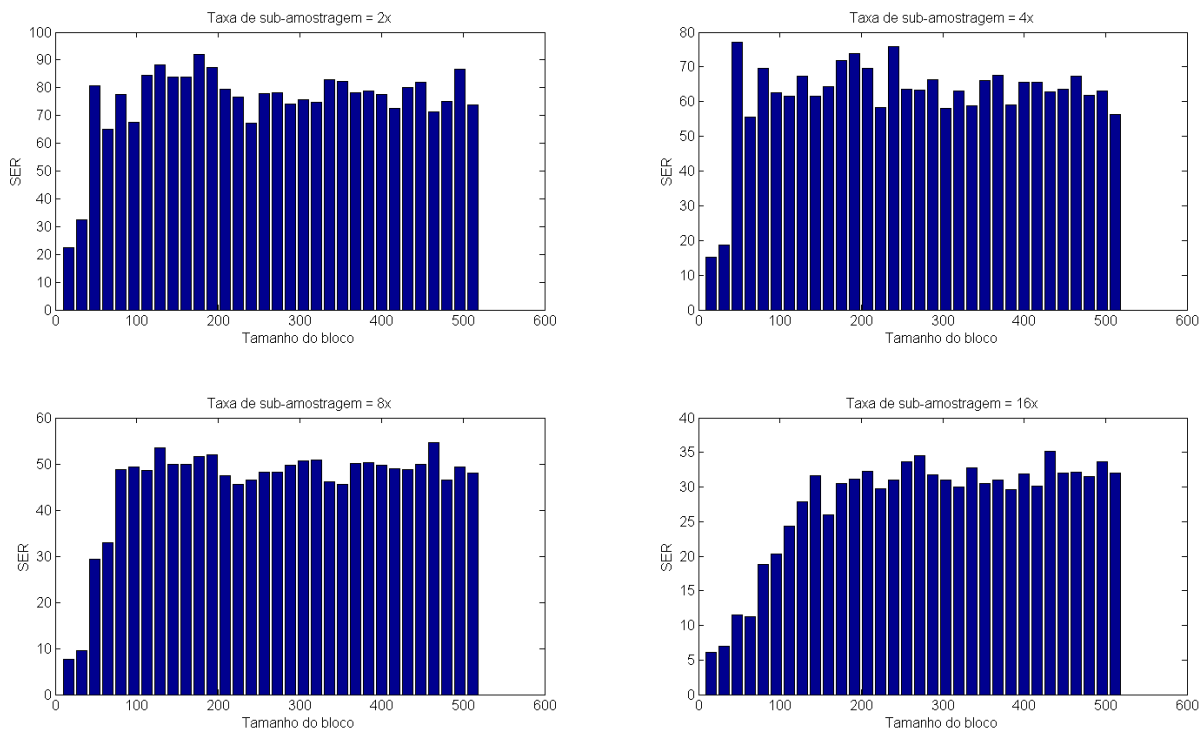


Figura 4.10. SER das reconstruções a partir do método PPM2 por blocos.

4.3 Vídeos Naturais

Os fantasmas mostrados nas Seções anteriores têm a característica de serem muito esparsos em relação ao operador de diferenças finitas de amostras consecutivas, o que é medido pelo operador TV. Então, os resultados com minimização TV são, de forma geral, bons. Infelizmente, os vídeos naturais não são tão esparsos em relação a esse operador como os vídeos sintéticos de fantasmas. Nesta seção, apresentamos teste de como os métodos de aquisição se comportam com vídeos naturais.

Os resultados em vídeos sintéticos forneceram boas direções sobre os parâmetros que levam aos melhores resultados:

- Vídeos de maior resolução espacial apresentam melhor qualidade de reconstrução.
- Aparentemente, o número de quadros a serem reconstruídos juntos não afeta consideravelmente a qualidade das reconstruções.
- Para maiores taxas de sub-amostragem e maiores níveis de ruído, o PPS2 mostrou um bom resultado.

Considerando isso, selecionamos para teste um conjunto de vídeos em resolução HD (720×1280 pixels, 50 quadros por segundo). De cada um dos vídeos, escolhemos um conjunto de 16 quadros para os testes de simulação da aquisição e reconstrução. A escolha

do número de quadros se deve, entre outros fatores, à dificuldade de processar vídeos com grande resolução espacial e com muitos quadros. Como, de acordo com os resultados obtidos para os fantasmas, a resolução espacial tem maior impacto sobre a qualidade da reconstrução do que o número de quadros, decidimos minimizar o número de quadros para poder aumentar a resolução espacial do vídeo sem aumentar drasticamente o tamanho dos vídeos (armazenamento em memória). Entretanto, o número de quadros depende da taxa de sub-amostragem. Para testar a taxa de sub-amostragem de 16x, precisamos de um mínimo de 16 quadros de um vídeo original. Desta forma, cada modo de amostragem vai gerar 1 quadro com as amostras, sendo este o número de amostras que são captadas pelos sensores da câmera de cada vez. Em suma, 16 é o número mínimo de quadros que devemos capturar para simular o processo de aquisição para uma taxa de sub-amostragem de 16x. Evidentemente, os métodos PPS1 e PPM permitem um número qualquer de quadros, mas como vamos comparar todos os métodos, os vídeos devem obedecer as condições compatíveis para todos os métodos.

Os vídeos escolhidos foram obtidos do banco de dados *The Consumer Digital Video Library* (CDVL). Os vídeos são não-comprimidos (formato avi) com resolução 720×1280 , capturados a 50 FPS e possuem sub-amostragem de cor a 4:2:0. Para testes deste tipo, é importante que os vídeos sejam de alta qualidade e sem degradações visíveis, a utilização de vídeos comprimidos poderia afetar os resultados por causa da introdução de degradação. Por exemplo, a blocagem de uma compressão JPEG cria bordas entre os blocos 8×8 , ao mesmo tempo que deixa as regiões constantes por partes [56]. A criação de regiões constantes poderia ajudar na reconstrução por minimização TV.

De cada um dos vídeos, selecionamos 16 quadros para os testes. Em alguns dos vídeos, os quadros eram seguidos, e de outros escolhemos os quadros com saltos maiores, porém constantes entre os quadros, com a ideia de simular uma movimentação maior entre alguns deles. A Tabela 4.4 mostra os vídeos e os quadros escolhidos de cada um. A Figura 4.11 mostra o primeiro e último quadro de cada um dos vídeos escolhidos. Quanto maior a diferença entre o primeiro e último quadro, maior é a atividade temporal de cada vídeo.

Nos testes anteriores realizados nos fantasmas, vimos que os melhores resultados para as maiores taxas de sub-amostragem foram obtidos para os métodos PPS2 e PPM2. Além disso, o método PPS2 mostrou a melhor robustez ao ruído. Como estes testes demandam uma considerável quantidade de tempo e espaço em disco, escolhemos apenas os métodos de aquisição PPS2 e PPM2 para estes testes.

Além disso, devido a problemas na solução inicial para o algoritmo utilizado para TV3D para o método PPM2, reconstruímos o vídeo obtido pelo método PPM2 pela TV3D com a solução inicial dada pela reconstrução da TV2D dos sub-quadros isoladamente. Como vimos na implementação descrita na Seção 3.5, as medidas obtidas podem ser

Tabela 4.4. Os quadros de início e final e os saltos dos vídeos, note que cada um contém 16 quadros.

Video	Quadro inicial	Salto em quadros	Quadro final
1	402	1	417
2	80	2	830
3	440	1	455
4	1	1	16
5	70	1	85
6	20	20	320
7	1100	50	1550
8	300	1	315
9	96	1	111
10	30	1	45
11	60	2	90
12	100	30	550

separadas, de tal forma que as primeiras M/k medidas correspondam ao primeiro sub-quadro, as próximas M/k medidas ao segundo sub-quadro, e assim sucessivamente até o último sub-quadro. Desta forma, podemos adquirir as medidas isoladamente de cada sub-quadro a partir das medidas do vídeo 3D. E assim, podemos reconstruir cada um dos sub-quadros isoladamente. Efetuamos a reconstrução da TV3D a partir das medidas PPM2, mas com a solução inicial dada pela reconstrução 2D dos sub-quadros isolados. Chamamos essa configuração de PPM3. Para fins de comparação, também mostramos nos resultados os sub-quadros reconstruídos isoladamente pela TV2D, que são tomados pelo PPM3 como solução inicial. Chamaremos essa solução inicial de PPM-2D.

Os resultados do SER das reconstruções para os métodos PPS2, PPM2, PPM-2D e PPM3, para as taxas de sub-amostragem de 4, 8 e 16 são apresentadas na Tabela 4.5. A última linha desta tabela apresenta o resultado médio para todos os vídeos. Para fins de melhor visualização, os resultados também são apresentados na Figura 4.12.



Figura 4.11. O primeiro e último quadro de cada um dos 12 vídeos escolhidos para os testes. Eles foram obtidos a partir de uma escolha de quadros de vídeos originais, de acordo com a Tabela 4.4.

Tabela 4.5. SER (dB) médio dos vídeos naturais reconstruídos.

