



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Transmissão Progressiva de Imagens Sintetizadas de Light Field

Wallace Bruno Silva de Souza

Dissertação apresentada como requisito parcial
para conclusão do Mestrado em Informática

Orientador

Prof. Dr. Bruno Luigi Machiavello

Brasília

2018

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

Coordenador: Prof. Dr. Coordenador Bruno Luiggi Machiavello

Banca examinadora composta por:

Prof. Dr. Bruno Luiggi Machiavello (Orientador) — CIC/UnB

Prof.^a Dr.^a Mylène C. Q. Farias — ENE/UnB

Prof.^a Dr.^a Carla L. Pagliari — DCT/IME

CIP — Catalogação Internacional na Publicação

Souza, Wallace Bruno Silva de.

Transmissão Progressiva de Imagens Sintetizadas de Light Field / Wallace Bruno Silva de Souza. Brasília : UnB, 2018.

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

Dissertação (Mestrado) — Universidade de Brasília, Brasília, 2018.

1. Light Field, 2. Transmissão, 3. Compressão, 4. Quantização,
5. Imagens Sintetizadas

CDU 004.4

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

Transmissão Progressiva de Imagens Sintetizadas de Light Field

Wallace Bruno Silva de Souza

Dissertação apresentada como requisito parcial
para conclusão do Mestrado em Informática

Prof. Dr. Bruno Luigi Machiavello (Orientador)
CIC/UnB

Prof.^a Dr.^a Mylène C. Q. Farias Prof.^a Dr.^a Carla L. Pagliari
ENE/UnB DCT/IME

Prof. Dr. Coordenador Bruno Luigi Machiavello
Coordenador do Mestrado em Informática

Brasília, 25 de julho de 2018

Dedicatória

À vida, à saúde e ao privilégio de poder aproveitar cada instante. À alegria dos bons momentos e à experiência proporcionada pelos momentos mais difíceis. À minha família abençoada.

Agradecimentos

Agradeço, máxime, a Deus por me proteger e por me dar sabedoria em tudo que faço ou penso em fazer. Agradeço a minha família pelo amor e carinho imensuráveis, pela compreensão, pelo incentivo e por sempre estar do meu lado: Priscilla, Maria, Marcondes e Matheus. Obrigado por me ajudarem a conquistar mais esta grande vitória.

A meu amigo Bruno, que foi meu mentor na UnB, sempre disposto a ensinar e também a aprender. Ao professor Zaghetto que, juntamente com o Bruno, me proporcionou ótimas descontrações no meio das inúmeras reuniões. Ao professor Mintsu pela boa vontade, executando uma parte da grande quantidade de testes que eu precisei realizar.

A meu amigo Silas que esteve por perto e me motivou mesmo quando ele não tinha a intenção. A meu amigo Clerimar por sua parceria. Aos meus amigos Cristiano Aveiro, Ezio, Schmidt, Ítalo, Losi, Rodopoulos e Santaguida. Ademais, agradeço a todos que também participaram de alguma maneira desta etapa da minha vida.

“Nada é verdade, tudo é permitido.”
(Bartol, V.)

“Busquem conhecimento.”
(Bilu, E.T.)

Resumo

Esta proposta estabelece um método otimizado baseado em taxa-distorção para transmitir imagens sintetizadas de *light field*. Resumidamente, uma imagem *light field* pode ser interpretada como um dado quadridimensional (4D) que possui tanto resolução espacial, quanto resolução angular, sendo que cada subimagem bidimensional desse dado 4D é tido como uma determinada perspectiva, isto é, uma imagem de subabertura (SAI, do inglês *Sub-Aperture Image*). Este trabalho visa modificar e aprimorar uma proposta anterior chamada de Comunicação Progressiva de *Light Field* (PLFC, do inglês *Progressive Light Field Communication*), a qual trata da sintetização de imagens referentes a diferentes focos requisitados por um usuário. Como o PLFC, este trabalho busca fornecer informação suficiente para o usuário de modo que, conforme a transmissão avance, ele tenha condições de sintetizar suas próprias imagens de ponto focal, sem a necessidade de se enviar novas imagens. Assim, a primeira modificação proposta diz respeito à como escolher a *cache* inicial do usuário, determinando uma quantidade ideal de imagens de subabertura para enviar no início da transmissão. Propõe-se também um aprimoramento do processo de seleção de imagens adicionais por meio de um algoritmo de refinamento, o qual é aplicado inclusive na inicialização da *cache*. Esse novo processo de seleção lida com QPs (Passo de Quantização, do inglês *Quantization Parameter*) dinâmicos durante a codificação e envolve não só os ganhos imediatos para a qualidade da imagem sintetizada, mas ainda considera as sintetizações subsequentes. Tal ideia já foi apresentada pelo PLFC, mas não havia sido implementada de maneira satisfatória. Estabelece-se ainda uma maneira automática para calcular o multiplicador de Lagrange que controla a influência do benefício futuro associado à transmissão de uma SAI. Por fim, descreve-se um modo simplificado de obter esse benefício futuro, reduzindo a complexidade computacional envolvida. Muitas são as utilidades de um sistema como este, podendo, por exemplo, ser usado para identificar algum elemento em uma imagem *light field*, ajustando apropriadamente o foco em questão. Além da proposta, os resultados obtidos são exibidos, sendo feita uma discussão acerca dos significativos ganhos conseguidos de até 32,8% com relação ao PLFC anterior em termos de BD-Taxa. Esse ganho chega a ser de até 85,8% em comparação com transmissões triviais de dados *light field*.

Palavras-chave: Light Field, Transmissão, Compressão, Quantização, Imagens Sintetizadas

Abstract

This work proposes an optimized rate-distortion method to transmit light field synthesized images. Briefly, light field images could be understood like quadridimensional (4D) data, which have both spatial and angular resolution, once each bidimensional subimage in this 4D image is a certain perspective, that is, a SAI (Sub-Aperture Image). This work aims to modify and to improve a previous proposal named PLFC (Progressive Light Field Communication), which addresses the image synthesis for different focal point images requested by an user. Like the PLFC, this work tries to provide enough information to the user so that, as the transmission progress, he can synthesize his own focal point images, without the need to transmit new images. Thus, the first proposed modification refers to how the user's initial cache should be chosen, defining an ideal amount of SAIs to send at the transmission beginning. An improvement of the additional images selection process is also proposed by means of a refinement algorithm, which is applied even in the cache initialization. This new selection process works with dynamic QPs (Quantization Parameter) during encoding and involves not only the immediate gains for the synthesized image, but either considers the subsequent synthesis. This idea already was presented by PLFC, but had not been satisfactorily implemented. Moreover, this work proposes an automatic way to calculate the Lagrange multiplier which controls the influence of the future benefit associated with the transmission of some SAI. Finally, a simplified manner of obtaining this future benefit is then described, reducing the computational complexity involved. The utilities of such a system are diverse and, for example, it can be used to identify some element in a light field image, adjusting the focus accordingly. Besides the proposal, the obtained results are shown, and a discussion is made about the significant achieved gains up to 32.8% compared to the previous PLFC in terms of BD-Rate. This gain is up to 85.8% in relation to trivial light field data transmissions.

Keywords: Light Field, Transmission, Compression, Quantization, Synthesized Images

Sumário

| | | |
|----------|--|-----------|
| 1 | Introdução | 1 |
| 1.1 | Descrição do problema | 2 |
| 1.2 | Objetivo geral | 4 |
| 1.3 | Objetivos específicos | 4 |
| 1.4 | Organização | 4 |
| 2 | Fundamentação Teórica | 6 |
| 2.1 | Câmeras convencionais | 6 |
| 2.2 | Câmeras <i>Light Field</i> | 9 |
| 2.2.1 | <i>Slope</i> | 11 |
| 2.3 | Imagens <i>Light Field</i> | 12 |
| 2.3.1 | Imagem integral | 16 |
| 2.3.2 | Mudança de perspectiva e de foco | 17 |
| 2.4 | Técnicas básicas de predição para a compressão de imagens e vídeos | 23 |
| 2.5 | H.264 | 24 |
| 2.6 | Padronização da compressão de <i>light field</i> | 25 |
| 3 | Revisão Bibliográfica | 26 |
| 4 | Metodologia | 30 |
| 4.1 | Comunicação Progressiva de Light Field | 30 |
| 4.1.1 | Usando apenas a <i>cache</i> | 32 |
| 4.1.2 | Selecionando novas imagens para transmissão | 32 |
| 4.1.3 | O sistema aprende a SAI | 33 |
| 4.2 | Método Proposto | 33 |
| 4.2.1 | Inicialização da <i>cache</i> | 35 |
| 4.2.2 | Algoritmo de Refinamento para seleção de SAIs | 37 |
| 4.2.3 | Seleção dinâmica de QP / contribuição de qualidade | 38 |
| 4.2.4 | Eliminação da combinação linear esparsa para o cálculo do benefício futuro | 38 |
| 4.2.5 | Cômputo automático do multiplicador de Lagrange, μ | 39 |
| 5 | Resultados | 40 |
| 5.1 | Banco de imagens | 40 |
| 5.2 | Métrica de avaliação de resultados | 43 |
| 5.3 | Configurações dos testes | 43 |
| 5.4 | Resultados Obtidos | 44 |

| | | |
|----------|---|-----------|
| 5.4.1 | Taxa acumulada | 44 |
| 5.4.2 | Taxa distorção | 49 |
| 5.5 | Discussão de resultados | 51 |
| 6 | Conclusões | 56 |
| | Referências | 58 |
| A | Apêndice A - Resultados Adicionais | 63 |
| A.1 | Taxa acumulada | 63 |
| A.2 | Taxa distorção | 75 |

Lista de Figuras

| | | |
|-----|---|---|
| 1.1 | Representação simplificada em duas dimensões de como a imagem de um ponto A é capturada pelo sensor de Lipmann, por meio de uma matriz de pontos. Percebe-se que cada microlente captura uma imagem a na mesma posição relativa. Imagem de [1]. | 1 |
| 1.2 | Ilustração de uma imagem sintetizada borrada, fora de foco. Imagem nítida sintetizada com profundidade expandida. Adaptada de [2]. | 3 |
| 1.3 | Imagens sintetizadas referentes a diferentes focos. Em (c) a imagem sintetizada com profundidade expandida é obtida pela combinação entre os dois focos anteriores [3], focalizando os dois planos ao mesmo tempo. Imagens originais de [2]. | 3 |
| 2.1 | Funcionamento de uma câmera convencional. Os raios luminosos capturados a partir de um ponto de um determinado objeto são mapeados para um único ponto da imagem. Se esse objeto estiver em foco, os raios são focalizados em um mesmo ponto do plano sensor. Adaptada de [4]. | 6 |
| 2.2 | Formação de um <i>pixel</i> em uma fotografia. O cone de luz representado em (a) intercepta o plano focal x em um mesmo ponto e é mapeado para o diagrama de raios como a tira vertical em (b). Essa tira corresponde a soma de todas as posições de u , gerando um <i>pixel</i> da imagem. Adaptada de [2]. | 7 |
| 2.3 | Mudança de foco em uma câmera convencional por meio da mudança de posição do plano focal. (a) Objetos mais próximos da câmera são focalizados, sendo que a imagem se forma depois do plano focal original x . No diagrama mais abaixo, o movimento dos raios através de u possui a mesma direção do movimento através do plano x . (b) A imagem se forma antes de x , o foco está em objetos mais distantes da câmera. No diagrama de raios, a direção do movimento das interseções no plano u é oposta com relação a direção dos raios em x . Adaptada de [2]. | 8 |
| 2.4 | Ilustração de câmeras <i>light field</i> Lytro. A câmera mostrada em (a) foi a primeira câmera <i>light field</i> voltada para o consumidor. As imagens utilizadas nos testes deste trabalho foram capturadas por uma câmera Lytro Illum, como a exibida em (b). Imagens originais de [5]. | 9 |
| 2.5 | Esquema bidimensional representativo de uma câmera <i>light field</i> . Percebe-se que os raios luminosos são capturados por sensores situados logo atrás da matriz de microlentes. Raios de diferentes direções atingem o plano sensor em diferentes <i>pixels</i> . Adaptada de [6]. | 9 |

| | | |
|------|---|----|
| 2.6 | Esquema ilustrativo dos diferentes <i>f-numbers</i> para a lente principal e para as microlentes. Quando o <i>f-number</i> da lente principal é maior que o <i>f-number</i> referente às microlentes, as imagens geradas são cortadas, havendo desperdício de resolução, ou seja, vários <i>pixels</i> escuros. Contrariamente, se o <i>f-number</i> da lente principal é menor, ocorre sobreposição. A situação ideal ocorre no topo desta figura, onde os <i>f-stops</i> são todos idênticos. Imagem adaptada de [7]. | 10 |
| 2.7 | Explicação do <i>slope</i> . O giro do anel focalizador de uma câmera convencional, indicado pelas setas circulares, é responsável por modificar a distância d original, exibida em (a), entre a lente da câmera (plano u) e o plano focal (x), alterando a profundidade a ser focalizada. Em (b), tem-se um d' maior que d , sendo que o foco está em um plano mais próximo da câmera com relação ao foco original. Em (c), nota-se um d'' menor que d , fazendo com que objetos mais distantes da câmera sejam focalizados. Adaptada de [8]. | 11 |
| 2.8 | <i>Light field</i> completo não processado: Sophie & Vincent . Notam-se grandes blocos de <i>pixels</i> bem definidos, os quais são chamados de <i>macropixels</i> | 12 |
| 2.9 | Ilustração de como uma imagem de subabertura é obtida a partir das diversas angulações possíveis para todas as microlentes. Verifica-se que a posição (u, v) da lente principal permanece inalterada, à medida que (i, j) percorre todo o plano das microlentes. Imagem adaptada de [7]. | 13 |
| 2.10 | Ilustração esquemática de duas diferentes imagens de sub abertura. Cada uma delas se refere a uma posição distinta (u, v) da lente principal (mostrada à esquerda), o que implica uma leve mudança na paralaxe vertical. Adaptada de [7]. | 13 |
| 2.11 | Imagens de subabertura referentes ao LF Railway Lines 1 . Percebe-se uma pequena mudança do ponto de vista, principalmente à direita da barra de ferro, onde o caminhão aparece, devido à paralaxe vertical. | 14 |
| 2.12 | Recorte do LF completo Sophie & Vincent . A visualização dos <i>macropixels</i> permite perceber como são geradas as SAIs: escolhendo-se uma mesma posição dentro de todos os <i>macropixels</i> | 14 |
| 2.13 | Ilustração de uma imagem obtida por uma das microlentes para ilustrar o conceito de <i>confidence</i> . Cores mais claras representam áreas onde a luz foi mais intensamente capturada. Nota-se que as bordas da imagem, principalmente, sofrem com problemas de captura da intensidade luminosa, portanto, tendem a ser menos confiáveis. Pode ser tida como um <i>zoom</i> de uma porção da imagem da Figura 2.14. Adaptada de [9]. | 15 |
| 2.14 | Exemplo de imagem branca de calibração capturada em um ambiente controlado. Em vermelho, nota-se o centro estimado para cada microlente. Imagem de [9]. | 15 |
| 2.15 | Representação bidimensional da relação entre o LF adquirido e uma imagem sintetizada LF'. Percebe-se o mapeamento feito entre uma das coordenadas da abertura original (u) e a abertura sintética (u'). O mesmo vale para os planos focais original (i) e sintético (i'). De modo implícito define-se ξ e φ . Adaptada de [7]. | 16 |