Video	PPS2			PPM2			PPM-2D			PPM3			
	k=	4	8	16	4	8	16	4	8	16	4	8	16
1		11.3	10.0	8.4	12.7	9.1	7.3	15.0	12.4	10.7	15.8	13.2	11.5
2		15.1	13.4	11.9	17.4	12.3	10.0	20.2	17.0	14.7	20.4	17.5	15.3
3		18.4	17.9	14.6	21.7	13.9	11.1	27.7	23.9	21.3	27.1	24.0	21.7
4		12.1	10.2	9.2	15.0	12.1	10.4	17.0	14.7	13.3	17.2	14.9	13.5
5		13.3	12.0	10.1	18.9	11.8	8.4	30.2	25.9	22.2	25.9	22.0	19.7
6		16.7	14.4	13.5	17.3	12.3	9.7	24.3	20.9	18.7	22.0	19.4	17.9
7		11.7	9.6	7.2	14.3	10.4	8.6	24.1	21.8	20.4	19.7	17.4	17.0
8		11.9	10.5	8.8	14.8	9.5	7.0	19.8	16.7	14.4	19.2	16.6	14.5
9		10.6	8.0	5.8	14.0	8.5	5.5	23.0	21.0	18.6	22.9	20.6	18.2
10		16.0	13.1	11.6	20.8	11.6	8.5	26.5	22.8	20.1	28.0	24.1	21.3
11		17.4	14.5	13.4	20.4	14.9	12.2	29.2	25.1	21.5	26.5	23.2	20.5
12		13.2	11.1	9.2	14.8	10.5	8.7	25.2	21.3	18.5	22.5	19.6	17.1
Média		14.0	12.5	10.3	16.8	11.4	9.0	23.5	20.3	17.9	22.3	19.4	17.3

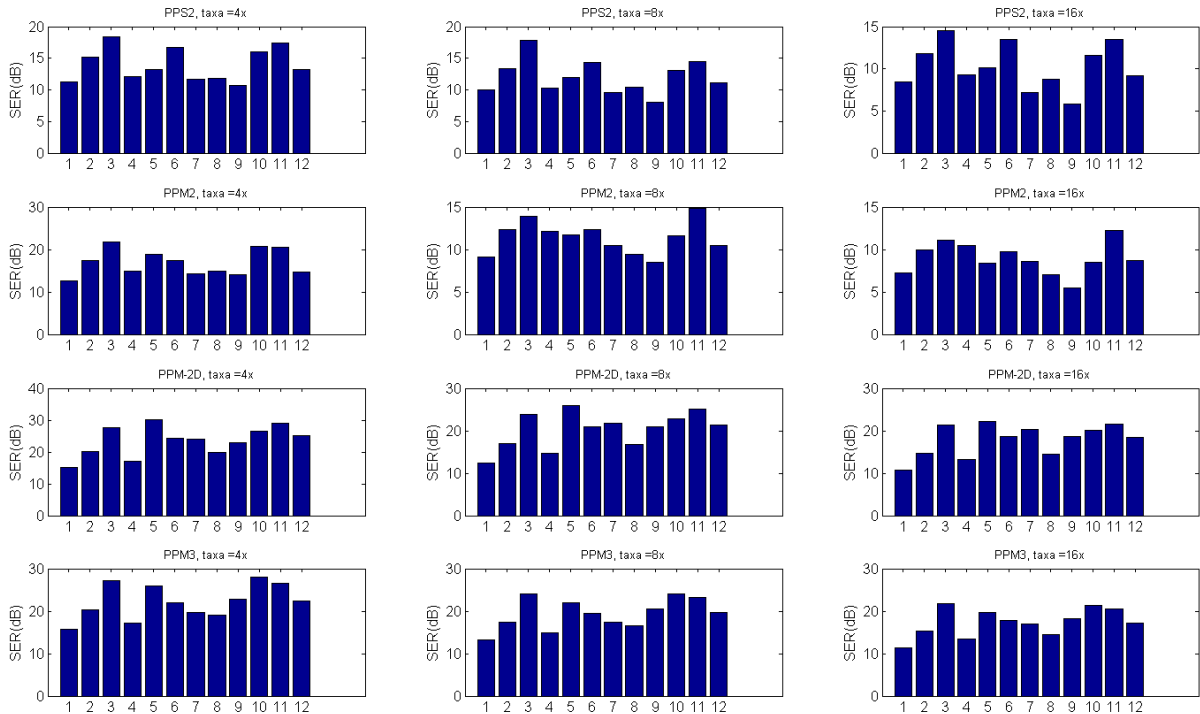


Figura 4.12. SER da reconstrução dos 12 vídeos para cada método testado. Em cada gráfico, o eixo x se refere ao número do vídeo, e o eixo y se refere ao SER (dB)

Comparando a reconstrução do PPS2 com o PPM2, para uma taxa de sub-amostragem 4x, o PPM2 teve melhor desempenho (SER médio de 14 para o PPS2 e 16,8 para PPM2). Porém, com o aumento da taxa, o PPS2 começou a ter um melhor desempenho do que o PPM2 para a maioria dos casos, apresentando um menor decaimento dos valores de SER.

Esse comportamento foi visível no teste de robustez ao ruído. Apesar do PPM2 apresentar melhores resultados do que o PPS2, quanto maior o ruído e taxa de sub-amostragem, melhor a qualidade de reconstrução utilizando o PPS2. De tal forma que o PPS2 e o PPM2 apresentam desempenhos equivalentes para os maiores níveis de ruído testados. O modelo utilizado para vídeos naturais assume que eles são aproximadamente esparsos, ou seja, seriam equivalentes a um sinal esparsos somado a um ruído de baixa energia. Ao se adquirir as medidas, o ruído também é medido. Como o PPM2 não mostrou boa robustez ao ruído como o PPS2, faz sentido que os vídeos naturais reconstruídos pelo PPM2 tenham uma qualidade mais baixa de que os reconstruídos com o PPS2 em taxas de sub-amostragem ainda maiores.

Porém, o PPS2 não gerou o melhor desempenho em todos os casos da taxa de sub-amostragem 16x. O PPM2 apresentou melhor desempenho para os vídeos 4 e 7. O vídeo 4 mostra dois quadros sem nenhuma informação visual relevante e 14 quadros com um vídeo normal de um homem caminhando. Sendo assim, há uma transição abrupta do conteúdo das cenas. O vídeo 7 mostra um vídeo com movimento de câmera. Inicialmente, a cena mostra uma casa com um campo ao lado. Em seguida, a câmera se aproxima da casa e das árvores, terminando de baixo da copa de uma das árvores, onde há um grupo de pessoas se movimentando. Para compor o vídeo de testes, tomamos um quadro a cada 50, com o objetivo de simular a aproximação da câmera de forma mais rápida. A transição de quadros é, portanto, muito grande, assim como no vídeo 4.

A Figura 4.13 ilustra o segundo e terceiro quadros do vídeo 4, e suas reconstruções para a taxa de sub-amostragem 16x. A Figura 4.14 ilustra o primeiro e décimo sexto quadro do vídeo 7 e suas respectivas reconstruções para a taxa de 16x.

Para o método PPS2, a reconstrução do segundo quadro do vídeo 4 mostra uma clara influência do quadro seguinte. Em particular, podemos ver claramente o efeito de *motion blur* em ambos os quadros. Para o método PPM2, o segundo e terceiro quadros ficaram bem ruidosos, porém, há muita redundância do conteúdo de um quadro em outro. Para o método PPM3, a reconstrução do segundo quadro sofre uma influência do terceiro em uma intensidade muito menor, e o terceiro quadro ficou mais fiel ao original. A solução inicial dada pelo PPM-2D, entretanto, não mostra influência dos outros quadros no segundo quadro, porém a qualidade do terceiro quadro é pior em qualidade visual em relação ao terceiro quadro do PPM3.

Como o vídeo 7 apresenta grande atividade temporal entre os quadros subsequentes,

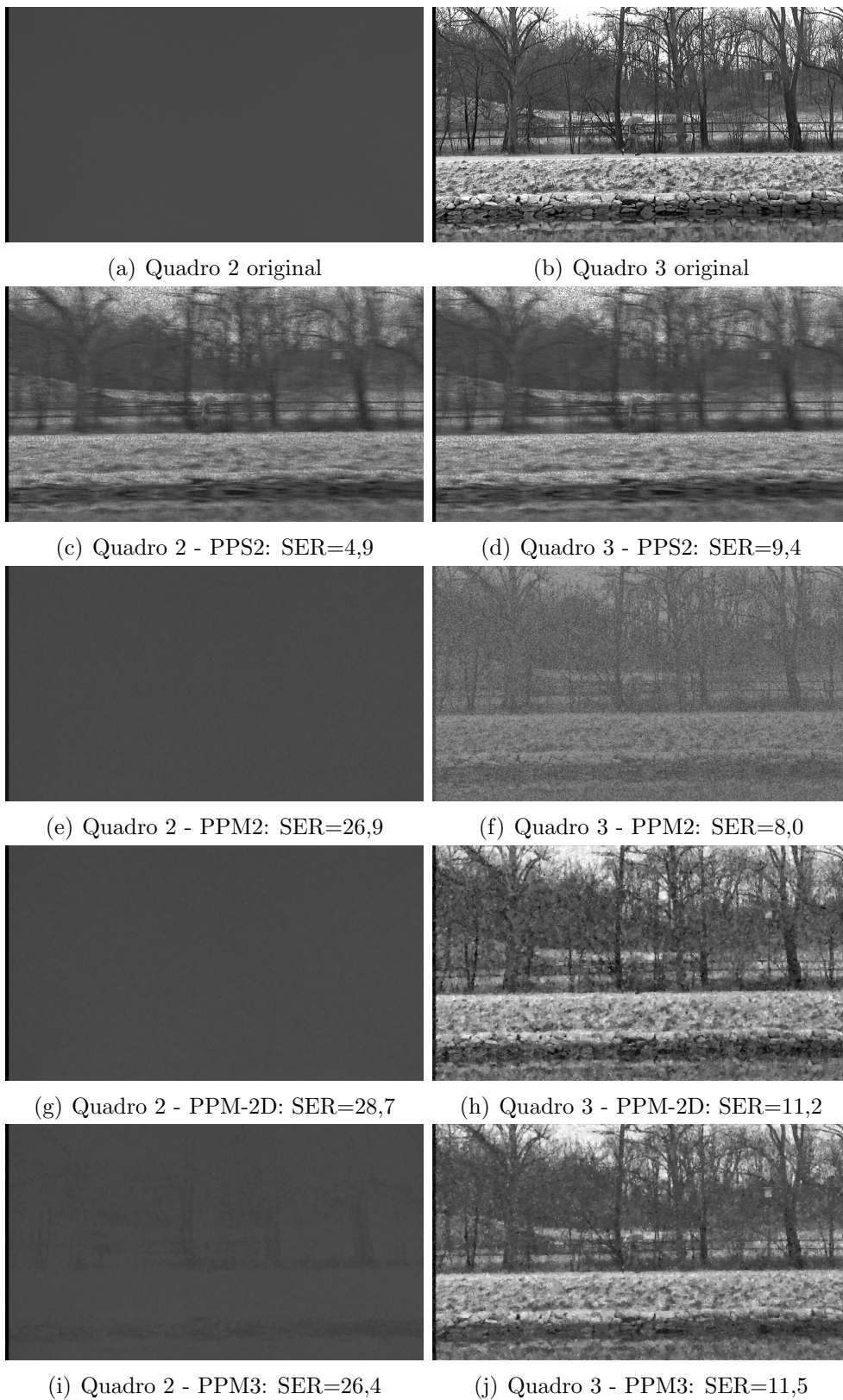


Figura 4.13. Reconstruções do segundo e terceiro quadros do vídeo 4 a partir das diversos métodos, a uma taxa de 16x (6,25% das amostras).



(a) Quadro 1 original



(b) Quadro 16 original



(c) Quadro 1 - PPS2: SER=7,6



(d) Quadro 16 - PPS2: SER=5,3



(e) Quadro 1 - PPM2: SER=6,7



(f) Quadro 16 - PPM2: SER=7,3



(g) Quadro 1 - PPM-2D: SER=22,5



(h) Quadro 16 - PPM-2D: SER=12,1



(i) Quadro 1 - PPM3: SER=20,2



(j) Quadro 16 - PPM3: SER=11,2

Figura 4.14. Reconstruções do primeiro e último quadros do vídeo 7 a partir das diversos métodos, a uma taxa de 16x (6,25% das amostras).

Tabela 4.6. Medida de movimento utilizada para medir a atividade temporal dos vídeos dada por 4.1. Quanto maior o valor obtido pela métrica, maior o movimento do vídeo.

Vídeo		10	3	11	2	5	9	6	1	8	4	12	7
Movimento		1849	2711	4911	5146	5483	5556	6873	7193	7332	8876	9572	11068

o método PPS2 apresentou um resultado pior, pois reconstrução dos quadros tem muita influência do conteúdo dos outros quadros. No método PPM2 as reconstruções ficaram muito ruidosas, porém, com os quadros bem definidos. Já o PPM3 ficou bem melhor que ambos, apresentando boa definição da informação nos quadros e pouco ruído. Entretanto, neste caso, a própria solução inicial dada pelo PPM-2D obteve um desempenho melhor que o PPM3. Observe que a mistura de conteúdo entre os quadros é inexistente, e a qualidade dos quadros isoladamente é superior.

Esse comportamento sugere que uma grande transição dos quadros subsequentes favorece o nosso método PPM, nas suas diversas variações. A partir disso, surgiu a ideia de medir a transição temporal entre os quadros e verificar se essa tendência é verdadeira. Além disso, podemos comparar o desempenho da solução inicial com o da solução final do PPM3 ao se variar a quantidade de movimento.

Para medir a quantidade de movimento, escolhemos uma medida simples, que soma os valores absolutos das diferença entre os quadros. Assim, quanto maior a diferença entre os quadros, ou seja, maior a transição entre os quadros, maior é o valor da medida. À medida, é adicionado um fator de normalização da energia do vídeo, pois vídeos com maior luminância média poderiam ter um peso maior na medida do que vídeos com menor luminância média. Para um vídeo V , de dimensões $p \times q \times r$, a métrica é calculada utilizando a seguinte equação:

$$\frac{\sum_{k=1}^{r-1} |V(:, :, k) - V(:, :, k - 1)|}{(r - 1) \sqrt{\sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r V(i, j, k) V(i, j, k)^*}}. \quad (4.1)$$

Com essa métrica, classificamos os vídeos de acordo com sua atividade temporal. A Tabela 4.6 mostra o resultado da aplicação da medida de movimento aos vídeos. Os resultados foram ordenados de acordo com a quantidade de movimento (ordem crescente). A Figura 4.15 mostra a relação entre a métrica e o SER da reconstrução.

Com taxa de sub-amostragem 4x, o PPM2 apresenta melhor desempenho que o PPS2. Porém, quando a taxa aumenta, o PPS2 começa a superar o PPM2. O aumento da quantidade de movimento sugere que a diferença entre o PPS2 e o PPM2 diminui. Entretanto, em todos os casos, o PPM3 e a inicialização PPM-2D, mostraram uma desempenho muito melhor que o PPM2 e o PPS2.

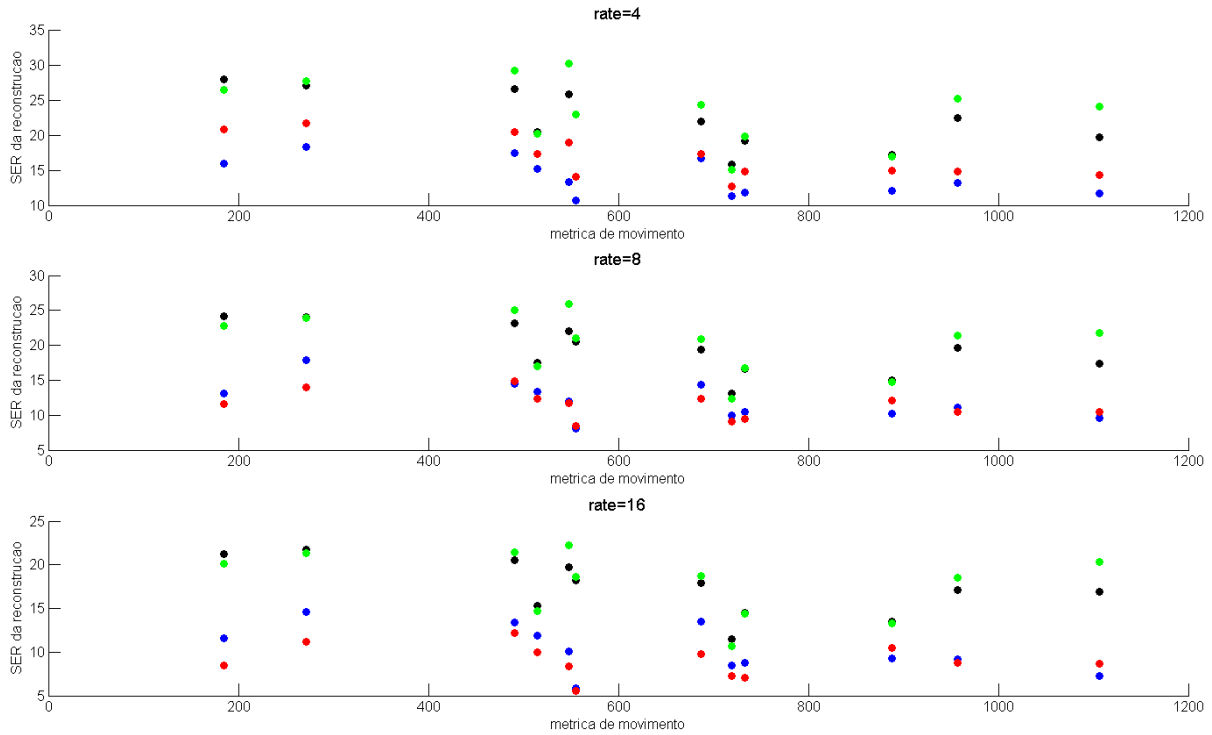


Figura 4.15. SER ordenados pela métrica de movimento para taxas de 4x, 8x e 16x. Em azul é o PPS2, em vermelho é o PPM2, em verde o PPM-2D, e o preto corresponde ao PPM3.

Para visualizar melhor este comportamento, analisamos um trecho em zoom de mais dos vídeos 9 e 8. A Figura 4.16(a) ilustra um trecho de uma área de 600×600 do primeiro quadro do vídeo 9. Note que existe uma sombra que cobre as costas do homem. Essa sombra se move durante os quadros. As Figuras 4.16(b), (c), (d) e (e) mostram a reconstrução utilizando PPS2, PPM2, PPM-2D e PPM3, respectivamente, com uma taxa de sub-amostragem de 16x (6.25% das amostras). No quadro reconstruído pelo PPS2, as costas do homem estão visíveis. Entretanto, as costas do homem só estariam visíveis nos quadros seguintes. Ou seja, o PPS2 não separa muito bem os quadros. Por outro lado, a reconstrução com o PPM2 mostrou-se muito ruidosa (Figura 4.16(c)). Tanto o PPM-2D quanto o PPM3 ficaram muito melhores. O PPM-2D apresentou um SER um pouco melhor, mas ainda é muito próximo do PPM3. Visualmente o PPM3 está um pouco melhor, com menos artefatos característicos da minimização TV.