| | | |
|------|---|----|
| 2.16 | Visualização de uma imagem <i>light field</i> sendo interpretada como uma coleção 2D de imagens 2D. Cada uma dessas fatias bidimensionais correspondem a uma posição diferente do observador. Imagem adaptada de [10]. | 18 |
| 2.17 | Imagens sintetizadas com diferentes focos. De cima para baixo, nota-se respectivamente: o foco nos objetos mais ao fundo (janela por exemplo); depois o foco mais ao meio (rosto do Vincent); e finalmente o foco mais à frente (rosto da Sophie). | 19 |
| 2.18 | Esquema de sintetização de uma fotografia E_0 com foco em um plano M_0 localizado bem atrás na cena. (a) Objetos mais ao fundo são focalizados. (b) O plano focal (plano da imagem) coincide com o plano das microlentes (MLA). Adaptada de [11, 12]. | 20 |
| 2.19 | Representação de como focalizar um plano M_1 mais próximo da câmera do que M_0 . (a) Imagem sintetizada E_1 com foco em objetos mais à frente na cena. (b) A imagem se forma atrás da matriz de microlentes. Adaptada de [11, 12]. | 21 |
| 2.20 | Mudança de profundidade entre planos paralelos e paralaxe entre diferentes imagens de subabertura. P é o plano de referência, isto é, a profundidade original e P' refere-se a uma profundidade sintética mais distante da câmera do que P . Os pontos de vista V_i e V_0 representam diferentes imagens de subabertura, com diferentes paralaxes. Essa paralaxe depende tanto do deslocamento ψ quanto da profundidade relativa $\frac{\Delta_z}{\Delta_z + Z_0}$. Adaptada de [13]. | 22 |
| 2.21 | Ilustração dos três diferentes tipos de <i>frames</i> referentes à codificação preditiva. Nas bordas da imagem, estão os <i>frames-I</i> , os quais são codificados independentemente dos outros <i>frames</i> do vídeo ao qual pertencem. Mais ao meio e hachurados estão os <i>frames-P</i> e os <i>frames-B</i> preditos. Imagens adaptada de [14]. | 24 |
| 3.1 | Ilustração do processo de decriptação (<i>demosaiicing</i>) de uma imagem <i>light field</i> no formato <i>raw</i> , capturada por uma câmera Lytro [15], e renderização de uma vista bidimensional. Imagem de [16]. | 27 |
| 4.1 | Ilustração do funcionamento do trabalho desenvolvido por Peixoto et al. [17]. A mesma ideia geral é válida para a proposta deste trabalho. A ser publicada em [18]. | 30 |
| 4.2 | Fluxogramas que estabelecem uma visão geral do PLFC proposto por este trabalho. (a) Processo de inicialização da <i>cache</i> . (b) Processo de sintetização de um novo ponto focal, sendo necessária ou não a transmissão de uma ou mais imagens adicionais. | 34 |
| 5.1 | À esquerda: LF completo. À direita: Uma das SAIs obtidas. | 40 |
| 5.2 | Lado esquerdo: LF completo, demonstrando os <i>macropixels</i> capturados. Lado direito: Ilustração de uma das 225 imagens de subabertura obtidas. | 41 |
| 5.3 | Lado esquerdo: <i>raw</i> LF com resolução completa de 9615×6750 , sendo possível verificar os <i>macropixels</i> capturados. Lado Direito: Ilustração de uma das 225 SAIs obtidas. | 42 |

| | | |
|------|--|----|
| 5.4 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Swans 2</i> , com PSNR alvo de 36 dB. Apresenta-se os resultados das transmissões triviais ILFS, do PLFC original [17] e do PLFC proposto por este trabalho. | 45 |
| 5.5 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Swans 2</i> , PSNR alvo de 38 dB. Compara-se os resultados do ILFS, do PLFC [17] e do novo PLFC. | 45 |
| 5.6 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Swans 2</i> , PSNR alvo de 40 dB. Nota-se os resultados do ILFS, do PLFC [17] e do PLFC proposto. | 46 |
| 5.7 | Resultado da transmissão de 100 <i>slopes</i> para PLFC [17], ILFS e PLFC proposto para o LF <i>Sophie & Vincent 1</i> , alvo 36 dB. A taxa acumulada, em <i>bits</i> , encontra-se no eixo vertical e o número de cada imagem requisitada (de 1 a 100) está no eixo horizontal. | 46 |
| 5.8 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Sophie & Vincent 1</i> , com PSNR alvo de 38 dB. A taxa acumulada, em <i>bits</i> , encontra-se no eixo vertical e o número de cada imagem requisitada (de 1 a 100) está no eixo horizontal. Os resultados do ILFS, do PLFC original e do PLFC proposto são apresentados. | 47 |
| 5.9 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Sophie & Vincent 1</i> , PSNR alvo de 40 dB. Eixo vertical: taxa acumulada, em <i>bits</i> . Compara-se os resultados do ILFS, do PLFC [17] e do novo PLFC. | 47 |
| 5.10 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Sophie Krios & Vincent</i> , com PSNR alvo de 36 dB. Comparação entre PLFC original, PLFC proposto e ILFS. | 48 |
| 5.11 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Sophie Krios & Vincent</i> , com PSNR alvo de 38 dB. Exibe-se a taxa acumulada conforme o número de imagens requisitadas pelo usuário. | 48 |
| 5.12 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Sophie Krios & Vincent</i> , com PSNR alvo de 40 dB. A taxa acumulada, em <i>bits</i> está representada no eixo das ordenadas e o número do <i>slope</i> pedido (de 1 a 100) está no eixo das abcissas. Resultados do ILFS, do PLFC original e do PLFC proposto são apresentados. | 49 |
| 5.13 | Resultado taxa distorção referente ao LF <i>Swans 2</i> depois de 100 imagens transmitidas. Quanto mais acima e à esquerda uma curva estiver, melhor é o resultado. Exibe-se os resultados do ILFS, do PLFC e do PLFC proposto. | 50 |
| 5.14 | Resultado taxa distorção referente ao LF <i>Sophie & Vincent 1</i> | 50 |
| 5.15 | Resultado taxa distorção referente ao LF <i>Sophie Krios & Vincent</i> | 51 |
| 5.16 | Caso particular do PLFC anterior [17]. Nota-se que com 100 imagens enviadas a qualidade mínima não foi atingida. Nessa situação o sistema chegava a mandar todas as 225 SAIs disponíveis no servidor e ainda assim a PSNR mínima não era alcançada. Por outro lado, o PLFC proposto envia menos taxa com relação ao momento retratado de [17] e muito menos imagens, suficientes para todos os 100 <i>slopes</i> requisitados. | 52 |

| | | |
|------|---|----|
| 5.17 | Ilustração do comportamento do método proposto em comparação com PLFC original [17]. Nota-se que conforme a qualidade requisitada aumenta, o PLFC original precisa mandar uma taxa significativamente maior em comparação com a qualidade imediatamente anterior. Por outro lado, o método proposto apresenta um salto menor em termos de taxa à medida que a qualidade aumenta. | 53 |
| 5.18 | Análise do efeito do benefício futuro, o qual é controlado pelo parâmetro μ da Equação (4.7), na transmissão do <i>light field Friends 3</i> . No caso que considera o benefício futuro, a taxa acumulada final foi menor e, mesmo assim, a qualidade atingida (PSNR) foi maior. | 54 |
| 5.19 | Análise do efeito do benefício futuro, controlado pelo μ da Equação (4.7), na transmissão do LF <i>Sophie Krios & Vincent</i> . Nota-se a diminuição da taxa acumulada e o aumento da qualidade atingida (PSNR) quando se considera também o benefício futuro. | 55 |
| A.1 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Friends 1</i> , com PSNR alvo de 36 dB. A taxa acumulada, em <i>bits</i> , encontra-se no eixo vertical e o número de cada imagem requisitada (de 1 a 100) está no eixo horizontal. São comparados os resultados obtidos pelo ILFS de referência, os resultados do PLFC anterior bem como os resultados do PLFC proposto. | 63 |
| A.2 | Resultado da transmissão de pontos focais sintetizados para o <i>light field Friends 1</i> , com PSNR alvo de 38 dB. | 64 |
| A.3 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Friends 1</i> , PSNR alvo de 40 dB. | 64 |
| A.4 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Friends 3</i> , PSNR alvo de 36 dB. | 65 |
| A.5 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Friends 3</i> , PSNR alvo de 38 dB. | 65 |
| A.6 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Friends 3</i> , PSNR alvo de 40 dB. | 66 |
| A.7 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Swans 1</i> , PSNR alvo de 36 dB. | 66 |
| A.8 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Swans 1</i> , PSNR alvo de 38 dB. | 67 |
| A.9 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Swans1</i> , PSNR alvo de 40 dB. | 67 |
| A.10 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Chain-link Fence 1</i> , com PSNR alvo de 36 dB. A taxa acumulada desde o início da execução do sistema, em <i>bits</i> , encontra-se no eixo vertical e o número de cada imagem requisitada (de 1 a 100) está no eixo horizontal. Compara-se os resultados PLFC proposto com os resultados do método referência ILFS. | 68 |
| A.11 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Chain-link Fence 1</i> , com PSNR alvo de 38 dB. Compara-se o resultado desta proposta com o resultado do ILFS. | 68 |

| | | |
|------|--|----|
| A.12 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Chain-link Fence 1</i> , com PSNR alvo de 40 dB. É feita a comparação entre este método e o ILFS. | 69 |
| A.13 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Billboards</i> , com PSNR alvo de 36 dB. Exibe-se o resultado do ILFS, do PLFC e do PLFC proposto. | 69 |
| A.14 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Billboards</i> , com PSNR alvo de 38 dB. | 70 |
| A.15 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Billboards</i> , PSNR alvo de 40 dB. | 70 |
| A.16 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Poopies</i> , PSNR alvo de 36 dB. | 71 |
| A.17 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Poopies</i> , PSNR alvo de 38 dB. | 71 |
| A.18 | Resultado da transmissão de pontos focais sintetizados referentes ao <i>light field Poopies</i> , PSNR alvo de 40 dB. | 72 |
| A.19 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Railway Lines 1</i> , PSNR alvo de 36 dB. | 72 |
| A.20 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Railway Lines 1</i> , PSNR alvo de 38 dB. | 73 |
| A.21 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Railway Lines 1</i> , PSNR alvo de 40 dB. | 73 |
| A.22 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Books</i> , PSNR alvo de 36 dB. | 74 |
| A.23 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Books</i> , PSNR alvo de 38 dB. | 74 |
| A.24 | Resultado da transmissão de pontos focais sintetizados referentes ao LF <i>Books</i> , PSNR alvo de 40 dB. | 75 |
| A.25 | Resultado taxa distorção referente ao LF <i>Friends 1</i> após 100 imagens transmitidas. Quanto mais acima e à esquerda uma curva estiver, melhor é o resultado. Exibe-se os resultados do ILFS, do PLFC e do PLFC proposto. | 75 |
| A.26 | Resultado taxa distorção referente ao LF <i>Friends 3</i> | 76 |
| A.27 | Resultado taxa distorção referente ao LF <i>Swans 1</i> | 76 |
| A.28 | Resultado taxa distorção referente ao LF <i>Chain-link Fence 1</i> | 77 |
| A.29 | Resultado taxa distorção referente ao LF <i>Billboards</i> | 77 |
| A.30 | Resultado taxa distorção referente ao LF <i>Poopies</i> | 78 |
| A.31 | Resultado taxa distorção referente ao LF <i>Railway Lines 1</i> | 78 |
| A.32 | Resultado taxa distorção referente ao LF <i>Books</i> | 79 |

Lista de Tabelas

- 4.1 Exemplo do mapeamento feito entre a qualidade alvo L_{QLD} e o parâmetro x para obter o valor de ϵ . Os dados da primeira coluna estão quantificados em decibéis em termos da relação sinal ruído, ou seja, quanto maior o valor, maior a qualidade. Os colchetes indicam que o intervalo é fechado e os parênteses indicam que o intervalo é aberto. Conforme a qualidade desejada aumenta, o fator de tolerância (ϵ) muda mais cedo. As constantes usadas nesse exemplo foram $a_1 = 5.81$ e $a_2 = 32$ 36
- 5.1 BD-Taxa [19] depois de um determinado número de imagens transmitidas, usando ILFS [20–24] como base. O resultado do PLFC anterior [17] aparece nas subcolunas da esquerda, ao passo que o resultado desta proposta encontra-se nas subcolunas da direita (prop). Valores positivos referem-se ao quão pior o método em questão é em comparação com o método base, sendo que valores negativos representam os ganhos obtidos em porcentagem. 52

Lista de Siglas, Abreviações e Acrônimos

| | |
|------------|--|
| 2D | Do inglês 2 <i>dimensional</i> , indica duas dimensões, bidimensional |
| 4D | Do inglês 4 <i>dimensional</i> , quatro dimensões, quadridimensional |
| AVC | Do inglês <i>Advanced Video Coding</i> em referência à proposta para estabelecer um padrão de codificação avançada de vídeo |
| BD | Métrica de avaliação definida por Bjøntegaard; computa o ganho médio em termos de qualidade (BD-PSNR, por exemplo) ou em termos de taxa (BD-Taxa) entre duas curvas de taxa-distorção |
| DCT | <i>Discrete Cosine Transform</i> , Transformada Discreta de Cosseno |
| H.264 | Padrão de codificação de vídeo desenvolvido pela União Internacional de Telecomunicações (ITU) e pela Organização Internacional de Padronização (ISO); também chamado de AVC |
| HEVC | <i>High Efficiency Video Coding</i> , em alusão ao padrão de codificação de vídeo de alta eficiência; também conhecido como H.265 |
| ILFS | <i>Interactive Light Field Streaming</i> , Transmissão Interativa de LF |
| JPEG | <i>Joint Photographic Experts Group</i> , padrão de compressão de imagens com perdas |
| JPEG PLENO | Variação do JPEG, haja vista que PLENO corresponde a <i>plenoptic</i> , em alusão às câmeras plenópticas; padrão ainda em desenvolvimento de compressão de imagens <i>light field</i> |
| LF | Imagens <i>Light Field</i> , traduzido para o português como Imagens de campo de luz, basicamente se referem a imagens que possuem informações adicionais a respeito da profundidade e do foco |
| MLA | <i>MicroLens Array</i> ; Matriz de Microlentes, Plano das Microlentes |
| PLFC | <i>Progressive Light Field Communication</i> , traduzida como Comunicação Progressiva de Light Field |
| PSNR | <i>Peak-Signal-to-Noise-Ratio</i> , Razão entre pico de sinal e ruído |
| QP | <i>Quantization Parameter</i> , Passo (Parâmetro) de Quantização |
| RGB | <i>Red Green Blue</i> , isto é, vermelho, verde e azul respectivamente |

Capítulo 1

Introdução

Câmeras *light field*, também chamadas câmeras plenópticas, são câmeras que armazenam a distribuição da luz captada em uma cena. Diferentemente de uma fotografia bidimensional (2D) convencional, a qual mede apenas a quantidade de luz que chega em cada ponto do sensor, o princípio das câmeras *light field* é essencialmente a captura de uma imagem quadridimensional (4D) [7], obtida a partir da função plenóptica completa de 7 dimensões [25]. Assim, identifica-se a quantidade de luz contida em cada raio luminoso que intercepta o sensor da câmera plenóptica, considerando diversos ângulos de incidência. Caso a informação de cor seja considerada, o dado obtido passa a ter 5 dimensões.

Resumidamente, pode-se dizer que a ideia da captura dessas duas dimensões extras com relação à câmera convencional é a possibilidade de se computar fotografias sintetizadas a partir da luz adquirida. Portanto, uma imagem convencional pode ser entendida como uma subimagem, a qual Levoy e Hanrahan [26] denominam *slice*, de uma imagem 4D. O processo de criação de uma imagem *light field*, a partir de um conjunto de imagens, implica a inserção de cada *slice* 2D na representação quadridimensional e, analogamente, criar novos pontos de vista envolve extrair e reamostrar uma dessas subimagens [26].

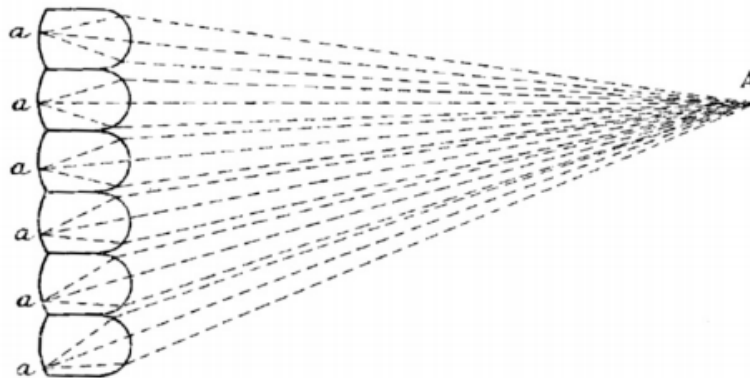


Figura 1.1: Representação simplificada em duas dimensões de como a imagem de um ponto A é capturada pelo sensor de Lippmann, por meio de uma matriz de pontos. Percebe-se que cada microlente captura uma imagem a na mesma posição relativa. Imagem de [1].

Convém destacar que em 1908, Lippmann estabeleceu um conceito conhecido como fotografia integral [1], complementado por Ives [27] um pouco depois. Tal conceito é a base do funcionamento das câmeras plenópticas [7].

Dito isso, é importante tratar dos sensores de Lippmann, ilustrados na Figura 1.1. Gorgiev et al. [28] resumem tal sensor como uma matriz de lentes que formam microcâmeras focadas em um plano determinado. Afirmam que, apesar dessas microlentes serem um separador de luz ideal, elas são usadas de fato para captar luz suficiente e evitar efeitos de difração.

Embora a primeira câmera *light field* tenha sido proposta por Lippmann [1] há mais de um século (1908), este tipo de câmera tem ganhado mais atenção nos dias de hoje. Somente em 1992, Adelson e Wang [29] propuseram um maneira de reduzir consideravelmente o problema de correspondência pertinente a essas câmeras, isto é, onde quais partes de uma imagem podem ser encontradas em outra imagem. Tal avanço foi possível pela colocação de uma matriz de microlentes no plano focal da lente principal da câmera, permitindo a análise de profundidade de uma cena.

Outra característica importante das câmeras *light field* é que, por conta da sua capacidade de medir tanto a cor quanto informações geométricas das cenas, tais câmeras podem operar em condições nas quais as câmeras convencionais não conseguiriam, como contra o brilho do sol ou debaixo da água. Assim, é pertinente apontar algumas aplicações referentes aos dados *light-field*:

- Renderização baseada em imagem: é possível reconstruir os raios de luz que atravessam qualquer ponto de vista, reamostrando os raios já adquiridos [26].
- Elaboração de fotografias de abertura sintética, pois as imagens *light field* podem ser usadas para simular o *defocus blur* de lentes convencionais, que se referem à mudança focal [30].
- Estimação da geometria de cenas, nas quais as imagens de subabertura podem ser utilizadas para estimar o aspecto tridimensional de um objeto [31].
- Possibilidade de se aplicar algumas outras operações na imagem depois de sua captura, interessantes para a visão computacional [32], como a remoção de oclusões.

Diante disso, é importante notar que existem diversas implementações de câmeras plenópticas [7, 33–35]. No entanto, dois tipos básicos de implementação são os principais [6]: (i) aquelas formadas por uma matriz de câmeras (ii) e aquelas que possuem microlentes em seu interior, do inglês *lenslet-based cameras*, como por exemplo Lytro [15] e Raytrix [36], as quais usam tais microlentes para capturar raios de luz individuais. A quantidade de dados capturados por câmeras do tipo (i) é ainda maior em comparação com as câmeras do tipo (ii).

Vale ressaltar que as imagens obtidas por essas câmeras ainda carecem de padronização, no que se refere à compressão e à transmissão. A despeito disso, pode-se destacar que vem sendo desenvolvida uma padronização, referida como JPEG PLENO (sendo JPEG a abreviação do inglês *Joint Photographic Experts Group* e PLENO referente a *Plenoptic*), a qual ainda está em desenvolvimento [37].

1.1 Descrição do problema

As câmeras convencionais apresentam uma limitação inerente ao foco, sendo necessário um alto nível de concentração no momento da captura para que se consiga um foco

adequado. Mesmo com a assistência fornecida pelos sistemas de autofoco, um pequeno erro já é capaz de comprometer toda a foto. Em outras palavras, uma fotografia mal focada pode ser irreversível, pois não se pode mudar o foco satisfatoriamente depois da captura [2].

Diante disso, as câmeras *light field* ganham evidência, pois elas conseguem armazenar a informação angular dos raios de luz, a qual é perdida no momento da aquisição da imagem nas câmeras convencionais [12]. Os raios luminosos originários de um ponto fora de foco se espalham por vários pontos do sensor, o que se transforma em uma região visivelmente borrada, como mostra a Figura 1.2 (a). A Figura 1.2 (b) ilustra uma imagem com profundidade expandida sintetizada a partir de outras imagens com diferentes focos [3], na qual cada pessoa da cena aparece focalizada ao mesmo tempo.



(a) Imagem desfocada

(b) Imagem nítida

Figura 1.2: Ilustração de uma imagem sintetizada borrada, fora de foco. Imagem nítida sintetizada com profundidade expandida. Adaptada de [2].



(a) Foco mais ao fundo

(b) Foco mais à frente

(c) Profundidade expandida

Figura 1.3: Imagens sintetizadas referentes a diferentes focos. Em (c) a imagem sintetizada com profundidade expandida é obtida pela combinação entre os dois focos anteriores [3], focalizando os dois planos ao mesmo tempo. Imagens originais de [2].

Assim, os dados *light field* podem ser utilizados para sintetizar focos depois do instante da captura, como mostram a Figura 1.2 e a Figura 1.3. Entretanto, essa e outras capacidades após a captura envolvem um custo maior de armazenamento, pois são adquiridos muito mais dados em comparação com uma fotografia convencional. Deve-se então desenvolver um método eficaz e eficiente para transmitir esses dados, uma vez que as imagens *light field* apresentam uma redundância característica.

1.2 Objetivo geral

Desenvolver um método de transmissão de imagens sintetizadas de um mesmo *light field* de forma a reduzir a taxa acumulada final, no qual um usuário possa visualizar diferentes pontos focais da imagem.

1.3 Objetivos específicos

- Estudar a sintetização de imagens a partir de imagens *light field*;
- Estudar o estado da arte referente à compressão e à transmissão de imagens *light field*;
- Estudar conceitos relacionados à otimização e modelos probabilísticos, para estabelecer quantas imagens o cliente (usuário) deve ter no início de uma transmissão;
- Analisar o impacto futuro de se transmitir determinada imagem, no sentido de que ela possa ser reutilizada para as próximas sintetizações de vista;
- Selecionar ferramentas apropriadas para alcançar satisfatoriamente o objetivo geral;
- Implementar um método de otimização em termos de taxa-distorção em que um servidor transmita imagens sintetizadas de *light field* para um cliente (usuário);
- Expandir a proposta, estipulando a quantidade de imagens que devem ser transmitidas inicialmente (*cache* inicial);
- Aprimorar o critério de seleção das imagens com base no benefício futuro;
- Fazer análise e comparação dos resultados obtidos com o estado da arte referente ao tema.

1.4 Organização

Este texto está dividido em 6 capítulos, sendo o presente capítulo o introdutório. No Capítulo 2, os conceitos básicos relacionados a câmeras e a imagens *light field* são apresentados. O Capítulo 3 descreve os trabalhos prévios que dizem respeito ao tema desta proposta, abordando a compressão e a transmissão de imagens *light field*. Outrossim, esse capítulo cita algumas aplicações de tais imagens, embora isso não seja o foco deste trabalho. O Capítulo 4 detalha a metodologia desta proposta, além de descrever o trabalho base utilizado como referência. No Capítulo 5, constam as imagens e as configurações

utilizadas para os testes realizados. Ademais, exibe-se e discute-se os resultados obtidos. O sexto e último capítulo aborda as contribuições e as conclusões deste trabalho, sugerindo também alguns trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Neste capítulo são explicados os conceitos necessários para o entendimento da proposta deste trabalho. Inicialmente é importante tratar das câmeras convencionais. Logo em seguida, o mecanismo das câmeras *light field* deve ser abordado, mais precisamente aquelas dotadas de microlentes, como a Lytro [15] utilizada neste trabalho.

Exposto o funcionamento dessas câmeras, finalmente é preciso descrever o dado 4D *light field* capturado e como podem ser feitas algumas operações a partir dele. Por fim, merecem referência técnicas básicas de predição para a compressão de imagens e vídeos, as quais podem ser aplicadas a imagens *light field*, se feitas as devidas adaptações.

2.1 Câmeras convencionais

A câmera fotográfica convencional é um dispositivo dotado de um sensor fotossensível posicionado à distância b_M atrás do plano da lente principal, de modo que todos os raios capturados a partir de um objeto a uma distância a_M converjam em um único ponto do plano sensor (ou plano focal, do inglês *film plane*), como exhibe a Figura 2.1.

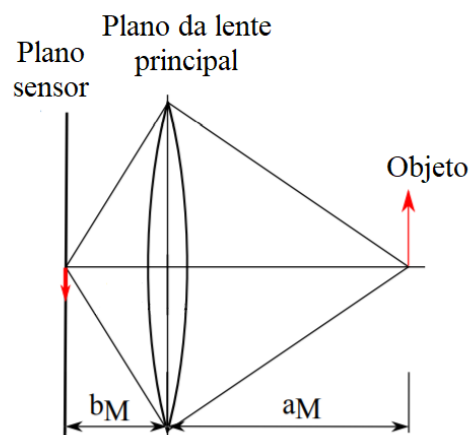


Figura 2.1: Funcionamento de uma câmera convencional. Os raios luminosos capturados a partir de um ponto de um determinado objeto são mapeados para um único ponto da imagem. Se esse objeto estiver em foco, os raios são focalizados em um mesmo ponto do plano sensor. Adaptada de [4].

Nesse sentido, a câmera convencional integra todos os raios de luz que chegam em um determinado ponto no sensor, vindos de todas as direções, formando uma projeção do mundo tridimensional para um plano bidimensional. Esse fato engendra uma perda de informação com relação à profundidade da cena [4].

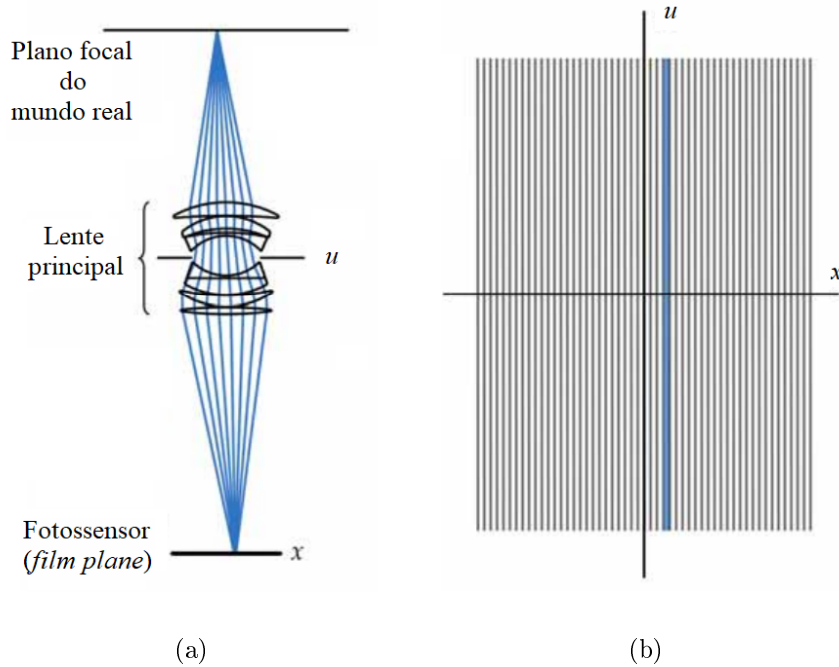


Figura 2.2: Formação de um *pixel* em uma fotografia. O cone de luz representado em (a) intercepta o plano focal x em um mesmo ponto e é mapeado para o diagrama de raios como a tira vertical em (b). Essa tira corresponde a soma de todas as posições de u , gerando um *pixel* da imagem. Adaptada de [2].

Assim, cada posição no plano sensor soma todos os raios luminosos que a interceptam [2], sendo que o peso de cada raio nessa soma depende da sua direção de incidência com relação ao plano sensor. No entanto, esses efeitos direcionais são indesejáveis e ocorrem por conta de limitações físicas ou por conta da implementação feita. A Figura 2.2 ignora tais efeitos e mostra uma fotografia, de certo modo, idealizada.

Assumindo que u represente o plano da lente da câmera e que a imagem se forme no plano focal x , cada tira azul vertical na Figura 2.2 (b) corresponde à soma de um conjunto de raios luminosos para formar um *pixel* (elemento de imagem, do inglês *picture element*) diferente da imagem capturada. Nesse sentido, o cone de luz (tridimensional) ilustrado na Figura 2.2 (a) pode ser projetado como a tira destacada na Figura 2.2 (b) (bidimensional), uma vez que os raios desse cone interceptam o plano sensor na mesma posição, mas assumem todas as direções possíveis na lente principal (u).

Diferentes *pixels* de uma fotografia interceptam o plano sensor em posições distintas, correspondendo a diferentes linhas verticais do diagrama bidimensional de raios. Sendo assim, para formar a fotografia completa, é preciso considerar todas as projeções verticais da Figura 2.2 (b), valendo a pena destacar que tais projeções preservam a localização espacial x dos raios, mas fazem com que a informação direcional u se perca.

Ainda sobre as câmeras convencionais, convém mencionar que é possível modificar a profundidade a ser focalizada. Para isso, é preciso alterar a distância entre o plano sensor e a lente da câmera [2]. Se a posição da câmera for mantida, pode-se mover o plano focal, modificando o local em que a imagem se forma (plano da imagem, do inglês *image plane*).

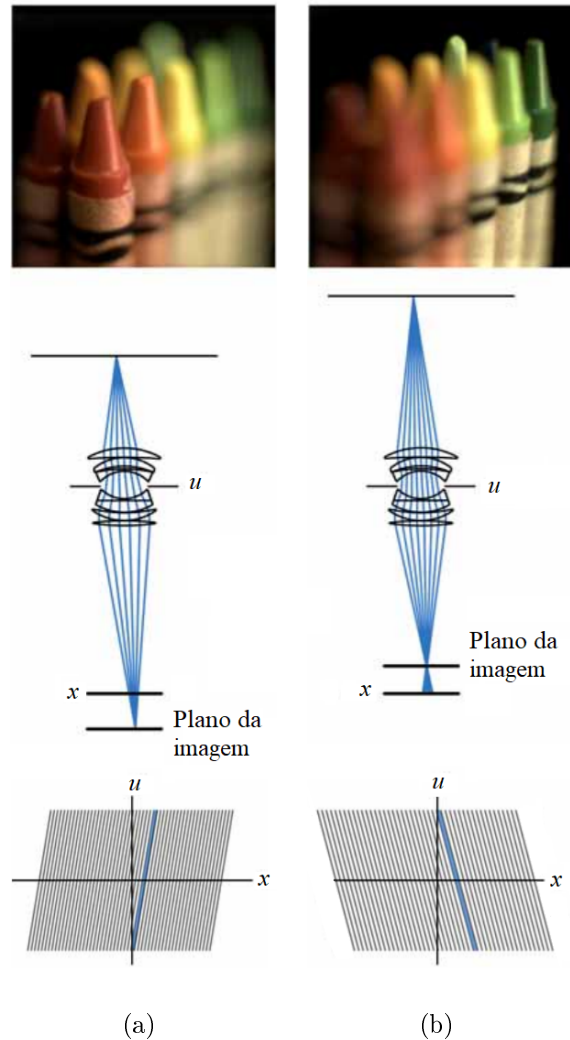


Figura 2.3: Mudança de foco em uma câmera convencional por meio da mudança de posição do plano focal. (a) Objetos mais próximos da câmera são focalizados, sendo que a imagem se forma depois do plano focal original x . No diagrama mais abaixo, o movimento dos raios através de u possui a mesma direção do movimento através do plano x . (b) A imagem se forma antes de x , o foco está em objetos mais distantes da câmera. No diagrama de raios, a direção do movimento das interseções no plano u é oposta com relação a direção dos raios em x . Adaptada de [2].

Na Figura 2.3 (a), focalizar objetos mais à frente em uma cena implica que o plano da imagem será localizado depois do plano focal original (x), sendo que o fundo da cena aparece borrado. Em contrapartida, a imagem se forma antes de x na Figura 2.3 (b), pois o foco está em objetos mais distantes com relação à câmera. Nesse caso, a porção borrada da imagem está localizada mais à frente da cena.

Quanto aos diagramas de raios presentes na parte debaixo da Figura 2.3, é possível perceber que à medida que as interceptações dos raios se movem linearmente pelo plano u da lente, se movem também de maneira linear através do plano focal original x [2]. Na Figura 2.3 (a), como a convergência se dá depois de x , o movimento através de u é na mesma direção que o movimento no plano x . Na Figura 2.3 (b) essas direções são opostas, pois a convergência ocorre antes de x . Portanto o ângulo de deslocamento depende da separação entre o plano da imagem e o plano focal original x .

2.2 Câmeras *Light Field*

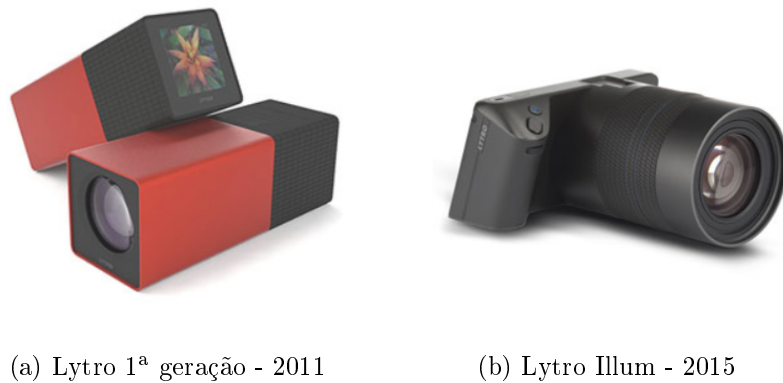


Figura 2.4: Ilustração de câmeras *light field* Lytro. A câmera mostrada em (a) foi a primeira câmera *light field* voltada para o consumidor. As imagens utilizadas nos testes deste trabalho foram capturadas por uma câmera Lytro Illum, como a exibida em (b). Imagens originais de [5].

Antes de expor as imagens propriamente ditas, convém explicar o funcionamento das câmeras *light field*. Basicamente, uma *lenslet-based camera* é formada por uma lente principal (similar às lentes das câmeras convencionais), por uma matriz de microlentes bem como por uma matriz de fotossensores [7], como mostra a Figura 2.5.

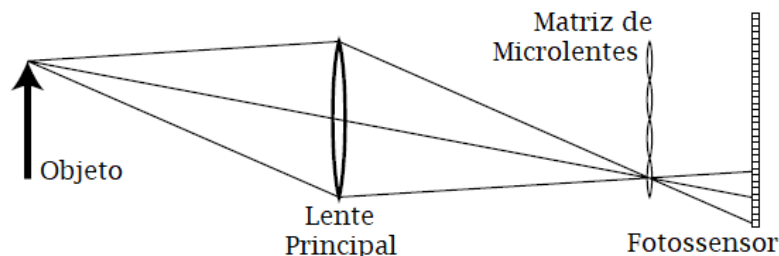


Figura 2.5: Esquema bidimensional representativo de uma câmera *light field*. Percebe-se que os raios luminosos são capturados por sensores situados logo atrás da matriz de microlentes. Raios de diferentes direções atingem o plano sensor em diferentes *pixels*. Adaptada de [6].