A Figura 4.17(a) mostra um trecho de 300×300 do vídeo 8. O vídeo contém uma multidão de pessoas correndo, apresentando muito movimento. As Figuras 4.17(b), (c), (d) e (e) mostram as reconstruções dos métodos PPS2, PPM2, PPM-2D e PPM3, respectivamente, com taxa de sub-amostragem de 4x (a partir de 25% das amostras). Na reconstrução PPS2 é possível ver muito *motion blur*. Como antes, a imagem reconstruída com o PPM2 está ruidosa, e os PPM-2D e PPM3 apresentam melhores qualidades visuais.



(a) original



(b) PPS2



(c) PPM2



(d) PPM-2D

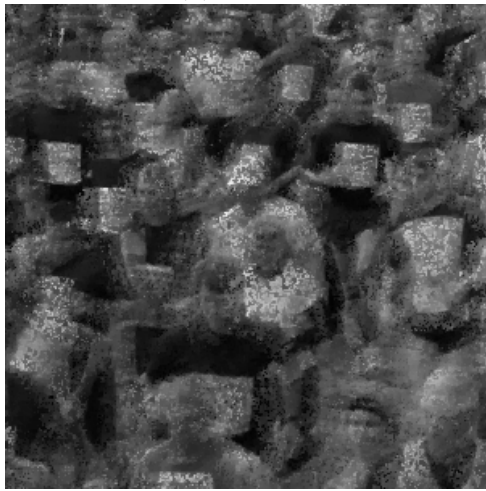


(e) PPM3

Figura 4.16. Resultados para os métodos PPS2, PPM2, PPM-2D e PPM3 para o vídeo 9, com taxa de sub-amostragem de 16x.



(a) original



(b) PPS2



(c) PPM2



(d) PPM-2D



(e) PPM3

Figura 4.17. Resultados para os métodos PPS2, PPM2, PPM-2D e PPM3 para o vídeo 8, com taxa de sub-amostragem de 4x.

É difícil entender porque o PPM-2D mostrou ótimos resultados. Testamos poucos números de quadros no teste com vídeos sintetizados (apenas 128 e 256) para deduzir que os resultados seriam melhores poucos quadros em vídeos naturais. Acreditamos que o resultado 2D ficou melhor pois o número de quadros reconstruídos no PPS2 e PPM2 era muito pequeno, e a resolução espacial 2D era grande suficiente pro PPM-2D apresentar um bom desempenho. Mais testes com vídeos naturais de tamanhos e número de quadros diferentes seriam necessários para investigar este desempenho. Infelizmente precisaríamos de mais recursos computacionais e mais tempo para testar essa hipótese.

4.4 Análise dos Tempos de Execução

Para todos os testes, o mesmo algoritmo de reconstrução foi utilizado, variando apenas a forma de aquisição. É importante analisar o tempo de execução dos testes para verificar se a forma de aquisição pode influenciar na velocidade em que a reconstrução é obtida.

Todos os testes foram feitos utilizando o mesmo computador: um PC com 4 *cores* do processador Intel(R) Xeon(R) CPU E31220 U em 3093 MHz, 64 bits e 8 gigabytes de memória ram. As implementações foram feitas em MATLAB. O tempo medido é relativo apenas ao algoritmo de reconstrução, ou seja, não medimos o tempo tomado na abertura dos dados, simulação da aquisição ou no salvamento dos dados. Além de serem muito pequenos relação ao tempo de reconstrução, podemos avaliar a influência do tipo de amostragem exatamente no tempo de reconstrução, e não outros fatores.

Como o algoritmo é de otimização, ele itera um certo número de vezes até que a solução obtida seja próxima suficiente das condições desejadas, ou quando um número máximo de iterações for atingido. Em nenhum caso mostrado aqui, o algoritmo alcançou o número máximo de 300 iterações, ou seja, para todos os testes apresentados a mínima TV foi alcançada.

Vídeos Sintetizados

Inicialmente avaliamos os tempos de execução dos testes para reconstrução dos vídeos sintetizados. As Figuras 4.18 e 4.19 apresentam os gráficos de tempo de reconstrução para os testes feitos a partir das reconstruções 2D das colunas pelo tempo e reconstrução 3D, respectivamente. Em geral, o tempo médio de reconstrução a partir dos métodos de amostragem foi proporcional ao tamanho do sinal reconstruído, com algumas poucas exceções.

No caso de reconstrução das linhas pelo tempo, o método proposto em sua primeira configuração apresentou um tempo de reconstrução um pouco maior nos casos em que

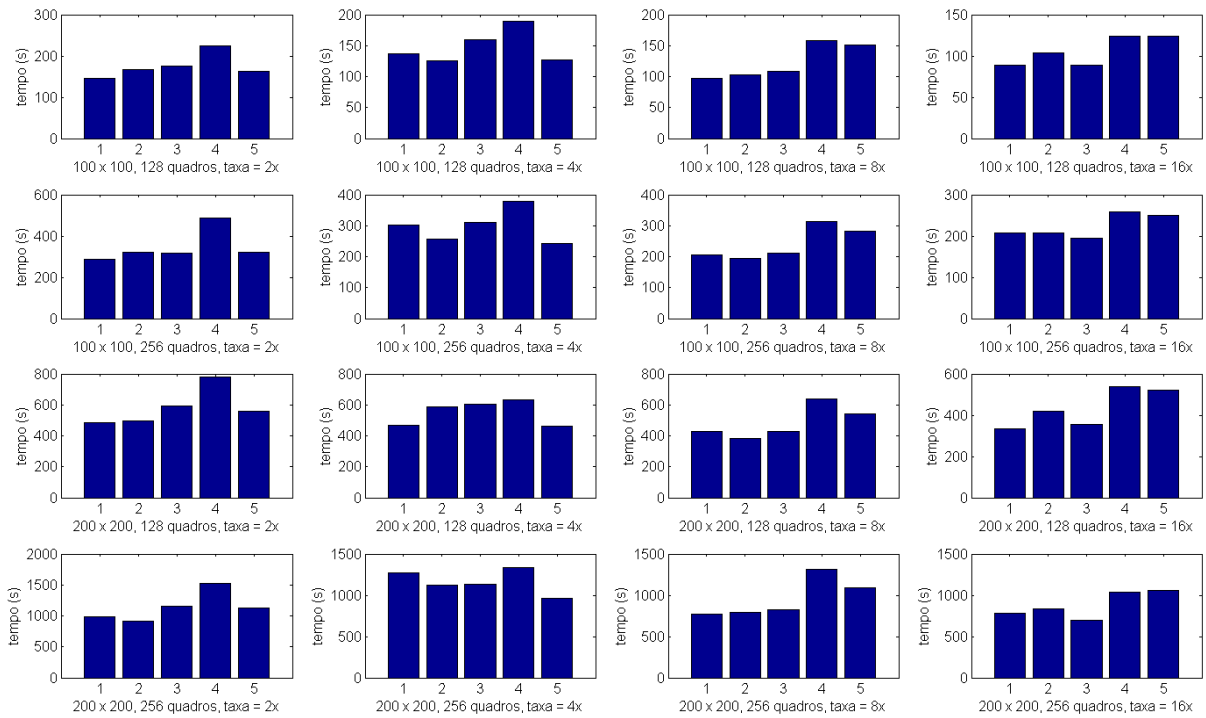


Figura 4.18. Tempos de reconstrução dos testes de reconstrução 2D das colunas pelo tempo.

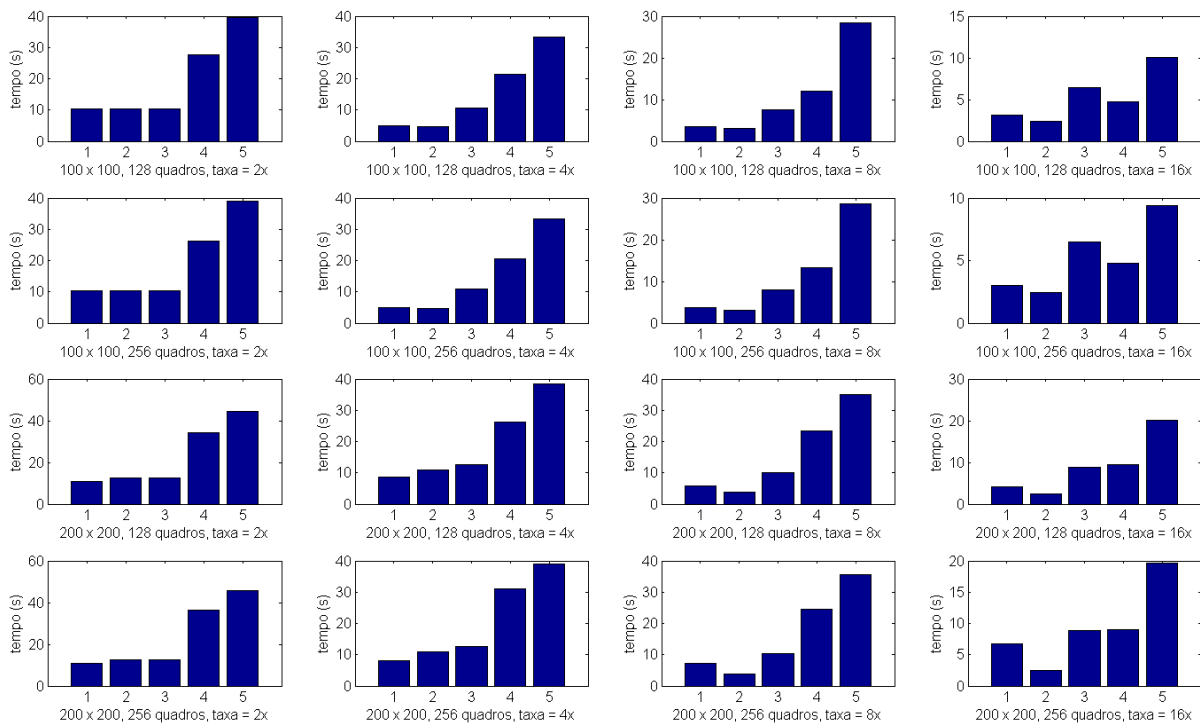


Figura 4.19. Tempos de reconstrução dos testes de reconstrução 3D para vídeos sintetizados.

o a taxa de sub-amostragem é maior. O PPM2 apresentou melhor desempenho quando comparado ao PPM1.