Verifica-se que, de acordo com a direção de incidência, os raios luminosos são separados quando chegam às microlentes, gerando uma imagem focalizada da abertura da lente principal na matriz de *pixels* localizada abaixo das microlentes. Essa separação do feixe de raios luminosos incidentes promovida pela matriz de microlentes faz com que raios de diferentes direções interceptem o sensor em diferentes posições. Isso preserva a informação angular pertinente a esses raios, diferentemente do que ocorre em uma câmera convencional [4], cuja imagem possui apenas resolução angular.

Assim, como uma câmera *light field* é um sistema de dois planos de lentes consecutivos, a lente principal é considerada como o ‘objeto’ a ser capturado pelas microlentes [4]. Essa lente principal deve então ser estruturada de modo a considerar tanto a clareza das imagens das microlentes, quanto o seu tamanho, a fim de cobrir a maior quantidade de *pixels* fotossensores possível.

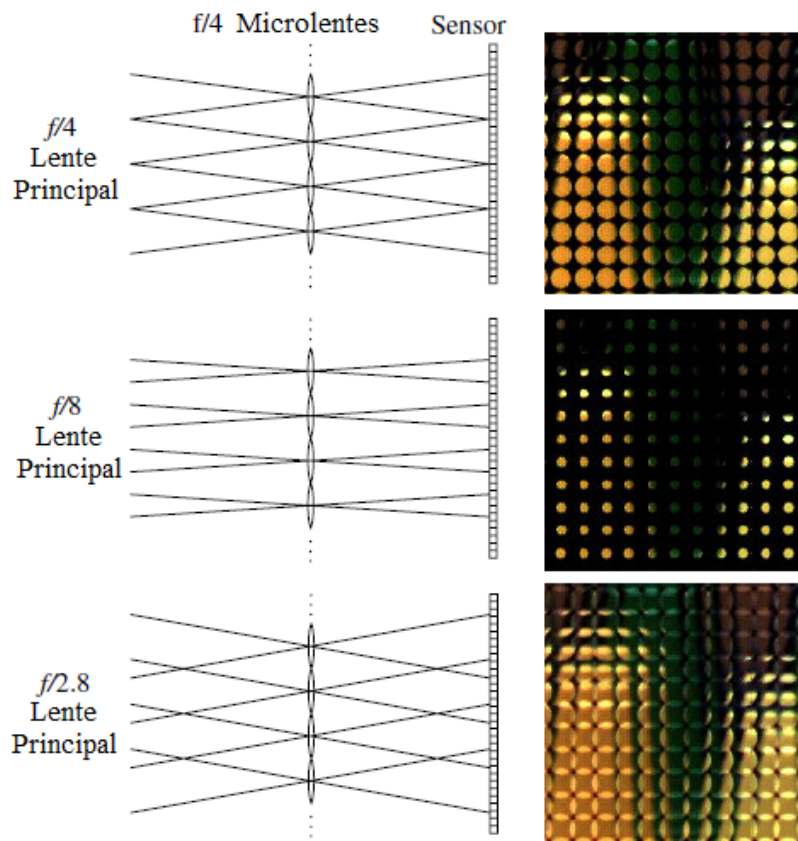


Figura 2.6: Esquema ilustrativo dos diferentes *f-numbers* para a lente principal e para as microlentes. Quando o *f-number* da lente principal é maior que o *f-number* referente às microlentes, as imagens geradas são cortadas, havendo desperdício de resolução, ou seja, vários *pixels* escuros. Contrariamente, se o *f-number* da lente principal é menor, ocorre sobreposição. A situação ideal ocorre no topo desta figura, onde os *f-stops* são todos idênticos. Imagem adaptada de [7].

O ideal é escolher o tamanho relativo entre a lente principal e as aberturas das microlentes que proporciona o maior tamanho possível para cada imagem, sem que haja sobreposição. Isso ocorre quando os *f-numbers* das microlentes são iguais ao *f-number* da lente principal, como ilustra a Figura 2.6.

Sucintamente, o *f-number* (também chamado *f-stop* ou razão focal, do inglês *focal ratio*) é um valor adimensional que se refere à razão entre a distância focal da lente e o diâmetro da pupila de entrada [38]. Esse número aparece geralmente vinculado à notação f/N , na qual N é de fato o *f-number*, valendo a pena destacar que, quanto maior o *f-number*, menor o tamanho da abertura da lente.

2.2.1 *Slope*

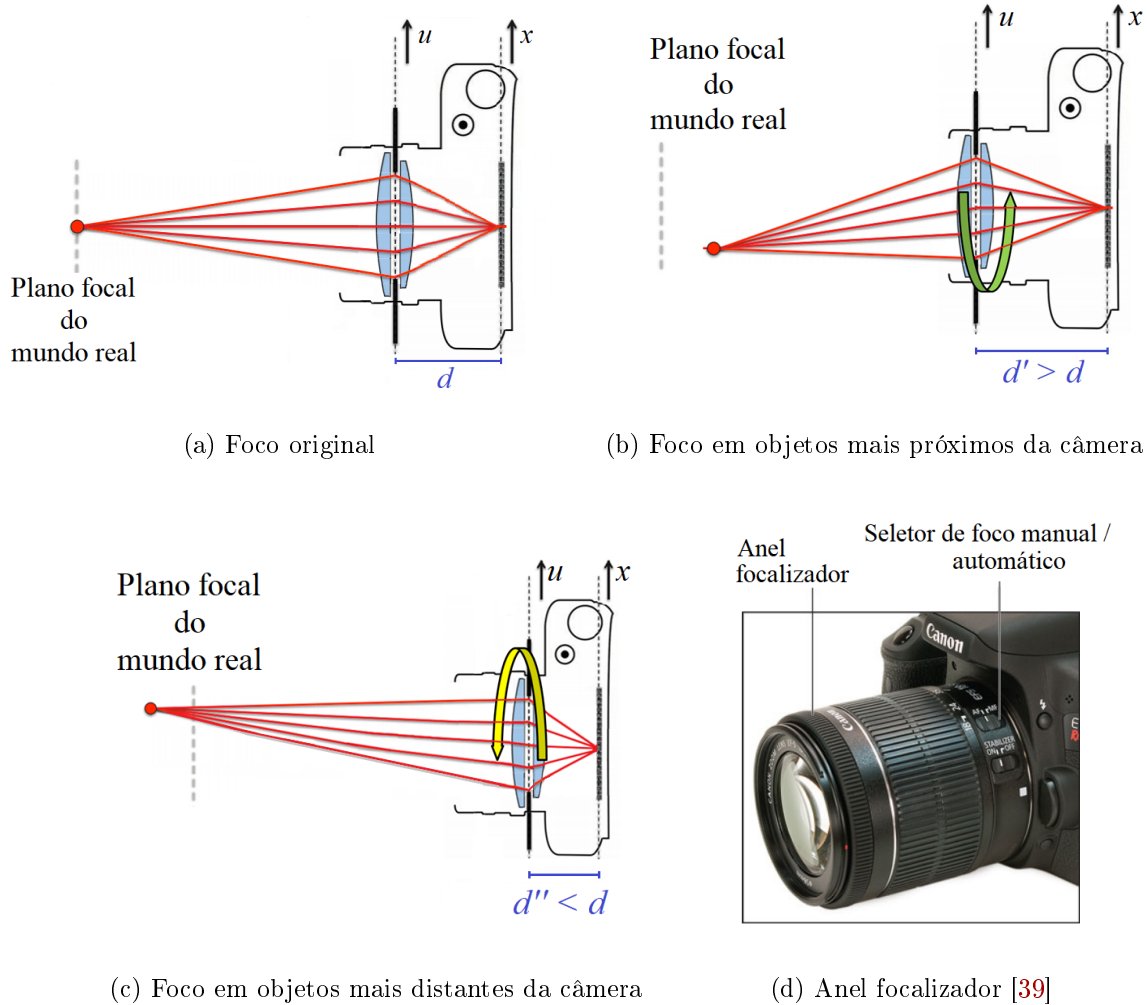


Figura 2.7: Explicação do *slope*. O giro do anel focalizador de uma câmera convencional, indicado pelas setas circulares, é responsável por modificar a distância d original, exibida em (a), entre a lente da câmera (plano u) e o plano focal (x), alterando a profundidade a ser focalizada. Em (b), tem-se um d' maior que d , sendo que o foco está em um plano mais próximo da câmera com relação ao foco original. Em (c), nota-se um d'' menor que d , fazendo com que objetos mais distantes da câmera sejam focalizados. Adaptada de [8].

Em uma câmera convencional, mudar o *slope* geralmente significa girar o anel focalizador [2], Figura 2.7(d), deslocando a lente da câmera de modo a aproximá-la ou a distanciá-la do plano focal. Isso faz com que uma profundidade diferente seja focalizada,

como ilustra a Figura 2.7. Por analogia, neste trabalho, o termo *slope* se refere ao ângulo de deslocamento aplicado aos dados *light field* para produzir imagens de um foco particular, mesmo que o ponto focal correspondente não tenha sido originalmente capturado.

Assim, em comparação ao exemplo dado na Seção 2.1 (Figura 2.3), a Figura 2.7 ilustra uma maneira diferente de se modificar a distância entre o plano focal e a lente da câmera. Nessa figura, ao invés de se manter fixa a posição da câmera, movendo-se o plano focal como feito antes; deixa-se o plano focal (x) imóvel, havendo o deslocamento do conjunto de lentes da câmera (u). De fato, movimentar o plano da lente principal é o modo mais comum de modificar o foco em uma câmera convencional [40].

A partir disso, girar o anel focalizador de modo a distanciar a câmera do plano focal resulta em focalizar objetos mais à frente de uma cena, como mostra a Figura 2.7(b). Se o giro desse anel promover uma aproximação entre a câmera e o plano focal, tem-se um foco em objetos mais ao fundo de uma cena, Figura 2.7(c).

Em grande parte das lentes fotográficas convencionais, o processo de mudança de foco (refoco óptico, do inglês *optical refocusing*) para um plano mais próximo da câmera acaba aumentando a distância focal efetiva (d) e pode causar uma diminuição do campo de visão da cena. Isso porque há um aumento na magnitude da imagem, o que significa um fator de ampliação maior e gera efeitos de “*zoom in*”. Por outro lado, uma mudança de foco a partir de imagens *light field* (refoco digital, do inglês *digital refocusing*) permite que o campo de visão permaneça constante, independentemente da profundidade desejada, ou seja, somente o foco se modifica [2].

2.3 Imagens *Light Field*



Figura 2.8: *Light field* completo não processado: Sophie & Vincent. Notam-se grandes blocos de *pixels* bem definidos, os quais são chamados de *macropixels*.

Seja LF o dado *light field* de 4 dimensões obtido a partir de uma câmera *lenslet-based*, como descrito anteriormente na Seção 2.2. Então, cada amostra $LF(u, v, i, j)$ pode ser interpretada como o raio de luz que intercepta a lente principal (também chamada de plano de abertura) na posição (u, v) e o plano das microlentes (MLA, do inglês *MicroLens Array*) na posição (i, j) , conhecido como plano focal de uma câmera LF.

Diante disso, mantendo fixa a posição (u, v) e variando o ângulo de incidência de modo a obter todas as posições (i, j) , obtém-se uma imagem chamada de *Sub-Aperture Image* (SAI) [7], que pode ser traduzida como imagem de subabertura. O conjunto de todas essas SAIs compõe a imagem *light field* completa, também chamada de *full aperture image* ou de *raw light field* (dados brutos *light field*) [2], como mostra a Figura 2.8.

Assim, modificar (u, v) na lente principal, significa selecionar um *pixel* diferente dentro de todas as microlentes, o que corresponde a escolher uma outra imagem de subabertura [7]. A Figura 2.9 e a Figura 2.10 ilustram como tais imagens são capturadas e a Figura 2.11 exemplifica algumas imagens de subabertura, entendidas como diferentes pontos de vista.

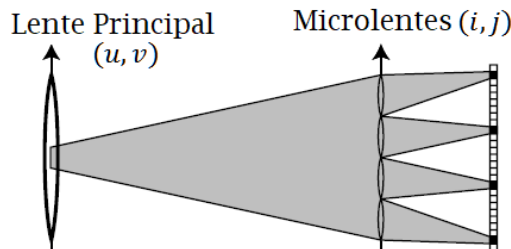


Figura 2.9: Ilustração de como uma imagem de subabertura é obtida a partir das diversas angulações possíveis para todas as microlentes. Verifica-se que a posição (u, v) da lente principal permanece inalterada, à medida que (i, j) percorre todo o plano das microlentes. Imagem adaptada de [7].

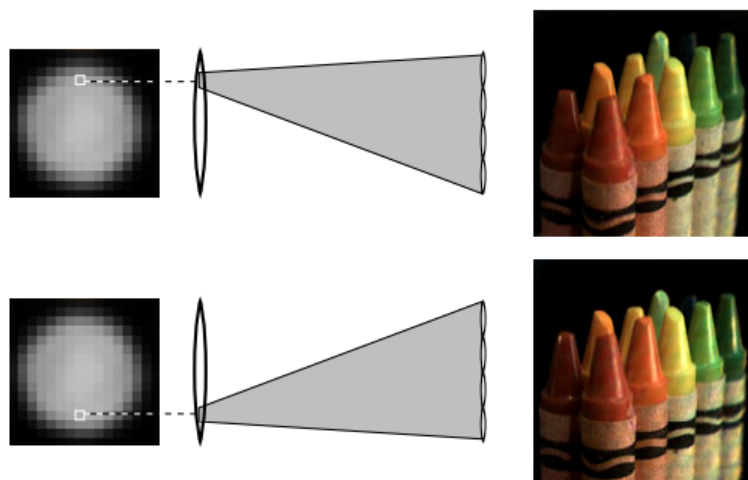


Figura 2.10: Ilustração esquemática de duas diferentes imagens de sub abertura. Cada uma delas se refere a uma posição distinta (u, v) da lente principal (mostrada à esquerda), o que implica uma leve mudança na paralaxe vertical. Adaptada de [7].

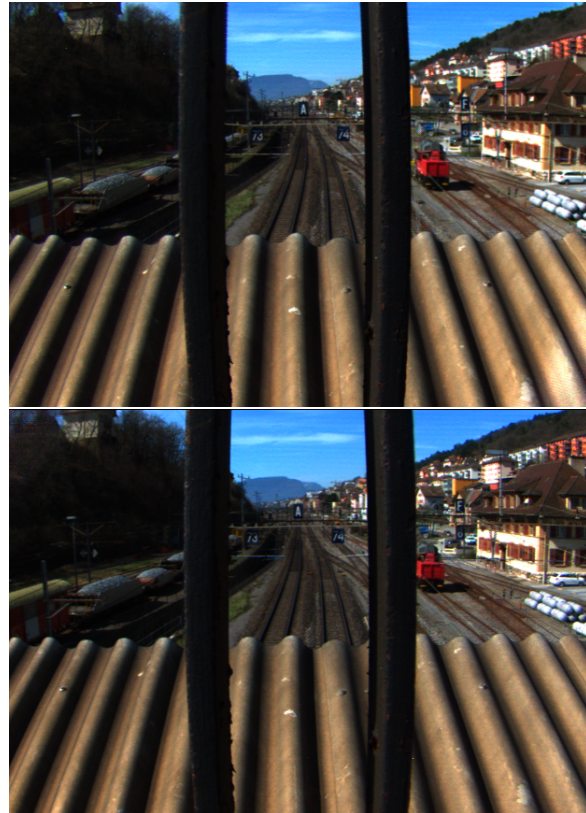


Figura 2.11: Imagens de subabertura referentes ao LF Railway Lines 1. Percebe-se uma pequena mudança do ponto de vista, principalmente à direita da barra de ferro, onde o caminhão aparece, devido à paralaxe vertical.

Nesse sentido, é importante ressaltar que cada *pixel* nessa *full aperture image* é denominado *macropixel*, haja vista que corresponde a um conjunto de *pixels* capturados por uma mesma microlente. Portanto, o número de *macropixels* capturados corresponde à quantidade de microlentes no interior da câmera, que também determina a resolução espacial máxima que cada imagem de subabertura pode ter [10].



Figura 2.12: Recorte do LF completo Sophie & Vincent. A visualização dos *macropixels* permite perceber como são geradas as SAIs: escolhendo-se uma mesma posição dentro de todos os *macropixels*.

A Figura 2.12 representa uma porção do LF completo *Sophie & Vincent*, sendo possível notar a delimitação de cada *macropixel*. Percebe-se que uma mudança na posição (u, v) da lente principal realmente pode gerar uma mudança do ponto de vista de observação. Isso pode ser visto principalmente nas bordas do nariz, pois sua localização pode ser diferente, dependendo de qual *pixel* será escolhido dentro do *macropixel*.