No caso da reconstrução 3D, houve muita variação no tempo de reconstrução com taxas de sub-amostragem mais baixas. Porém, para taxas de sub-amostragem maiores, o tempo de reconstrução foi parecido para todas os métodos de aquisição. Então, a partir dos dados, não é possível concluir que um dos método leva vantagem em tempo de execução da reconstrução em relação aos outros.

Influência do Ruído no Tempo de Reconstrução

Vamos analisar agora os testes da robustez ao ruído, quanto ao seu tempo de reconstrução. O gráficos na Figura 4.20 ilustram os tempos de execução para cada uma dos métodos, para diferentes níveis de ruído.

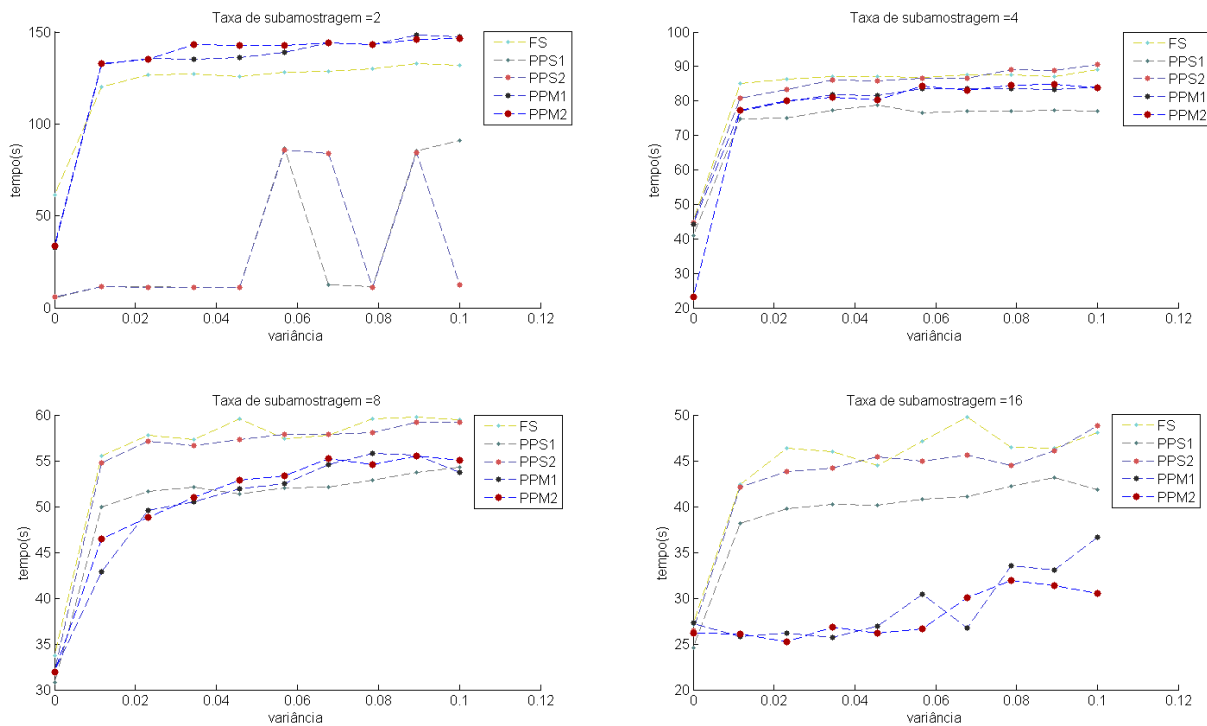


Figura 4.20. Tempos de reconstrução dos testes de reconstrução 3D com ruído nas medidas.

Para a reconstrução sem ruído, o tempo de execução não variou muito. Porém, mesmo ao adicionar um ruído de baixa intensidade, o tempo das reconstruções aumentou bastante. Na maioria dos casos, o tempo de reconstrução mais do que dobrou. Com taxas de sub-amostragem menores, o método proposto apresenta os piores resultados. Mas, surpreendentemente, com o aumento da taxa de sub-amostragem, o método proposto apresenta o melhor resultado no que diz respeito ao tempo de reconstrução.

Como já foi mencionado, com as maiores taxas de sub-amostragem, o método proposto se mostrou menos robusta ao ruído que por exemplo o PPS2. Tal comportamento sugere um compromisso entre o tempo de execução e a qualidade de reconstrução.

Influência do Tamanho do Bloco do Método PPM2 no Tempo de Reconstrução

Na Seção 3.3.1 propomos uma alteração no método PPM2, na qual as medidas são escolhidas a partir de amostras que estão próximas, em um mesmo bloco. Então, medimos a qualidade de reconstrução ao variar o tamanho do bloco. Vamos agora analisar se o tamanho do bloco pode trazer alguma vantagem no tempo de reconstrução. A Figura 4.21 ilustra os tempos de reconstrução para esse teste.

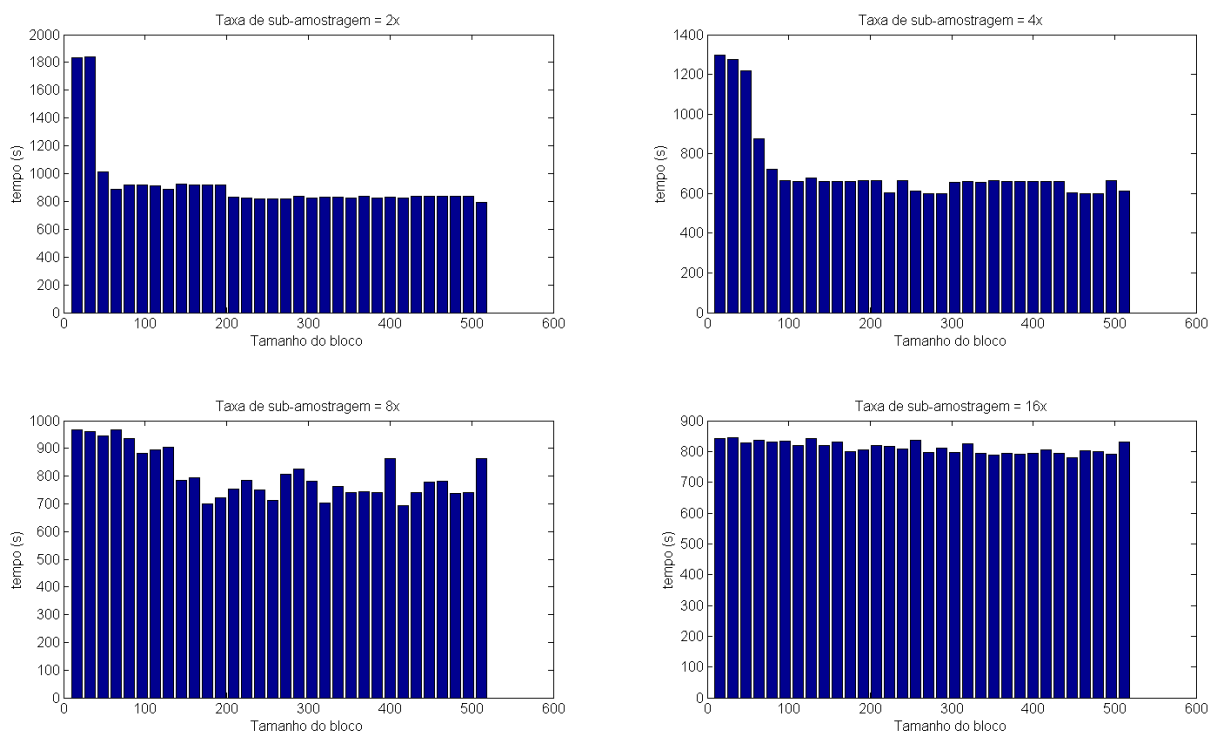


Figura 4.21. Tempos de reconstrução dos testes de reconstrução do PPM2 ao variar o tamanho do bloco.

Peculiarmente, o comportamento do tempo de execução é muito parecido com o comportamento da qualidade da reconstrução. Nos casos em que a qualidade de reconstrução tem o pior desempenho, o tempo de reconstrução também tem o pior desempenho (menores tamanhos de bloco). A partir de um certo tamanho de bloco (que varia de acordo com a taxa de sub-amostragem), a qualidade e tempo de reconstrução se estabilizam.

Isso reforça a ideia de que ao dividir o vídeo em blocos muito pequenos o ganho em tempo não é muito. Só é interessante dividir em blocos em relação tanto a qualidade e tempo de reconstrução quando o tamanho do bloco é grande, caso em que não há nenhuma

Tabela 4.7. Average SER (dB): reconstructed natural videos.