Portanto, pode-se estabelecer que a matriz quadridimensional (4D) *light field* pode ser decomposta em $U \times V$ imagens de subabertura, cada uma com resolução espacial $I \times J$, sendo $0 \leq u \leq U$, $0 \leq v \leq V$, $0 \leq i \leq I$ e $0 \leq j \leq J$. Logo a imagem de subabertura que diz respeito à posição $(i = 0, j = 0)$ será referenciada como $S_{0,0}(u, v)$.

Convém destacar também que existe uma certa incompatibilidade inerente à forma geométrica das lentes (tanto da lente principal, quanto da matriz de microlentes) com relação aos sensores que captam a informação luminosa de fato. Por conseguinte, alguns *pixels* de um mesmo *macropixel* podem apresentar valores de *confidence* diferentes, isto é, alguns desses *pixels* podem ser mais confiáveis, no sentido de que capturam raios de luz mais intensamente do que outros [9], como mostram a Figura 2.13 e a Figura 2.14.

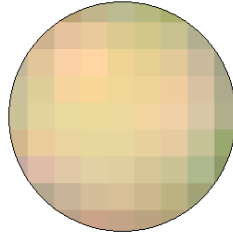


Figura 2.13: Ilustração de uma imagem obtida por uma das microlentes para ilustrar o conceito de *confidence*. Cores mais claras representam áreas onde a luz foi mais intensamente capturada. Nota-se que as bordas da imagem, principalmente, sofrem com problemas de captura da intensidade luminosa, portanto, tendem a ser menos confiáveis. Pode ser tida como um *zoom* de uma porção da imagem da Figura 2.14. Adaptada de [9].

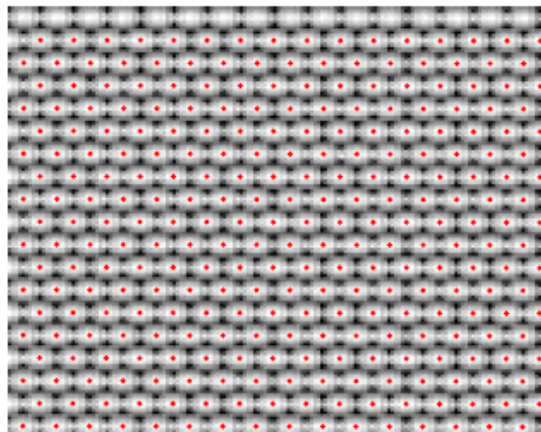


Figura 2.14: Exemplo de imagem branca de calibração capturada em um ambiente controlado. Em vermelho, nota-se o centro estimado para cada microlente. Imagem de [9].

Pode-se então lidar com tal diferença de intensidade se uma imagem completamente branca for capturada em ambiente controlado, como ilustra a Figura 2.14. Essa imagem

pode ser utilizada como um arquivo de configuração para calibrar um sistema que lide com LF, pois, como a imagem é totalmente branca, todos os sensores deveriam capturar a mesma quantidade de luz teoricamente. No entanto, não é o caso, já que a própria localização espacial de cada sensor pode influenciar a aquisição da luz [9].

Dito isso, é conveniente normalizar o valor obtido por cada sensor a partir de uma imagem branca. Este valor pode ser visto como um peso (*weight*, em inglês), o qual é proporcional à quantia de intensidade de luz que cada sensor pode capturar. Inclusive, tais pesos variam entre 0 e 1 e podem ser organizados em um dado bidimensional do tipo $w_{i,j}(u, v)$ [9], correspondendo aos valores de *confidence* mencionados anteriormente.

2.3.1 Imagem integral

A imagem que se forma dentro de uma câmera é proporcional a irradiância que chega no plano focal. Dessa maneira, pode-se usar o dado LF adquirido para se computar fotografias como se elas tivessem sido capturadas por uma câmera convencional posicionada e focada de uma maneira diferente da qual a câmera de aquisição estava [7].

Essa fotografia sintetizada pode ser obtida a partir da luz $LF'(u', v', i', j')$ que passa através da posição (u', v') no plano de abertura sintético e na posição (i', j') no plano focal sintético. A partir disso, a irradiância da imagem formada pode ser descrita pela integral ponderada [41]:

$$E(i', j') = \frac{1}{D^2} \iint LF'(u', v', i', j') A(u', v') \cos^4(\Theta) du' dv' \quad (2.1)$$

na qual D é a distância entre o plano focal e a abertura da lente, A é uma função que descreve a abertura e Θ é o ângulo de incidência do raio (u', v', i', j') com ao plano focal.

Ainda é possível eliminar o termo $\cos^4(\Theta)$ bem como ignorar a constante $1/D^2$ [2], simplificando a equação como:

$$\bar{E}(i', j') = \iint LF'(u', v', i', j') A(u', v') du' dv' \quad (2.2)$$

Definida a equação da imagem integral, é pertinente expressá-la em função do dado LF obtido, $LF(u, v, i, j)$, sendo (u, v) e (i, j) o plano de abertura e o plano focal originais respectivamente. A relação entre LF e LF' é demonstrada na Figura 2.15.

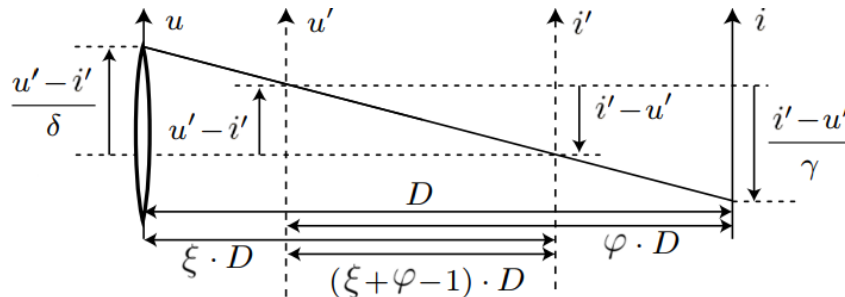


Figura 2.15: Representação bidimensional da relação entre o LF adquirido e uma imagem sintetizada LF' . Percebe-se o mapeamento feito entre uma das coordenadas da abertura original (u) e a abertura sintética (u'). O mesmo vale para os planos focais original (i) e sintético (i'). De modo implícito define-se ξ e φ . Adaptada de [7].

Usando o ξ e φ estabelecidos nessa figura, pode-se definir:

$$\delta = \frac{\xi + \varphi - 1}{\xi} \quad e \quad \gamma = \frac{\xi + \varphi - 1}{\varphi} \quad (2.3)$$

A partir disso e do diagrama da Figura 2.15, verifica-se que o raio o qual intercepta u' e i' também intercepta o plano u no ponto $i' + (u' - i')/\delta$ e o plano i no ponto $u' + (i' - u')/\gamma$. Outrossim o mesmo raciocínio é válido para v e j , resultando em:

$$LF'(u', v', i', j') = LF\left(i' + \frac{u' - i'}{\delta}, j' + \frac{v' - j'}{\delta}, u' + \frac{i' - u'}{\gamma}, v' + \frac{j' - v'}{\gamma}\right) \quad (2.4)$$

Utilizando a Equação 2.4 para substituir o LF' da Equação 2.2, obtém-se a equação de uma fotografia sintetizada em termos do dado LF capturado:

$$\bar{E}(i', j') = \iint LF\left(i' + \frac{u' - i'}{\delta}, j' + \frac{v' - j'}{\delta}, u' + \frac{i' - u'}{\gamma}, v' + \frac{j' - v'}{\gamma}\right) A(u', v') du' dv' \quad (2.5)$$

Assim, nota-se que o processo de sintetização de imagens a partir de um *light field* pode ser entendido como uma simulação fisicamente baseada de uma câmera convencional sintética. Portanto, trata-se de uma imagem integral dos campos canônicos de luz, em que a trajetória da projeção depende da profundidade na qual essa fotografia é focalizada [2]. Essa denominação é pertinente, uma vez que para um mesmo feixe de raios incidentes na lente principal, deve-se ‘somar’ os respectivos *pixels* capturados pelo sensor atrás das microlentes. Mais detalhes a respeito dessa definição podem ser verificados em [2] e [7].

2.3.2 Mudança de perspectiva e de foco

Com o intuito de obter tanto uma nova imagem sintetizada correspondente a uma nova perspectiva quanto uma mudança na abertura da fotografia, é preciso estimar uma imagem sintetizada a partir da imagem original *light field*. Dito isso, a criação de um novo ponto de vista pode ser feita com um sistema multivista, do inglês *multiview system*, ao passo que a mudança do foco digital da cena é um problema mais complexo [10].

É importante esclarecer que imagens de ponto focal (*focal point images*) referem-se ao foco da câmera, isto é, pontos focais distintos podem destacar melhor diferentes objetos ou áreas de uma determinada cena. Esse conceito difere das chamadas imagens de ponto de vista (do inglês *viewpoint image*), as quais são relacionadas a diferentes paralaxes [7], quando uma mesma cena é observada a partir de posições distintas.

Primeiramente, para se analisar diferentes pontos de vista, é preciso considerar diferentes imagens de subabertura. Como dito anteriormente, isso significa selecionar o mesmo *pixel* em cada uma das microlentes e, portanto, cada imagem de subabertura é referente à uma perspectiva específica [7], como já ilustrado na Figura 2.11.

No entanto, pode ser que um observador solicite um ponto de vista que nem tenha sido capturado, ou seja, um ponto de vista sintetizado. Assim, se o conjunto original das posições do observador for suficientemente grande, é possível construir novas perspectivas

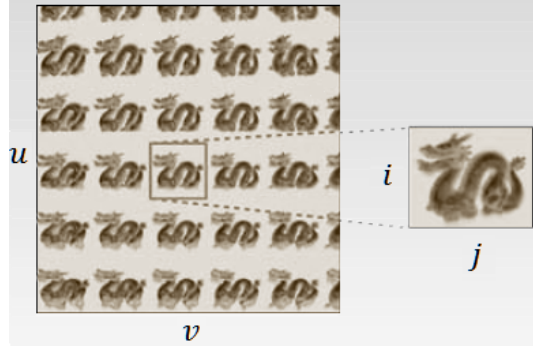


Figura 2.16: Visualização de uma imagem *light field* sendo interpretada como uma coleção 2D de imagens 2D. Cada uma dessas fatias bidimensionais correspondem a uma posição diferente do observador. Imagem adaptada de [10].

de uma imagem, mesmo que o observador não tenha estado nessas posições. Isso pode ser feito selecionando *pixels* apropriados da imagem *light field*, possivelmente utilizando interpolação entre *pixels* próximos [10].

De fato, a cena pode ser observada a partir de qualquer posição desejada, sendo que quanto maior a quantidade de *pixels* atrás de cada microlente, maior será o tamanho da amostra *light field*, promovendo uma discriminação mais fina das perspectivas. Levoy [10] afirma que o cálculo de um novo ponto de vista de uma cena pode ser interpretado como extrair uma fatia 2D posicionada e orientada adequadamente a partir da matriz 4D LF, como ilustra a Figura 2.16.

Por outro lado, mudar o foco em questão significa modificar a profundidade do plano focal (*film plane*), no qual a imagem da lente principal é projetada. Assim, para promover essa mudança de foco (também conhecida como *digital refocusing* ou *synthetic aperture photography* [7, 10]), apenas o plano focal sintético se move, ou seja, $\varphi = 1$. Nesse caso, usando uma abertura completa ($A(u', v') = 1$), tem-se $\delta = 1$ e $\gamma = \xi$, sendo que a equação de uma fotografia sintetizada pode ser simplificada como [7]:

$$\bar{E}(i', j') = \iint LF \left(u', v', u' + \frac{i' - u'}{\xi}, v' + \frac{j' - v'}{\xi} \right) du' dv' \quad (2.6)$$

Analisando a Equação 2.6, verifica-se que o processo de *digital refocusing* se trata de uma soma de versões deslocadas das imagens que se formam através de todos os possíveis pontos (fixando u' e v' e variando i' e j') da abertura uv na lente principal. De maneira quantizada, isso corresponde ao deslocamento e a adição de SAIs [7, 13, 42].

Em suma, a ideia de somar todas as imagens de subabertura, já projetadas no novo plano focal sintético, é que os objetos que estiverem no plano em questão, apareçam mais bem definidos para o observador, enquanto que os objetos correspondentes a outras profundidades apareçam borrados devido à paralaxe [13], que pode ser entendida como um deslocamento aparente de um objeto quando se muda o ponto de observação [43, 44].

A Figura 2.17 mostra imagens com diferentes focos, ao passo que a Figura 2.18 e a Figura 2.19 esclarecem esse processo de mudança de foco. Assim, supondo que o objeto em foco esta no plano M_0 e abreviando o plano da lente principal como u (ao invés de u, v) e o plano focal original como i (ao invés de i, j), a imagem ilustrada na Figura 2.18(a) se formará no plano E_0 (referente à iluminância que o intercepta).

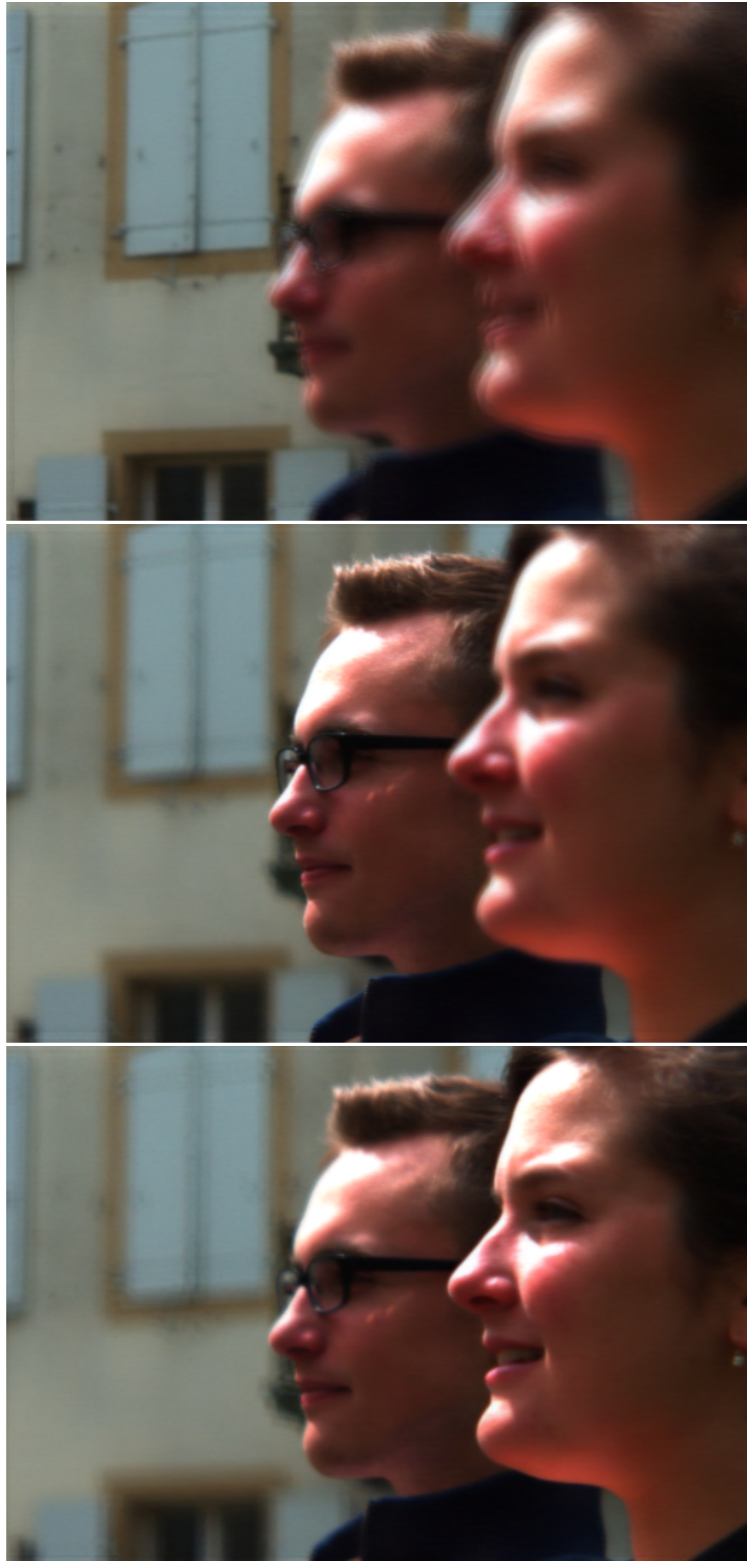
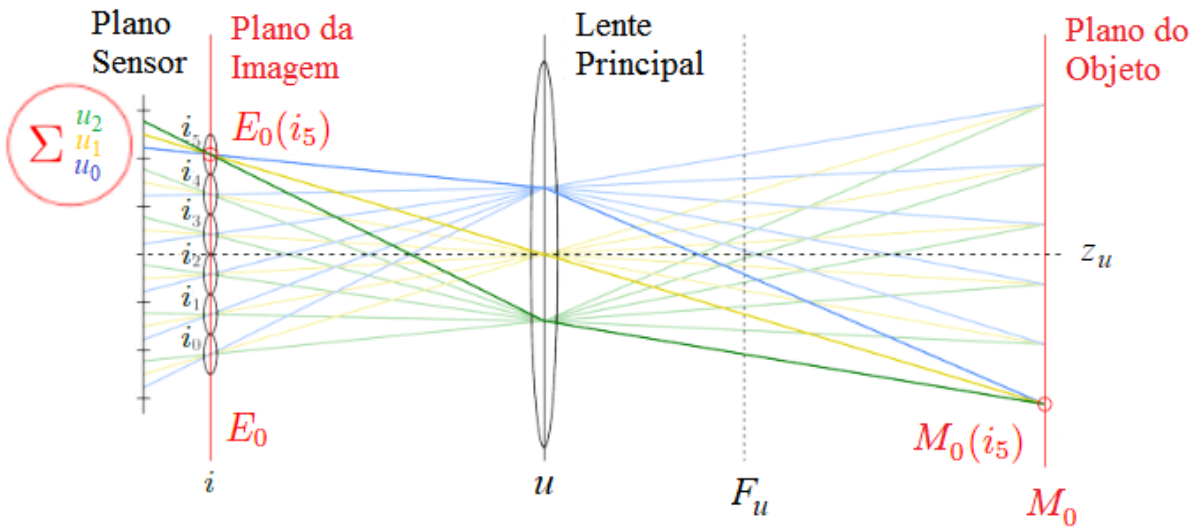


Figura 2.17: Imagens sintetizadas com diferentes focos. De cima para baixo, nota-se respectivamente: o foco nos objetos mais ao fundo (janela por exemplo); depois o foco mais ao meio (rosto do Vincent); e finalmente o foco mais à frente (rosto da Sophie).

E_0 

(a)



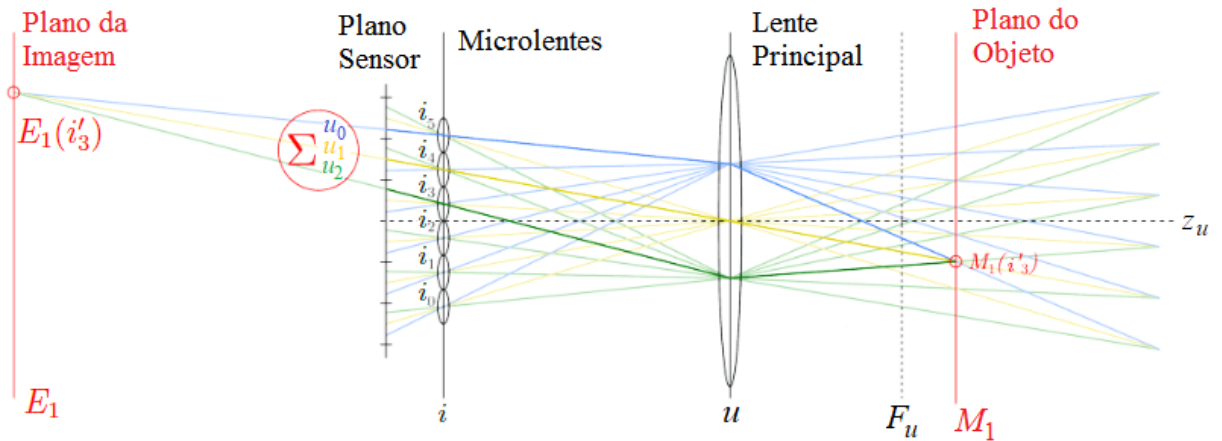
(b)

Figura 2.18: Esquema de sintetização de uma fotografia E_0 com foco em um plano M_0 localizado bem atrás na cena. (a) Objetos mais ao fundo são focalizados. (b) O plano focal (plano da imagem) coincide com o plano das microlentes (MLA). Adaptada de [11, 12].

Para formar essa imagem E_0 é preciso combinar as porções da imagem formadas em $E_0(i_0)$, $E_0(i_1)$, $E_0(i_2)$, $E_0(i_3)$, $E_0(i_4)$ e $E_0(i_5)$ [11]. A Figura 2.18(b) exhibe apenas $E_0(i_5)$ por simplicidade. Nesse sentido, os *pixels* que devem ser somados à cada direção de incidência do feixe de raios no plano u estão atrás do plano das microlentes. Assim, para formar $E_0(i_0)$, por exemplo, é preciso considerar os raios (u_0, i_0) , (u_1, i_0) e (u_2, i_0) .

E_1 

(a)



(b)

Figura 2.19: Representação de como focalizar um plano M_1 mais próximo da câmera do que M_0 . (a) Imagem sintetizada E_1 com foco em objetos mais à frente na cena. (b) A imagem se forma atrás da matriz de microlentes. Adaptada de [11, 12].

Comparativamente, se o foco estiver em um plano M_1 mais próximo da câmera, uma imagem E_1 , ilustrada pela Figura 2.19(a), se formará atrás do MLA. Desse modo, para compor E_1 em um plano focal sintético i' é preciso considerar todas as porções $E_1(i')$. A formação de $E_1(i'_3)$, por exemplo, leva em consideração os raios (u_2, i_3) , (u_1, i_4) e (u_0, i_5) , como mostra a Figura 2.19(b). Nesse caso, os *pixels* que devem ser selecionados para gerar o foco em questão estão localizados à frente do plano da imagem e estão espalhados em mais de uma microlente. Mais detalhes podem ser verificados em [45].

Resumindo todo esse processo, para sintetizar uma imagem referente a um novo foco é preciso:

1. Dado um conjunto $\mathcal{S}(u, v)$ de imagens de subabertura, aplicar os respectivos pesos luminosos w_i em todas as SAIs $\mathbf{x} \in \mathcal{S}(u, v)$, ou seja:

$$\mathcal{S}^*(u, v) = \sum_i \mathbf{x}_i \times w_i; \quad (2.7)$$

2. Calcular a profundidade relativa $d_\alpha = \frac{\Delta_z}{\Delta_z + Z_0}$, sendo Δ_z a distância entre o plano P' que define a nova profundidade e o plano P original (com P' paralelo a P), ao passo que Z_0 se refere à distância entre P e a lente principal, como ilustra a Figura 2.20;

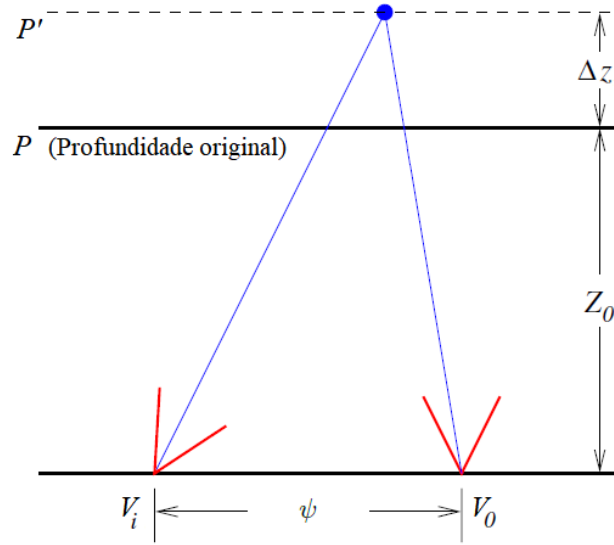


Figura 2.20: Mudança de profundidade entre planos paralelos e paralaxe entre diferentes imagens de subabertura. P é o plano de referência, isto é, a profundidade original e P' refere-se a uma profundidade sintética mais distante da câmera do que P . Os pontos de vista V_i e V_0 representam diferentes imagens de subabertura, com diferentes paralaxes. Essa paralaxe depende tanto do deslocamento ψ quanto da profundidade relativa $\frac{\Delta_z}{\Delta_z + Z_0}$. Adaptada de [13].

3. Calcular a translação espacial $\Delta_\alpha^u = \psi_u \cdot d_\alpha$ e $\Delta_\alpha^v = \psi_v \cdot d_\alpha$ para cada *pixel* dos *macropixels*, sendo ψ_u e ψ_v os deslocamentos relativos dos *pixels* dentro do *macropixel* em cada uma das respectivas direções;
4. A partir do conjunto $\mathcal{S}^*(u, v)$, utilizando interpolação linear, deve-se computar os valores interpolados de \mathcal{S}^* na posição $(u + \Delta_\alpha^u, v + \Delta_\alpha^v)$, gerando um novo conjunto $\mathcal{S}^*(u_\alpha, v_\alpha) = \mathcal{S}^*(u + \Delta_\alpha^u, v + \Delta_\alpha^v)$ de imagens de subabertura deslocadas.
5. Adicionar todas as SAIs $\mathbf{y}_i \in \mathcal{S}^*(u_\alpha, v_\alpha)$, obtendo $I_\alpha(u, v) = \sum_i \mathbf{y}_i$;
6. Normalizar a imagem resultante I_α correspondente a um novo ponto focal.

Entretanto, neste trabalho não se tem a informação precisa da mudança de profundidade aplicada, isto é, a distância Δ_α existente entre P original e P' . O que se tem é o *slope* alvo α , já explicado na Subseção 2.2.1, o qual é mapeado para a profundidade relativa d_α da seguinte maneira:

$$d_\alpha^r = \alpha \cdot \mathbf{tamanho}_r \cdot k_r \quad (2.8)$$

sendo $\mathbf{tamanho}_r$ a quantidade de *pixels* na direção r e k_r um vetor composto por $\mathbf{tamanho}_r$ valores discretos. Essa profundidade é específica para cada SAI, uma vez que essas imagens podem ter sido deslocadas para outro *slope* anteriormente.

Assim, é preciso computar tanto o deslocamento d_α^i referente às linhas (na direção i , ou seja, $r = i$), quanto o deslocamento d_α^j pertinente às colunas ($r = j$), assumindo que i e j sejam as coordenadas do plano das microlentes. Se uma imagem de subabertura possuir $I \times J$ *pixels*, existem I *pixels* na direção $r = i$ ($\mathbf{tamanho}_i = I$) e J *pixels* na direção j ($\mathbf{tamanho}_j = J$).

Além disso, sabe-se que o vetor k_r possui $\mathbf{tamanho}_r$ valores discretos e que, para este trabalho, eles variam de $-0,5$ até $0,5$, sendo 0 o valor que representa a imagem de subabertura do ponto central [9]. Convém esclarecer que os valores mínimo e máximo do vetor k_r dependem da câmera LF utilizada.

2.4 Técnicas básicas de predição para a compressão de imagens e vídeos

Fundamentalmente, convém descrever o conceito de codificação preditiva. Gonzalez e Woods [46] a definem como uma técnica que possibilita uma boa compressão sem um custo computacional significativo, podendo ser com ou sem perdas. Assim, essa codificação busca eliminar redundâncias de amostras com pouco espaçamento entre si, seja de acordo com o tempo e/ou com o espaço. Codifica-se apenas as diferenças entre o valor atual e o valor previsto de uma amostra, em particular um *pixel*.

Dito isso, a entrada do processo de codificação pode ser um bloco convencional de uma imagem, ou a diferença entre um bloco convencional e uma predição baseada em blocos similares de *frames* (quadros) anteriores ou subsequentes de um mesmo vídeo [47]. Portanto, deve-se comprimir três tipos básicos de *frames*, mostrados na Figura 2.21: os *frames-I*, os *frames-P* e os *frames-B*.

Os *frames-I* são os chamados *frames* independentes ou *intraframes* e referem-se a *frames* comprimidos de maneira independente de todos os outros quadros de vídeo anteriores e futuros, sendo que a predição de um *pixel* é feita internamente, a partir de *pixels* vizinhos. Em geral a compressão desse tipo de quadro não é tão eficiente quanto a compressão dos chamados *interframes* (*frames P* ou *B*). Normalmente *frames-I* são semelhantes a imagens codificadas por JPEG [48] e são usados como referência para as predições futuras *P* e *B*.

Por outro lado, *frames-P* são os quadros preditivos e dependem de um ou mais quadros anteriores. De fato, são a diferença comprimida entre o *frame* atual e uma predição dele baseada no *frame-I* ou *-P* anterior [47]. Finalmente, os *frames-B* são os *frames* bidirecionais, isto é, baseiam-se tanto em *frames* anteriores quanto em *frames* futuros.

Esses *frames-B* possuem menos *bits* que os demais, mas envolvem uma complexidade computacional maior.

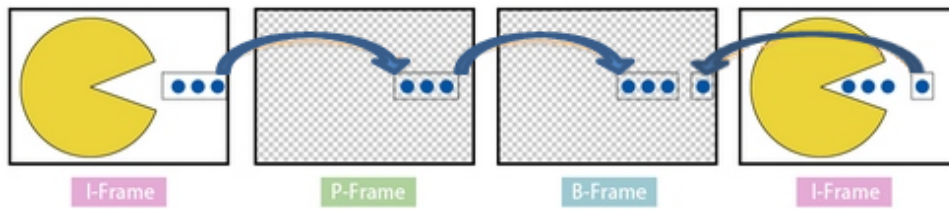


Figura 2.21: Ilustração dos três diferentes tipos de *frames* referentes à codificação preditiva. Nas bordas da imagem, estão os *frames-I*, os quais são codificados independentemente dos outros *frames* do vídeo ao qual pertencem. Mais ao meio e hachurados estão os *frames-P* e os *frames-B* preditos. Imagens adaptada de [14].

2.5 H.264

O H.264, também chamado de AVC e de MPEG-4 Part 10 (*Moving Picture Experts Group*), é um padrão de codificação de vídeo que apresenta uma melhoria significativa com relação a padrões anteriores como MPEG-1/2 [49]. Nesse esquema existem quadros intra, cuja codificação não considera quadros anteriores, e quadros inter, os quais são obtidos pela subtração de uma predição compensada de movimento a partir do quadro original. Os resíduos são então transformados para o domínio da frequência e os coeficientes dessa transformada são escaneados, quantizados e codificados usando-se códigos de comprimento variável. Um decodificador local reconstrói o quadro para que ele possa ser utilizado para predições subsequentes.

A estrutura de macroblocos consiste em quatro blocos de luminância 8×8 e dois blocos de crominância, sendo que um número inteiro de macroblocos pode ser combinado. Nos padrões anteriores a menor subdivisão de um macrobloco era de 8×8 sub-blocos, mas o padrão H.264 permite que sejam feitas partições nos macroblocos 8×8 para que sejam divididos posteriormente em submacroblocos de tamanho 8×4 , 4×8 e 4×4 [49]. Esses blocos menores podem ser usados para a predição de compensação de movimento, possibilitando uma busca por detalhes bem mais refinada do que a feita pelos outros padrões [49]. Além da partição 8×8 , os referidos macroblocos podem ser particionados em dois blocos de 8×16 ou de 16×8 , havendo a possibilidade de combiná-los para formar macroblocos de 16×16 .

Distintivamente dos esquemas de codificação anteriores, os quais usam uma transformada DCT 8×8 , o AVC utiliza uma matriz 4×4 semelhante à DCT, descrita por:

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & 2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \quad (2.9)$$

É importante saber que o processo de codificação adotado pelo AVC usa um quantizador escalar uniforme que apresenta 52 passos de quantização (QP) possíveis. Por fim, pode-se dizer que o H.264 é consideravelmente mais flexível que os padrões prévios, com uma gama maior de aplicações [49]. Diante desses padrões, ele apresenta melhores capacidade de predição de movimento e método de codificação de entropia, conseguindo reduzir em aproximadamente 50% a taxa de transmissão para qualidades equivalentes [50].

2.6 Padronização da compressão de *light field*

Embora não exista um padrão estabelecido para a compressão de imagens *light field*, há uma nova iniciativa denominada JPEG PLENO [37], a qual visa padronizar a representação de novas modalidades de imagens, como *light field*, *point-cloud* e imagens holográficas. A ideia é, não só melhorar a compressão dessas imagens, mas também promover ferramentas avançadas para a manipulação desses dados, metadados, privacidade e segurança. Assim, o JPEG PLENO busca acompanhar a inovação de tais imagens, levando também em consideração os formatos legados JPEG [48], como JPEG 2000 [51].

Nesse sentido, essa representação, que ainda está em desenvolvimento, precisa se adaptar, pois as imagens convencionais são representadas como matrizes de duas dimensões em um dado instante de tempo, no qual alguns atributos, como o foco, são fixos e relacionados a uma determinada captura. Portanto, a referida adaptação busca estabelecer um modo de gravar simultaneamente os valores de intensidade e de cor referentes a múltiplas perspectivas e distâncias [37].

Adicionalmente, o JPEG PLENO considera fundamental que esses novos formatos de representação possuam interoperabilidade com os formatos de imagens mais difundidos, como JPEG e JPEG 2000 [37]. Porém, a interação pode inclusive ser limitada, no sentido de que tal compatibilidade pode permitir, por exemplo, que um *browser* existente ou algum dispositivo seja capaz de visualizar imagens convencionais a partir de uma imagem *light field* [37]. Por fim, essa padronização almeja, se possível, estabelecer um formato unificado para as “novas” modalidades de imagem, o que facilitaria uma possível troca de informações entre elas.

Apesar dessa iniciativa, diversos outros trabalhos vêm sendo desenvolvidos com o intuito de comprimir e transmitir os dados LF de maneira eficaz e eficiente. Alguns deles serão abordados no próximo capítulo, que trata da revisão bibliográfica.