Taxa	4x	8x	16x
PPS2	1760	1313	1096
PPM2	1776	1499	1178
PPM-2D	1499	1173	976
PPM3	1499+1373=2871	1173+1139=2316	976+1015=1991

vantagem aparente. E mesmo assim, o desempenho apenas se equipara ao caso em que não há divisão em blocos.

Vídeos Naturais

Fizemos testes com 12 vídeos avaliando a qualidade de reconstrução dos métodos PPS2, PPM2 e PPM3, que é baseada no PPM2 com inicialização dos quadros a partir da TV2D. Nesta Seção avaliamos o tempo da reconstrução desses métodos aplicados a vídeos naturais. No caso do PPM3, foi adicionado ao tempo de reconstrução o tempo tomado para a inicialização dos quadros. Os resultados podem ser vistos na Figura 4.22, e a média dos resultados dos vídeos na Tabela 4.7.

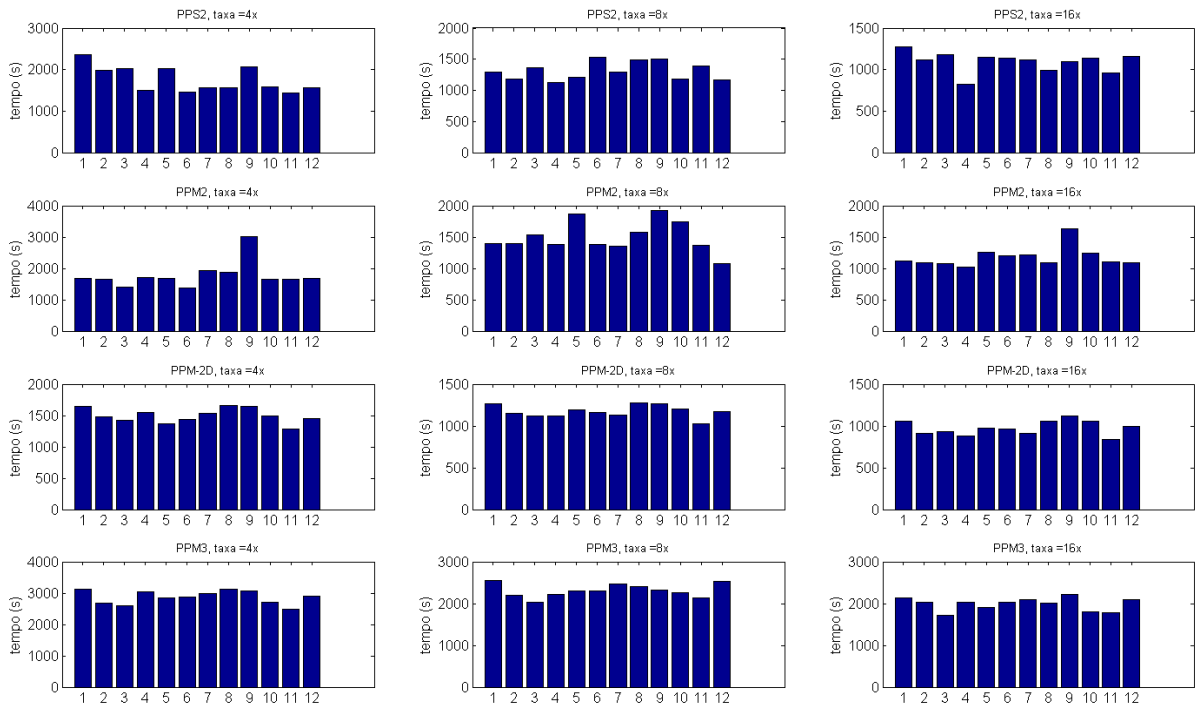


Figura 4.22. Tempos de reconstrução dos testes de reconstrução 3D para vídeos naturais para os métodos PPS2, PPM2, PPM3 e PPM-2D.

Um comportamento que podemos observar é que o tempo diminui ao aumentar a taxa de sub-amostragem. Os métodos PPS2 e PPM2 apresentaram melhores tempos de reconstrução que o PPM3, pois a solução inicial não é obtida por vários passos de um algoritmo de reconstrução. O PPM-2D é de 10 a 30% mais rápido do que o PPS2 e o PPM2, e isto se deve ao fato de termos 16 reconstruções 2D ao invés de 1 reconstrução 3D 16 vezes maior. O PPM3 foi mais lento porque ele tem como entrada o resultado de reconstrução do PPM-2D. Sendo assim, para o PPM3, além do tempo de reconstrução, contabilizamos também o tempo da reconstrução da entrada. Observa-se, entretanto que o tempo de reconstrução sem a inicialização é menor do que os tempos para o PPS2 e o PPM2 e equivalente ao do PPM-2D. Concluímos que o PPM-2D apresenta um ótimo resultado de reconstrução em termos do tempo de execução. Em aplicações que exigem menos tempo, o PPM-2D pode ser a alternativa mais viável.

4.5 Discussão dos Resultados

Fizemos diversas simulações de reconstrução a partir de diversos métodos de amostragem, inclusos os da literatura e o método proposto e suas variações. Avaliamos, principalmente, a qualidade dos vídeos reconstruídos a partir dos métodos de aquisição.

Para a reconstrução de vídeos com apenas 1 quadro (imagem), o método PPM mostrou o melhor resultado na qualidade de reconstrução (SER entre 5 e 8 dB acima) em relação ao outro método testada (PPS1). Nos casos em que a resolução espacial é maior, o resultado é ainda melhor. Para a maior resolução, a reconstrução dos quadros separados é boa o suficiente para taxas de sub-amostragem de 4x.

Foram feitos também testes para vídeos sintetizados bastante esparsos no domínio espacial e no temporal. Testamos com todos os métodos de aquisição, e os resultados mostram que o PPM2 obteve sempre as melhores qualidades de reconstrução, tanto para o teste de linhas pelo tempo, como para a reconstrução 3D. A diferença em dB foi grande, mostrando que para estas condições (vídeos esparsos no domínio TV e com maior resolução espacial) conseguimos qualidades de reconstrução com o PPM2 com defeitos visualmente imperceptíveis, mesmo a uma taxa de sub-amostragem de fator 16x. Isto significa que, se o conteúdo do vídeo for suficientemente esparsos, a partir de medidas de uma câmera que implementa o PPM2, poderíamos obter um vídeo com o conteúdo perceptualmente intacto e 16 vezes mais rápido que a taxa de captura da câmera. Isto é algo que não conseguimos com os outros métodos para o mesmo conteúdo de vídeo.

Porém, nem sempre temos o caso de esparsidade ideal. Para isto testamos a robustez ao ruído. Este ruído adicionado serve como teste tanto para medir a influência do ruído de medida no resultado, como para tornar o vídeo menos esparsos no domínio TV. Todos

os métodos sofreram uma queda de qualidade da reconstrução ao adicionarmos um pouco de ruído às medidas. O método proposto não apresentou um bom comportamento para maiores níveis de ruído, enquanto que o PPS2 (método que preserva a posição espacial das medidas, mas tem dependência temporal) manteve a qualidade nos maiores níveis de ruído. Por outro lado, a qualidade do PPM1 e do PPM2 (independentes temporalmente, porém dependentes espacialmente) caíram rapidamente.

Os vídeos naturais não são realmente muito esparsos no domínio TV. Nas simulações para vídeos naturais, o método PPM2 obteve resultados piores do que o PPS2 na maioria dos casos (em alguns poucos casos, nos vídeos de maior movimentação, o PPM2 obteve melhores SER), quando a taxa de sub-amostragem é maior. Porém, ao introduzir a reconstrução quadro a quadro como solução inicial para o método proposto (PPM3), o resultado ficou muito melhor em todos os casos, com um SER de cerca de 70% maior. De fato, a própria reconstrução 2D quadro a quadro já fornece um resultado muito bom, com um SER equivalente, o qual foi melhorado utilizando a TV 3D em alguns poucos casos, pelo menos visualmente.

Analisamos também o tempo de execução. Em geral, o algoritmo de reconstrução teve tempo de execução aproximadamente linear com o tamanho do vídeo. De forma geral, houve pouca variação quanto ao tempo de execução a partir dos diferentes os métodos de amostragem. A maior diferença foi para o PPM3, pois este método reconstrói duas vezes, levando aproximadamente o dobro do tempo que as outros métodos. Em termos de qualidade, o PPM3 apresentou um desempenho parecido com a reconstrução 2D. Então consideramos que o método PPM-2D apresenta o melhor compromisso entre tempo e qualidade de reconstrução.

Capítulo 5

Conclusão e Trabalhos Futuros

Nesta dissertação focamos em desenvolver e avaliar comparativamente um novo método de aquisição de vídeo com sensores comuns, com o objetivo de reconstruir vídeos de alta velocidade, denominado *per-pixel mirror-based measuring* (PPM). Neste capítulo, apresentamos as nossas conclusões sobre o trabalho desenvolvido, incluindo as principais contribuições do método proposto. Apresentamos também uma análise das limitações do nosso método e sugestões para trabalhos futuros.

5.1 Conclusão

Na introdução discutimos os problemas dos métodos de aquisição para vídeo de alta velocidade na literatura. Entre os problemas mais relevantes, estão a baixa eficiência luminosa e a dependência temporal entre as medidas. Nossa hipótese inicial é que uma técnica de aquisição que apresente vantagens em relação a esses quesitos pode levar a um melhor resultado de reconstrução. Com isto em vista, propomos um novo método de reconstrução. O método PPM, mistura espacialmente as intensidades da luz por espelhos, de tal forma que nenhuma luz é jogada fora. A luz é mantida separada para cada subquadro distinto, o que garante a propriedade de independência temporal. O método ainda permite algumas variações em sua configuração, que permitem maior incoerência, e até super-resolução.