Capítulo 3

Revisão Bibliográfica

Neste capítulo serão descritos os trabalhos previamente realizados pertinentes à transmissão de imagens sintetizadas de *light field* (LF). Inicialmente, é preciso tratar da compressão dessas imagens, haja vista que o volume de dados envolvido nesse caso é muito maior com relação a uma imagem RGB (*Red Green Blue*) convencional de resolução espacial semelhante. Dito isso, várias maneiras de codificar imagens LF têm sido propostas.

Alves et al. [16] usam os principais codificadores disponíveis para comprimir imagens obtidas a partir do *light field* completo. Assim, os autores comparam o desempenho do HEVC [52] (*High Efficiency Video Coding*) Intra, do H.264/AVC [53] Intra (do inglês *Advanced Video Coding*), do JPEG e do JPEG2000, concluindo que o HEVC Intra se mostra como o codificador mais eficiente em termos de BD-Taxa [19] (métrica de avaliação definida por Bjøntegaard).

Os autores detalham que seus experimentos foram feitos com imagens obtidas por uma câmera do tipo Lytro [15], as quais são de um formato *raw* (basicamente são dados brutos, sem tratamento) e necessitam de um processo de deciptação (*demosaicing*). Essa deciptação permite a obtenção da informação de cor dos dados obtidos, gerando-se imagens RGB, com a mesma resolução.

Diante disso, o processo de exibição de uma imagem particular 2D, a partir de uma imagem *light field*, em um *display* convencional requer dois passos básicos: deciptação e renderização, ilustrados na Figura 3.1. Como o primeiro passo já foi descrito, cabe destacar que o passo de renderização recebe uma imagem RGB deciptada e fornece uma vista bidimensional renderizada referente a uma perspectiva ou a um foco específicos.

É importante tratar também do trabalho feito por Aron et al. [21], os quais afirmam que, além da compressão eficiente, é interessante ter um fácil acesso randômico à essas imagens no momento da renderização. Porém, isso geralmente não é possível por conta das próprias dependências criadas entre os dados devido à predição associada ao processo de compressão. Assim, os autores propõem um codificador que elimine essa predição, sendo que as imagens são codificadas independentemente entre si por um codificador Wyner-Ziv [54]. No decodificador, as imagens previamente reconstruídas são usadas como informação adicional para explorar as similaridades entre as imagens.

Em suma, as imagens *light field* são divididas em imagens chave (K), que são comprimidas usando padrões convencionais de codificação, e imagens W , que são codificadas usando um codificador Wyner-Ziv. No codificador, os autores aplicam a Transformada Discreta de Cosseno, do inglês *Discrete Cosine Transform* (DCT), à imagem W , gerando

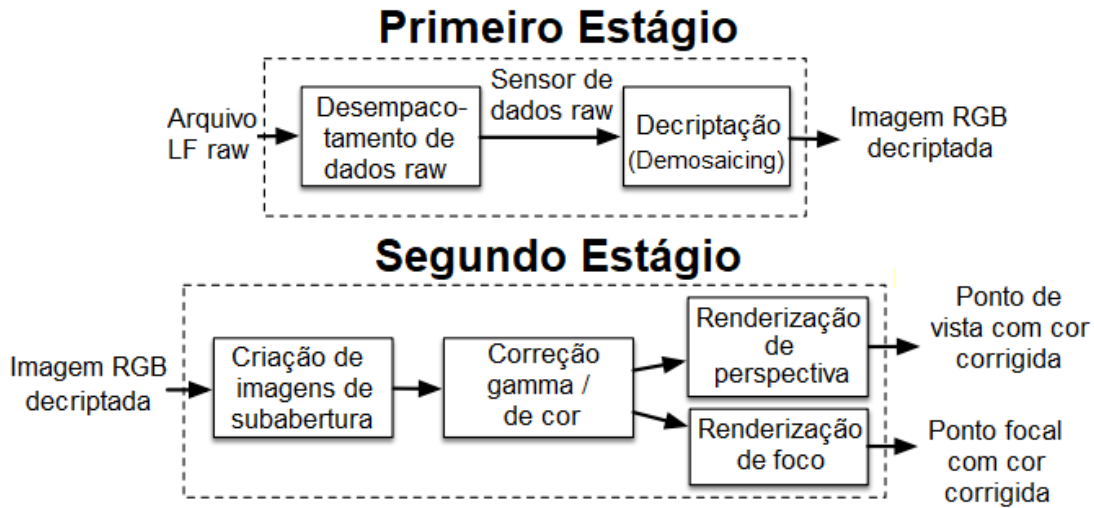


Figura 3.1: Ilustração do processo de decodificação (*demosaicing*) de uma imagem *light field* no formato *raw*, capturada por uma câmera Lytro [15], e renderização de uma vista bidimensional. Imagem de [16].

uma imagem X . Com isso, formam-se bandas de coeficientes X_k , sendo k o número de determinado coeficiente. Portanto, cada banda de coeficientes de transformação são codificados independentemente.

Esses coeficientes passam por um codificador Slepian-Wolf [55], o qual produz *bits* de paridade, que são armazenados como a versão comprimida da imagem Wyner-Ziv. Aron et al. [21] explicam que tais *bits* de paridade são utilizados para decodificar as imagens e que, por conta dessa codificação independente, afirmam que esse esquema possui características semelhantes ao acesso randômico apresentado pela codificação intra convencional.

Chang et al. [56] aplicam uma compressão *wavelet* nos dados LF com base na compensação de disparidades (do inglês *disparity-compensated lifting*) e na identificação de formas. Usando como referência um codificador de mapas de textura [57], apresentam ganhos de 6 dB, para uma mesma quantidade de *bits* transmitidos. Para uma mesma qualidade, os autores conseguem uma redução de 70% em termos de taxa.

Santos et al. [58] promovem uma compressão por meio de transformações de cores reversíveis e conseguem uma taxa de até 6,41 bpp (*bits por pixel*), a qual é 0,91 bpp menor em comparação com HEVC, 1,72 bpp menor frente ao JPEG-LS e 2,43 bpp menor com relação ao JPEG2000.

Tran et al. [59] propõem um *framework* de compressão que se baseia na compensação de movimentos (*motion-compensated lifting*), utilizando a transformada *wavelet*. O ganho médio auferido em termos da compressão sem perdas é de até 7,3% em comparação com o JPEG2000 e de até 25% com relação ao JPEG-LS. Os autores ainda relatam que, no caso da compressão com perdas, os resultados do *framework* proposto apresentam uma menor distorção frente ao JPEG2000.

Chao et al. [60] propõem um esquema de compressão o qual é aplicado diretamente no *raw LF* original, isto é, nos dados brutos LF antes do processo de decodificação/*demosaicing*, reduzindo em mais de 55% o tamanho dos dados em comparação com a codificação individual das imagens de subabertura.

Zhao e Chen [61] propõem um codificador que utiliza aproximação linear e conseguem um ganho médio de BD-Taxa [19] de 37,41% em comparação com o HEVC.

Mesmo utilizando essas técnicas de compressão, a transmissão do LF completo é desafiadora [17]. Assim, se um servidor transmitir todo o conjunto de dados LF para o cliente por meio de redes com bandas limitadas, antes mesmo da sintetização e da observação do lado do usuário, o atraso de inicialização será significativo.

Para evitar essa transmissão do LF inteiro, trabalhos anteriores acerca de transmissão interativa de *light field* (ILFS, do inglês *Interactive Light Field Streaming*) [20–24] foram apresentados. Nessas propostas, um usuário periodicamente requisita uma vista específica, sendo que o servidor responde transmitindo uma vista pré-sintetizada e codificada para ser observada. A vantagem é que o usuário pode, de maneira imediata, verificar as imagens de ponto de vista apenas utilizando uma decodificação padrão, sem haver um longo tempo de inicialização relacionado à transmissão do conjunto completo de dados [17]. Mais recentemente, Yuan et al. [62] apresentaram uma versão atualizada do ILFS.

A despeito disso, a transmissão de apenas imagens sintetizadas codificadas faz com que o usuário nunca “aprenda” as imagens de subabertura utilizadas no processo de sintetização, isto é, o cliente não recebe e sequer sabe quais SAIs foram usadas. Portanto, o usuário sempre espera que o servidor transmita cada uma das imagens de ponto de vista, resultando em um grande custo agregado [17] de transmissão, à medida que o número de requisições de ponto de vista aumenta.

Diante dessa conjuntura, Peixoto et al. [17] propuseram uma alternativa denominada Comunicação Progressiva de *Light Field* (PLFC, do inglês *Progressive Light Field Communication*). Tal proposta é a base deste trabalho e nela, para cada imagem de ponto de vista (ou ponto focal) sintetizada transmitida, o usuário pode “decodificar” e recuperar a imagem de subabertura correspondente. Essa SAI recuperada é inserida na *cache* do usuário para sintetizações futuras.

Os autores definem ainda que essa *cache* deve ser inicializada com um determinado número de SAIs e enviada ao cliente no início da transmissão, o que implica um atraso de inicialização (*startup delay*) que não existe no ILFS. Mesmo assim esse tempo é expressivamente menor que o tempo necessário para transmitir o LF completo. Em contrapartida, Peixoto et al. [17] afirmam que a grande vantagem é que, conforme mais imagens vão sendo requisitadas, a *cache* local se torna grande o suficiente para que o usuário sintetize as vistas por conta própria, o que reduz o custo de transmissão acumulado.

Outro trabalho que merece destaque relacionado à transmissão de LF é o desenvolvido por Liu et al. [63], o qual propõe um sistema de transmissão composto por um servidor e por vários clientes. Essas *streams* de vídeo múltiplas são codificadas em tempo real, de modo que cada cliente possa selecionar as *streams* que quiser para a renderização de uma nova vista. Os autores acreditam inclusive que a sua proposta se trata do primeiro sistema de *streaming* em tempo real que consegue simultaneamente garantir o livre controle dos pontos de vista em tempo real, permitir armazenamento de dados bem como sustentar um número arbitrário de usuários.

Finalmente, embora não seja o foco desta proposta, convém citar alguns trabalhos os quais abordam aplicações práticas envolvendo LF, ao invés de tratarem da compressão e da transmissão dessas imagens. Kara et al. [64] discutem a viabilidade, escalabilidade do cinema LF, comentando sobre realidade virtual e também a respeito da possibilidade do cinema tridimensional sem uso de óculos.

Moghaddam et al. [65] tratam do reconhecimento facial em LF. Os autores propõem um método que utiliza a informação angular do LF como um descritor de padrões binários LFLBP (do inglês *Light Field Local Binary Patterns*), que pode ser combinado com um descritor espacial SLBP (do inglês *Spatial Local Binary Pattern*), para melhorar o desempenho do processo de reconhecimento facial.

Wang et al. [66] lidam com a detecção de objetos, segmentação de imagens e interpolação de vistas em imagens *light field*. O reconhecimento de materiais em imagens LF realizado pelos autores apresentou uma taxa de sucesso 7% maior em relação à classificação de materiais utilizando apenas uma vista, uma imagem convencional 2D.

O Capítulo 4 apresentará o método proposto por este trabalho para transmitir progressivamente imagens sintetizadas de *light field*, o qual é baseado no referido PLFC [17].

Capítulo 4

Metodologia

Este capítulo explicita o método usado para transmitir imagens sintetizadas *light field*. Ele é baseado em um trabalho prévio [17], que será resumido no início do capítulo. Feito isso, serão abordadas as modificações propostas neste trabalho. Visa-se, essencialmente, aperfeiçoar o processo de escolha das imagens de subabertura que devem ser enviadas, propondo-se também uma nova maneira de inicializar a *cache*.

Mais especificamente, a metodologia pode ser dividida em 5 grandes passos: (i) modelar a inicialização da *cache* com a quantidade mínima de SAIs requeridas para uma qualidade particular, (ii) promover um algoritmo de refinamento para a seleção de SAIs, (iii) selecionar dinamicamente o QP (Passo de Quantização, do inglês *Quantization Parameter*) durante a codificação, (iv) diminuir a carga computacional requerida, modificando levemente a maneira de calcular o benefício futuro e (v) definir uma maneira para determinar o multiplicador de Lagrange automaticamente durante um processo de otimização.

4.1 Comunicação Progressiva de *Light Field*

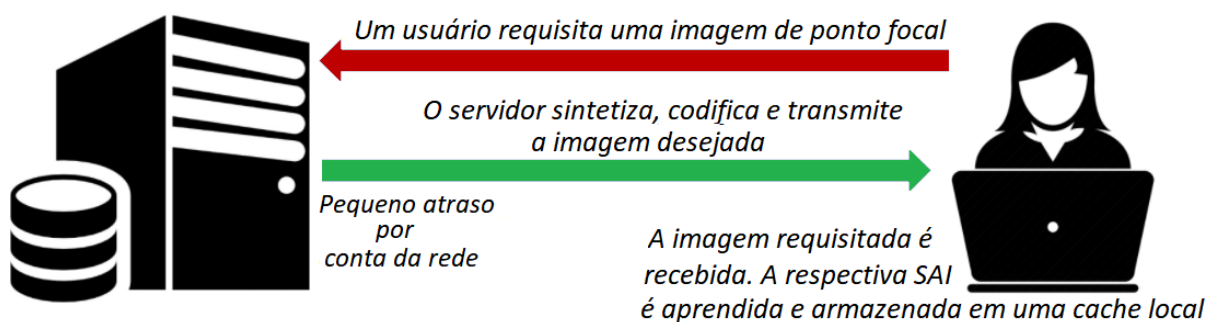


Figura 4.1: Ilustração do funcionamento do trabalho desenvolvido por Peixoto et al. [17]. A mesma ideia geral é válida para a proposta deste trabalho. A ser publicada em [18].

Antes de expor as contribuições deste trabalho convém descrever a Comunicação Progressiva de *Light Field* (PLFC, do inglês *Progressive Light Field Communication*). Como a Figura 4.1 mostra, basicamente um usuário define uma qualidade alvo e requisita diversas imagens de ponto focal (seta superior), por meio da escolha do *slope* correspondente,

o qual já foi explicado na Seção 2.2.1. Para cada requisição, o servidor tenta sintetizar a imagem desejada utilizando as SAIs na *cache* de usuário, \mathcal{C} .

Se \mathcal{C} for suficiente para conseguir a qualidade alvo, o servidor precisa apenas instruir o usuário em como sintetizar o ponto focal requisitado, sem a necessidade de transmitir nenhuma imagem adicional. Caso contrário, o servidor precisa selecionar uma nova SAI com o intuito de alcançar a qualidade pedida (seta inferior).

Uma vez que os dados LF podem ser decompostos em imagens de subabertura, pode-se definir \mathcal{S}^o como o conjunto completo de todas as SAIs e $|\cdot|$ como o comprimento de um conjunto de imagens. Imagens de diferentes vistas ou de diferentes pontos focais podem ser sintetizadas usando \mathcal{S}^o . Vale destacar que o PLFC original (assim como esta proposta) trabalha apenas com a sintetização de novos pontos focais, mas a adaptação do sistema para gerar novas vistas é possível e até relativamente mais simples.

Para gerar um novo foco, pode-se utilizar uma combinação linear de versões deslocadas dessas SAIs para se obter a respectiva imagem sintetizada [7, 13, 17, 26]. Assuma que o usuário deseje uma nova imagem de ponto focal representada pelo *slope* α . Como resultado do processo de mudança de foco explicado na Subseção 2.3.2, a imagem sintetizada, $\mathbf{v}(\alpha, \mathcal{S}^o)$, pode ser obtida por:

$$\mathbf{v}(\alpha, \mathcal{S}^o) = \frac{\sum_{i=1}^{|\mathcal{S}^o|} w_i^\alpha \cdot \mathbf{x}_i^\alpha}{\sum_{i=1}^{|\mathcal{S}^o|} w_i^\alpha} \quad (4.1)$$

na qual w_i é uma matriz de pesos associada a cada SAI \mathbf{x}_i , $\mathbf{x}_i \in \mathcal{S}^o$, e \mathbf{x}_i^α significa a versão deslocada de \mathbf{x}_i para o *slope* α . As matrizes w_i são específicas de cada câmera e dizem respeito a pesos luminosos, os quais são proporcionais à intensidade de luz adquirida por cada *subpixel* dentro de um *macropixel*. Desse modo, w_i^α referem-se aos respectivos pesos ajustados aplicados à versão deslocada \mathbf{x}_i^α .