Fizemos simulações em que testamos os métodos de medidas para vídeo de alta velocidade existentes na literatura e o método proposto. Na maioria dos casos, o método proposto obteve melhor qualidade de reconstrução. Os testes consistiam em obter uma quantidade reduzida de medidas de um vídeo de alta velocidade a partir dos diversos métodos de aquisição. Os vídeos foram reconstruídos com o mesmo algoritmo (minimização *total variation*) de reconstrução para todos os métodos. Com o resultado da reconstrução

em mãos, o comparamos com o vídeo de alta velocidade original com a métrica *signal-to-error ratio* (SER) e avaliamos a qualidade da reconstrução.

Primeiramente, testamos os métodos utilizando vídeos sintetizados que são muito esparsos no domínio TV, favorecendo a reconstrução. Para diversos tamanhos de vídeo, para algumas variações do algoritmo de reconstrução 2D e 3D e para diversas taxas de sub-amostragem, o método proposto apresentou um resultado melhor em relação aos outros métodos, tanto visualmente quanto em SER. Porém, no teste em que adicionamos ruído às medidas, a técnica não mostrou tal vantagem.

Utilizamos também vídeos reais nos nossos testes para um conjunto reduzido de cenários: testamos apenas a segunda melhor técnica da literatura (PPS2) com a nossa proposta. O PPS2 mostrou melhores resultados do que o PPM2 para vídeos com baixo movimento, enquanto que o PPM2 apresentou resultados semelhantes ao PPS2 para vídeos com maior quantidade de movimento. Ao verificarmos que nossa técnica não divergia muito da solução inicial, resolvemos utilizar a solução 2D dos quadros individuais como solução inicial. Utilizamos parte das medidas para reconstruir cada quadro separadamente (PPM-2D). Ao final, o resultado ficou muito bom, com menos ruído e sem dependência de conteúdo. A reconstrução 3D a partir dessa solução (PPM3) ficou com qualidade muito próxima, porém não houve melhora em relação ao PPM-2D nos casos de maior movimento. Escolhemos o PPM-2D e o PPM3 como a melhores métodos, pois seus resultados apresentam pouco ruído e se comportam bem com movimento. Quando o vídeo apresenta um pouco mais de movimento, o PPM-2D apresenta um desempenho um pouco melhor. Além disso, o PPM-2D mostrou-se a mais rápido.

Contribuição

Desenvolvemos o PPM, uma nova forma de aquisição para *compressive sensing* de vídeo de alta velocidade que produz medidas temporalmente independentes, e é luminosamente eficiente. O método proposto apresentou resultados muito bons quando comparado aos outros métodos disponíveis na literatura. O trabalho desta dissertação será apresentado na conferência EUSIPCO 2014 que acontecerá em Lisboa em setembro [33].

Ao invés de se jogar amostras fora, o PPM forma as medidas a partir da soma de amostras de posições distintas. Isto pode levar a um novo modo de adquirir medidas em sistemas de *compressive sensing*. Geralmente, em *compressive sensing*, as medidas são alguns coeficientes em algum domínio transformado de um sinal. Mas existem casos em que temos o sinal completo e fazemos o descarte das medidas. Nestes casos, ao invés de descartar as medidas, podemos obtê-las como no PPM, somando amostras no lugar de descartá-las. Vimos que para esse tipo de amostragem proposto, utilizando a DCT como domínio transformado, apresenta uma incoerência pelo menos tão boa quanto ao descarte

de amostras para sinais unidimensionais de tamanho 100 (não pudemos testar para sinais maiores). Isso dá o indicativo de que o PPM pode ser utilizada também para outras aplicações de *compressive sensing* na qual não descartamos medidas.

Limitações

A principal limitação do método proposto é que ainda não existe o HPDMD, o dispositivo físico utilizado na implementação do PPM. É preciso ainda averiguar se é possível adaptar o dispositivo DMD de tal forma que ele possa ter uma resolução suficiente para direcionar a luz de um espelho para outros pixels.

O método proposto apresenta independência temporal, mas não dependência espacial. Com isso, a luz de posições espaciais distintas é misturada em suas medidas. Sendo assim, as medidas perdem sua posição espacial e não é possível visualizar uma representação em imagem dos quadros reconstruídos, como nos outros métodos. Além disso, alguns casos, quando existe maior ruído nas medidas, a qualidade da reconstrução diminui, chegando a níveis de SER próximos aos obtidos a partir dos outros métodos.

5.2 Trabalhos Futuros

5.2.1 Outras Variações do PPM

O método proposto é adaptável o suficiente para apresentar variações em sua configuração, das quais pudemos testar algumas. Porém existem ainda mais variações a serem testadas. Entre as variações, estamos considerando avaliar métodos de aquisição híbridos e aquisição para super-resolução.

Métodos Híbridos

Embora o PPM tenha mostrado resultados muito bons, este resultado não pode ser generalizado. Por exemplo, nos casos em que há muito ruído nas medidas o seu desempenho cai consideravelmente. O dispositivo HPDMD permite implementar todos os métodos de aquisição para vídeos de alta velocidade mostrados neste trabalho. Uma possível extensão deste trabalho consiste em propor métodos híbridos que utilizem parte das medidas de um dos métodos, e outra parte das medidas de outro método. Assim podemos aproveitar as vantagens de dois ou mais métodos.

Super-resolução

Na seção 3.3.2 vimos a possibilidade do método proposto ser utilizada para super-resolução. Fizemos testes equivalentes ao que seriam ao reconstruir os sinais em uma resolução espacial maior, como no teste de imagens (na Seção 4.1) e na reconstrução das soluções iniciais no PPM-2D (Seção 4.3). Surpreendentemente, os resultados foram satisfatórios. Logo, um dos próximos passos inclui testar o método proposto com o objetivo específico de obter super-resolução, e comparar com os outros métodos utilizados para o mesmo fim. Além disso, observamos que a implementação em hardware do dispositivo proposto para a aplicação de super-resolução é mais simples que para a aquisição de vídeo de alta velocidade, o que torna a aplicação atrativa.

5.2.2 Reconstrução

Até agora exploramos apenas a reconstrução por minimização da variação total no domínio espacial. Próximos passos incluem testar outros tipos de reconstrução. Entre eles cogitamos inicialmente a reconstrução por treinamento de dicionário e por minimização em outros domínios.

Reconstrução por Treinamento de Dicionário

Uma outra forma de reconstrução é por treinamento de dicionário, para representação esparsa [2, 25, 35, 26]. As imagens naturais representadas nas bases como DCT ou *wavelet* não são exatamente esparsas, ou seja, possuem muitos coeficientes com baixa energia. O ideal é buscarmos uma base que represente as imagens e vídeos de tal forma que os sinais sejam exatamente esparsos, ou seja, com poucos coeficientes de alta energia e que a grande maioria dos coeficientes seja nulo.

Podemos então treinar um dicionário com uma grande quantidade de imagens e vídeos. Assim, qualquer vídeo natural poderia ser representado por uma combinação linear das bases desse dicionário. O vídeo seria reconstruído pela substituição de cada bloco do vídeo pelo item do dicionário que melhor o represente. Essa representação seria exatamente esparsa, pois apenas um número pequeno de blocos seriam escolhidos. Além de que qualquer vídeo teria a mesma esparsidade de qualquer outro de mesmo tamanho, pois os coeficientes que correspondem aos blocos não escolhidos seriam todos iguais a zero.

Reconstrução por Minimização no Domínio Transformado

Uma outra ideia para reconstrução seria utilizar medidas neste mesmo domínio espaço-temporal, mas reconstruir em algum domínio transformado. O algoritmo de minimização

escolhido minimiza o problema equivalente a:

$$\hat{s} = \operatorname{argmin}_{s'} (\|s'\|_1) \text{ sujeito a } y = \Phi\Psi s', \quad (5.1)$$

com $\Psi = D_i^{-1}$. Podemos escolher um algoritmo diferente que, ao invés de minimizar a TV, minimize a métrica ℓ_1 de algum outro domínio transformado. Desta forma, a matriz Ψ pode ser escolhida com mais liberdade e assim podemos escolher uma que leve a maiores valores de incoerência.

5.2.3 Implementação em Hardware

Para que o PPM seja realizável nas câmeras comerciais, é necessário que o dispositivo HPDMD esteja implementado. A ideia da implementação da HPDMD surge a partir de um dispositivo DMD. Em trabalhos futuros, podemos observar o dispositivo DMD e verificar qual é a precisão dos dispositivos mecânicos que alteram a angulação dos espelhos. Com isso podemos avaliar se é possível atingir o nível de precisão necessária para a implementação do HPDMD. Caso contrário, podemos cogitar utilizar outras tecnologias na implementação de um dispositivo equivalente ao HPDMD.

Referências

- [1] Amit Agrawal, Mohit Gupta, Ashok Veeraraghavan, and Srinivasa G Narasimhan. Optimal coded sampling for temporal super-resolution. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 599–606. IEEE, 2010. 5, 6, 25, 45
- [2] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, 2006. 90
- [3] Richard G Baraniuk. Compressive sensing [lecture notes]. *Signal Processing Magazine, IEEE*, 24(4):118–121, 2007. 13, 15
- [4] Glenn Baxter, Steven Frisken, Dmitri Abakoumov, Hao Zhou, Ian Clarke, Andrew Bartos, and Simon Poole. Highly programmable wavelength selective switch based on liquid crystal on silicon switching elements. In *Optical Fiber Communication Conference*, page OTuF2. Optical Society of America, 2006. 26
- [5] Gil Bub, Matthias Tecza, Michiel Helmes, Peter Lee, and Peter Kohl. Temporal pixel multiplexing for simultaneous high-speed, high-resolution imaging. *Nature methods*, 7(3):209–211, 2010. 25
- [6] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006. 13, 15
- [7] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006. 13, 20, 21
- [8] Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52(12):5406–5425, 2006. 13
- [9] Wai Lam Chan, Kriti Charan, Dharmpal Takhar, Kevin F Kelly, Richard G Baraniuk, and Daniel M Mittleman. A single-pixel terahertz imaging system based on compressed sensing. *Applied Physics Letters*, 93(12):121105–121105, 2008. 6
- [10] Wai Lam Chan, Hou-Tong Chen, Antoinette J Taylor, Igal Brener, Michael J Cich, and Daniel M Mittleman. A spatial light modulator for terahertz beams. *Applied Physics Letters*, 94(21):213511, 2009. 27

- [11] Hou-Tong Chen, Willie J Padilla, Joshua MO Zide, Arthur C Gossard, Antoinette J Taylor, and Richard D Averitt. Active terahertz metamaterial devices. *Nature*, 444(7119):597–600, 2006. 27
- [12] Roger N. Clark. Digital camera sensor performance summary. <http://www.clarkvision.com/articles/digital.sensor.performance.summary>. Acessado em Março, 2013. 3
- [13] Andrew Davidhazy. Basics of focusing schlieren systems. 1998. 2
- [14] David Leigh Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006. 13, 15
- [15] Michael Douglass. Dmd reliability: a mems success story. In *Micromachining and Microfabrication*, pages 1–11. International Society for Optics and Photonics, 2003. 25, 30
- [16] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk. Single-pixel imaging via compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):83–91, 2008. 5, 6
- [17] Dana Dudley, Walter M Duncan, and John Slaughter. Emerging digital micromirror device (dmd) applications. In *Micromachining and Microfabrication*, pages 14–25. International Society for Optics and Photonics, 2003. 27
- [18] Torsten Edeler, Kevin Ohliger, Stephan Hussmann, and Alfred Mertins. Super-resolution model for a compressed-sensing measurement setup. *Instrumentation and Measurement, IEEE Transactions on*, 61(5):1140–1148, 2012. ix, 30, 45
- [19] Michael Elad. *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer, 2010. 17
- [20] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T Roweis, and William T Freeman. Removing camera shake from a single photograph. In *ACM Transactions on Graphics (TOG)*, volume 25, pages 787–794. ACM, 2006. 5
- [21] Rafael C Gonzalez, Richard E Woods, and Steven L Eddins. *Digital image processing using MATLAB*, volume 2. Gatesmark Publishing Tennessee, 2009. ix, 22
- [22] Exposure guide. Exposure. <http://www.exposureguide.com/exposure.htm>. Acessado em Julho, 2014. 24
- [23] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G Narasimhan. Flexible voxels for motion-aware videography. In *Computer Vision–ECCV 2010*, pages 100–114. Springer, 2010. 5, 6, 25, 31
- [24] Zachary T Harmany, Roummel F Marcia, and Rebecca M Willett. Spatio-temporal compressed sensing with coded apertures and keyed exposures. *arXiv preprint arXiv:1111.7247*, 2011. 5

- [25] Yasunobu Hitomi, Jinwei Gu, Mohit Gupta, Tomoo Mitsunaga, and Shree K Nayar. Video from a single coded exposure photograph using a learned over-complete dictionary. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 287–294. IEEE, 2011. 5, 7, 22, 25, 31, 90
- [26] J Holloway, AC Sankaranarayanan, A Veeraraghavan, and S Tambe. Flutter shutter video camera for compressive sensing of videos. In *Computational Photography (ICCP), 2012 IEEE International Conference on*, pages 1–9. IEEE, 2012. 5, 6, 25, 28, 47, 90
- [27] Photography – Determination of ISO speed, 2006. 23
- [28] Kristina M Johnson, Douglas J McKnight, and Ian Underwood. Smart spatial light modulators using liquid crystals on silicon. *Quantum Electronics, IEEE Journal of*, 29(2):699–714, 1993. 26
- [29] Chengbo LI. Tval3: Tv minimization by augmented lagrangian and alternating direction algorithms. <http://www.caam.rice.edu/~optimization/L1/TVAL3/>. Accessed em Março, 2014. 47
- [30] Chengbo Li. *An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing*. PhD thesis, Citeseer, 2009. 20, 47
- [31] Chengbo Li. *Compressive sensing for 3D data processing tasks: applications, models and algorithms*. PhD thesis, Rice University, 2011. 48
- [32] Chengbo Li, Wotao Yin, and Yin Zhang. User’s guide for tval3: Tv minimization by augmented lagrangian and alternating direction algorithms. *CAAM Report*, 2009. 20, 46, 47
- [33] Jonathan A. Lima, Cristiano J. Miosso, and Mylène C.Q. Farias. Per-pixel mirror-based acquisition method for video compressive sensing. In *European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal. 88
- [34] Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic resonance in medicine*, 58(6):1182–1195, 2007. 16, 17
- [35] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 689–696. ACM, 2009. 90
- [36] Roummel Marcia and Rebecca M Willett. Compressive coded aperture video reconstruction. In *Proc. European Signal Processing Conf.(EUSIPCO)*, 2008. 5
- [37] Douglas J McKnight, Kristina M Johnson, and Royinn A Serati. 256×256 liquid-crystal-on-silicon spatial light modulator. *Applied Optics*, 33(14):2775–2784, 1994. 26

- [38] Sanjit Kumar Mitra and Yonghong Kuo. *Digital signal processing: a computer-based approach*, volume 2. McGraw-Hill New York, 2006. 10, 13
- [39] Shree K Nayar, Vlad Branzoi, and Terry E Boult. Programmable imaging: Towards a flexible camera. *International Journal of Computer Vision*, 70(1):7–22, 2006. 25
- [40] Mark Nelson. *Data compression book*. IDG Books Worldwide, Inc., 1991. 14, 20
- [41] Nokia. Nokia 808 pureview. <http://www.nokia.com/global/products/phone/808pureview/>. Acessado em Março, 2013. 4
- [42] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE*, 20(3):21–36, 2003. 44
- [43] Vishal M Patel, Glenn R Easley, Dennis M Healy Jr, and Rama Chellappa. Compressed synthetic aperture radar. *Selected Topics in Signal Processing, IEEE Journal of*, 4(2):244–254, 2010. 16
- [44] Phonerated. Top rated best overall camera phones global. <http://www.phonerated.com/top-rated-best-overall-camera-phones-global>. Acessado em Março, 2013. 4
- [45] Photron. Photron. <http://www.photron.com/>. Acessado em Março, 2013. ix, 2, 3, 4
- [46] Tomaso Poggio and Vincent Torre. *Ill-posed problems and regularization analysis in early vision*. Massachusetts Institute of Technology, 1984. 15
- [47] Dikpal Reddy, Ashok Veeraraghavan, and Rama Chellappa. P2c2: Programmable pixel compressive camera for high speed imaging. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 329–336. IEEE, 2011. 4, 5, 7, 25, 31, 45
- [48] William E Ross, Demetri Psaltis, and Robert H Anderson. Two-dimensional magneto-optic spatial light modulator for signal processing. *Optical Engineering*, 22(4):224485–224485, 1983. 26
- [49] William E Ross, Demetri Psaltis, and Robert H Anderson. Two-dimensional magneto-optic spatial light modulator for signal processing. *Optical Engineering*, 22(4):224485–224485, 1983. 26
- [50] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992. 46
- [51] Aswin C Sankaranarayanan, Pavan K Turaga, Richard G Baraniuk, and Rama Chellappa. Compressive acquisition of dynamic scenes. In *Computer Vision–ECCV 2010*, pages 129–142. Springer, 2010. 5, 6

- [52] Dharmpal Takhar, Jason N Laska, Michael B Wakin, Marco F Duarte, Dror Baron, Shriram Sarvotham, Kevin F Kelly, and Richard G Baraniuk. A new compressive imaging camera architecture using optical-domain compression. In *Electronic Imaging 2006*, pages 606509–606509. International Society for Optics and Photonics, 2006. 5, 26
- [53] Ashok Veeraraghavan, Dikpal Reddy, and Ramesh Raskar. Coded strobing photography: Compressive sensing of high speed periodic videos. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(4):671–686, 2011. 6, 25
- [54] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Mounsi G Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3:745, 2012. 2, 4
- [55] Michael Wakin, Jason Laska, Marco Duarte, Dror Baron, Shriram Sarvotham, Dharmpal Takhar, Kevin Kelly, and Richard G Baraniuk. Compressive imaging for video representation and coding. In *Picture Coding Symposium*, 2006. 5, 6
- [56] Gregory K Wallace. The jpeg still picture compression standard. *Communications of the ACM*, 34(4):30–44, 1991. 69
- [57] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Marc Levoy, and Mark Horowitz. High-speed videography using a dense camera array. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–294. IEEE, 2004. 4
- [58] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy. High performance imaging using large camera arrays. *ACM Transactions on Graphics*, 24(3):765–776, 2005. 4
- [59] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *Image Processing, IEEE Transactions on*, 19(11):2861–2873, 2010. 45