Dito isso, se um subconjunto de imagens de subabertura $\mathcal{S} \subset \mathcal{S}^o$ fosse utilizado, ao invés do conjunto completo \mathcal{S}^o , então a imagem virtual $\mathbf{v}(\alpha, \mathcal{S})$ não será idêntica à $\mathbf{v}^o(\alpha, \mathcal{S}^o)$. Logo, existe uma distorção intrínseca à imagem de ponto focal \mathbf{v} ao se usar uma combinação linear de SAIs em \mathcal{S} com relação a \mathcal{S}^o :

$$D_{\mathbf{v}}(\mathcal{S}) = \|\mathbf{v}(\alpha, \mathcal{S}) - \mathbf{v}^o(\alpha, \mathcal{S}^o)\|_2^2. \quad (4.2)$$

Assim, a Equação (4.1) pode ser modificada para:

$$\mathbf{v}(\alpha, \mathcal{S}) = \frac{\sum_{i=1}^{|\mathcal{S}|} p_i \cdot w_i^\alpha \cdot \mathbf{x}_i^\alpha}{\sum_{i=1}^{|\mathcal{S}|} w_i^\alpha} \quad (4.3)$$

sendo p_i pesos escalares aplicados a todos os *pixels* em \mathbf{x}_i^α . Tais valores são obtidos por meio de um processo de otimização com o intuito de minimizar a distorção D entre $\mathbf{v}(\alpha, \mathcal{S})$ e $\mathbf{v}^o(\alpha, \mathcal{S}^o)$ [17].

A partir disso, uma *cache* inicial é construída e transmitida para o usuário antes mesmo de qualquer imagem ser requisitada. No PLFC original essa *cache* inicial possuía um tamanho fixo o qual dependia da qualidade exigida pelo usuário. Tal valor fixo era determinado empiricamente.

Portanto, uma vez que o usuário (‘decodificador’, ou receptor) solicita uma imagem de ponto focal, o servidor (‘codificador’) deve decidir se a *cache* já é suficiente ou não, como será explicado a seguir.

4.1.1 Usando apenas a *cache*

Se as imagens contidas na *cache* forem suficientes para que a qualidade desejada seja alcançada, então o servidor precisa somente instruir o cliente a como sintetizar o foco correspondente. Nesse caso, apenas os valores adequados de p_i , sendo $1 \leq i \leq |\mathcal{C}|$, são transmitidos para o usuário de modo que se faça a devida combinação linear das SAIs, conforme a Equação (4.3).

4.1.2 Selecionando novas imagens para transmissão

Se a *cache* atual não for suficiente, há a necessidade de transmissão. Antes de enviar uma imagem, o servidor precisa selecionar uma SAI, denominada \mathbf{z} , fazendo uma análise entre os benefícios imediato e futuro que cada SAI pode gerar.

Supondo que o usuário tenha a *cache* \mathcal{C} , a \mathbf{z} selecionada deve contribuir para síntese atual da imagem de ponto focal \mathbf{v} . Isso é denominado o benefício imediato e pode ser descrito por:

$$B_v^i(\mathcal{C}, \mathbf{z}) = D_v(\mathcal{C}) - D_v(\mathcal{C} \cup \{\mathbf{z}\}) \quad (4.4)$$

Em contrapartida, \mathbf{z} deve também contribuir para os pontos focais sintetizados que serão requisitados no futuro. Se \mathcal{V} representar as imagens de ponto focal já requisitadas pelo usuário e \mathcal{V}^o representar todas as imagens que podem ser solicitadas pelo usuário em uma sessão de *streaming* (transmissão) tem-se que o benefício futuro associado a \mathbf{z} pode ser calculado como:

$$B^f(\mathbf{z}, \mathcal{C}, \mathcal{V}) = \sum_{\mathbf{u} \in \mathcal{V}^o \setminus \mathcal{V}} \theta(\mathbf{u}, \mathbf{v}, t) (D_{\mathbf{u}}(\mathcal{C}) - D_{\mathbf{u}}(\mathcal{C} \cup \{\mathbf{z}\})) \quad (4.5)$$

sendo que $\theta(\mathbf{u}, \mathbf{v}, t)$ é a importância dos pontos focais \mathbf{u} ainda não requisitados e t é a quantidade de *slopes* já requisitados.

Mais detalhadamente, essa importância θ foi definida a partir de um modelo de interação visual proposto por [17]. Assumindo que o usuário solicite um total de T pontos focais (dentro os $|\mathcal{V}^o|$ possíveis), tem-se que a importância de um ponto focal $\mathbf{u} \in \mathcal{V}^o \setminus \mathcal{V}$ durante a *streaming* pode ser computada como:

$$\theta(\mathbf{u}, \mathbf{v}, t) = \sum_{\tau=1}^{T-t} [\mathbf{1}_v \mathbf{P}^\tau]_{\mathbf{u}} \quad (4.6)$$

onde $\mathbf{1}_v$ é o vetor linha canônico de tamanho $|\mathcal{V}^o|$ cujo valor de todas as entradas é 0, com exceção da \mathbf{v} -ésima entrada, que é 1. O operador $[\cdot]_{\mathbf{u}}$ designa a \mathbf{u} -ésima entrada de um vetor e, além disso, a probabilidade de haver mudança de uma vista \mathbf{a} para uma vista \mathbf{b} é $p_{\mathbf{a}, \mathbf{b}}$, que corresponde à (\mathbf{a}, \mathbf{b}) -ésima entrada da matriz \mathbf{P} de transição de probabilidade. Traduzindo em palavras, a Equação (4.6) reflete a soma das probabilidades de que um usuário selecione um ponto focal \mathbf{u} em cada uma das T mudanças [17].

As considerações acima acerca dos dois benefícios, Equação (4.4) e Equação (4.5), podem então ser combinadas em um único critério para escolher \mathbf{z} :

$$\max_{\mathbf{z} \in \mathcal{S} \setminus \mathcal{C}} B_{\mathbf{v}}^i(\mathcal{C}, \mathbf{z}) + \mu B^f(\mathbf{z}, \mathcal{C}, \mathcal{V}) \quad (4.7)$$

em que μ é o multiplicador de Lagrange que controla o *trade-off* entre os benefícios atual e futuro, sendo também empiricamente determinado. Vale destacar que, embora tenha proposto esse benefício futuro, o PLFC não o implementa de maneira satisfatória, atribuindo um peso muito maior ao benefício imediato.

Uma vez que \mathbf{z} é finalmente selecionada, o respectivo ponto focal sintetizado é codificado e transmitido para o usuário por meio de codificação residual. Esse processo pode se repetir até que a qualidade alvo seja atingida.

4.1.3 O sistema aprende a SAI

Assumindo que um servidor e um usuário conheçam um subconjunto \mathcal{C} de SAIs e que todas as matrizes de peso w_i , o servidor adiciona uma nova SAI \mathbf{z} a \mathcal{C} . Se o *slope* atual for β , para gerar uma imagem de um novo ponto focal $\mathbf{v}(\beta, \mathcal{C} \cup \mathbf{z})$ o sistema aplica:

$$\mathbf{v}(\beta, \mathcal{C} \cup \mathbf{z}) = \frac{(\sum_{i=1}^{|\mathcal{S}|} p_i \cdot w_i^\beta \cdot \mathbf{x}_i^\beta) + \mathbf{z}}{(\sum_{i=1}^{|\mathcal{S}|} w_i^\beta) + w_{\mathbf{z}}} \quad (4.8)$$

sendo que cliente recebe e usa essa imagem sintetizada para estimar a \mathbf{z} escolhida, a qual será incorporada à *cache* \mathcal{C} local.

Adicionalmente, convém resumir a codificação adotada tanto pelo PLFC quanto por esta proposta. Assim, para transmitir essas imagens sintetizadas, utiliza-se a vista $\mathbf{v}(\beta, \mathcal{C})$ como preditor para a codificação de $\mathbf{v}(\beta, \mathcal{C} \cup \mathbf{z})$. Quanto à transmissão da *cache* \mathcal{C} inicial, o primeiro *frame* é codificado como um *frame-I*, ao passo que os seguintes são diferencialmente codificados como *frames-P*. Todos os quadros são codificados usando compressão com perdas. Mais detalhes a respeito dessa codificação podem ser encontrados em [17].

4.2 Método Proposto

A proposta deste trabalho, que também será referenciado como ‘PLFC proposto’ [18], consiste em expandir [17], partindo de dois pontos principais:

1. O número de imagens de subabertura que devem estar presentes inicialmente na *cache* \mathcal{C} ;
2. O controle da influência do benefício futuro no processo de seleção de novas imagens.

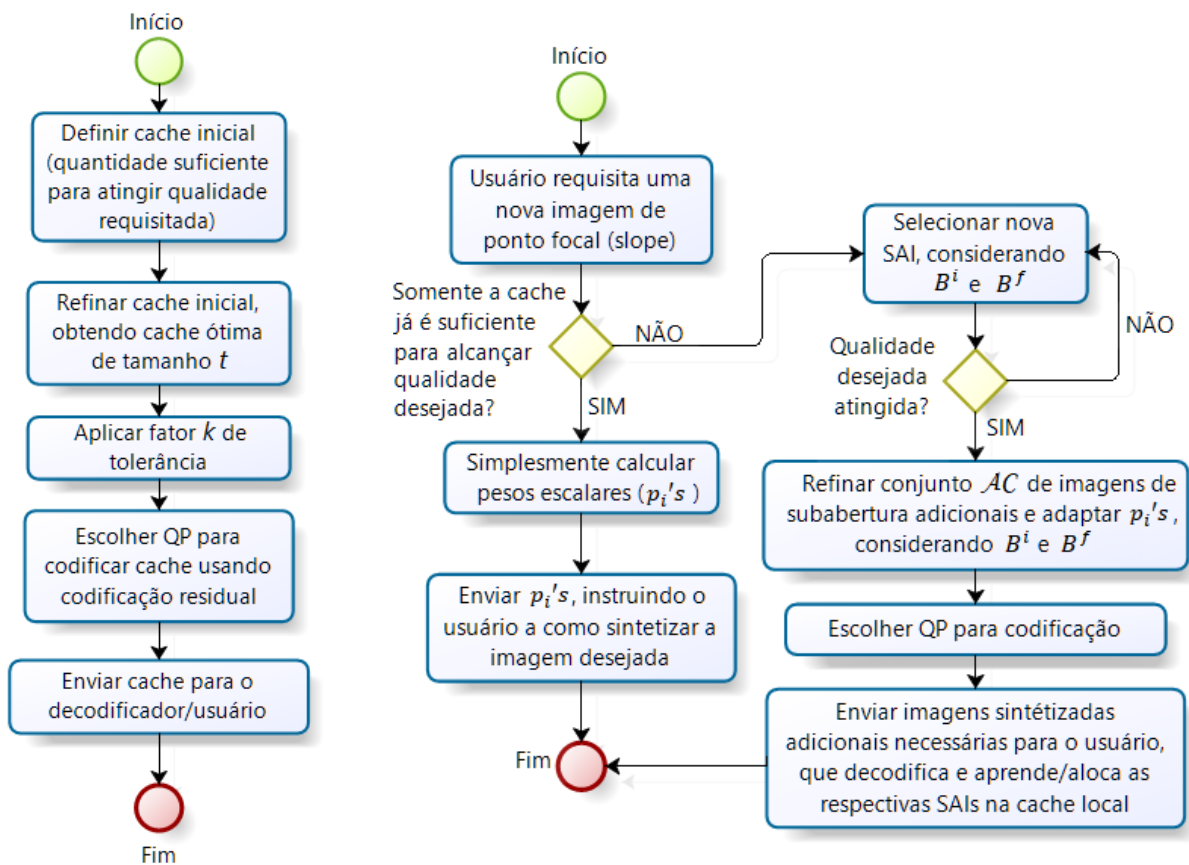
Primeiramente, o tamanho da *cache* inicial é importante pois deve ser escolhido de modo a minimizar o *overhead* de taxa da transmissão inicial. Porém, é preciso que essa quantidade seja suficiente para sintetizar imagens de acordo com a qualidade requerida pelo usuário, sem aumentar consideravelmente a distorção.

Dito isso, o segundo aspecto que merece atenção é o μ da Equação 4.7, que diz respeito ao ganho futuro relacionado a uma determinada imagem de subabertura. No trabalho de

Peixoto et al. [17], o valor usado do coeficiente μ é praticamente irrelevante, prevalecendo o benefício imediato.

Em contrapartida, se muitas imagens sintetizadas forem requisitadas, o benefício futuro associado a transmissão de uma imagem tende a ser maior, o que implica um μ também maior. Além disso, os pontos 1 e 2 não são descorrelacionados, pois é possível notar uma tendência de que sejam transmitidas poucas imagens se a *cache* inicial for maior. Adicionalmente, o impacto do benefício futuro tende a ser mais significativo quando a *cache* é pequena.

No entanto, a metodologia apresentada não se limita apenas a 1 e 2. Aperfeiçoa-se de maneira significativa o processo de seleção de novas imagens como um todo, propondo-se um refinamento dessas SAIs selecionadas para transmissão, no caso em que a *cache* do usuário não seja suficiente para alcançar a qualidade desejada. Uma visão geral desta proposta é apresentada pelos fluxogramas da Figura 4.2, a qual exhibe o processo (a) de inicialização da *cache* e o processo (b) de requisição/sintetização de uma nova imagem de ponto focal. Todos os detalhes desses processos encontram-se a seguir.



(a) Inicialização da cache

(b) Sintetização de um novo ponto focal

Figura 4.2: Fluxogramas que estabelecem uma visão geral do PLFC proposto por este trabalho. (a) Processo de inicialização da *cache*. (b) Processo de sintetização de um novo ponto focal, sendo necessária ou não a transmissão de uma ou mais imagens adicionais.

4.2.1 Inicialização da *cache*

Esta proposta de definição da *cache* inicial corresponde ao passo A de [18] e difere do PLFC original, o qual utilizava uma *cache* de tamanho fixo no início das transmissões que era codificada usando-se um QP igualmente fixo.

Sucintamente, o QP está inversamente ligado à qualidade, isto é, quanto maior o QP, maior o tamanho do bloco a ser codificado e consequentemente maior seja a compressão, o que implica uma qualidade menor; ao passo que um QP menor indica uma qualidade maior. A situação ideal em termos de compressão se dá quando se obtém a melhor qualidade, usando a menor taxa possível [49].

Dito isso, nesta proposta assume-se que o usuário estipula uma qualidade desejada antes mesmo de solicitar a primeira imagem de ponto focal. Posteriormente, o servidor seleciona uma *cache* ótima (ou refinada) de tamanho t , ou seja, o conjunto que contém o número mínimo de imagens necessárias com intuito de atingir a qualidade desejada para o *slope* 0 (esse *slope* é a referência para o início da transmissão, mesmo que esse *slope* nunca seja solicitado pelo usuário). Essa *cache* refinada \mathcal{C} é obtida tal que:

$$|\mathcal{C}| = t \times k \quad (4.9)$$

sendo k um fator de tolerância. Se, por exemplo, houver a necessidade de complementar a *cache* refinada com 50% da sua quantidade inicialmente estipulada, então $k = 1,5$.

Uma vez que \mathcal{C} esteja pronta, o servidor precisa escolher também um QP ótimo para codificar e enviar esse conjunto para o usuário. No sistema implementado, todas as imagens são enviadas usando-se codificação residual. É pertinente ressaltar que *cache* inicial é transmitida antes de qualquer outra imagem e sua qualidade é muito importante para o restante da transmissão, portanto, selecionar o QP apropriado é fundamental.

Suponha que \mathcal{Q} seja o conjunto de todos os valores disponíveis de QP, sendo QP_i um valor único QP em \mathcal{Q} , e $1 \leq i \leq |\mathcal{Q}|$. Considere também uma função genérica $QLD(.)$ que represente uma métrica objetiva da qualidade de um sinal. Por conseguinte, tem-se que $QLD(\mathcal{I}, QP_i)$ representa a qualidade de uma imagem sintetizada a partir de um conjunto de imagens \mathcal{I} , utilizando um QP_i . Assumindo L_{QLD} como o limiar de qualidade mínima requisitada pelo o usuário, o servidor precisa encontrar um índice idx :

$$idx = \arg \max_i \{QP_i : QLD(\mathcal{C}, QP_i) > L_{QLD}\} \quad (4.10)$$

para auferir um QP ótimo:

$$QP_{cache} = QP_{idx} - \epsilon \quad (4.11)$$

sendo que o objetivo é encontrar o maior QP o qual atinge a qualidade alvo. O QP final é ajustado por ϵ , fornecendo alguma tolerância devido a variações de qualidade.

Esse ϵ é obtido a partir de um parâmetro x que controla a frequência de mudança desse valor:

$$x = \frac{L_{QLD} \times \log_{10}(L_{QLD})}{a_1} \quad (4.12)$$

$$\epsilon = \left\lfloor \frac{x\sqrt{x}}{a_2} \right\rfloor \quad (4.13)$$

sendo a_1 e a_2 constantes. Quanto maiores forem essas constantes, mais tarde acontece a mudança de valor do QP adicional ϵ . Percebe-se que a_1 controla a influência do logaritmo e que a_2 controla a influência da raiz quadrada (a qual cresce relativamente mais depressa para valores suficientemente grandes).

Substituindo o x da Equação (4.13) pela Equação (4.12), o ϵ final é descrito por:

$$\epsilon = \lfloor c_1 \times \sqrt{[L_{QLD} \times \log_{10}(L_{QLD})]^{c_2}} \rfloor \quad (4.14)$$

onde c_1 e c_2 são constantes que controlam o crescimento de ϵ com relação à qualidade desejada. As funções logarítmica e raiz quadrada foram usadas para balancear o crescimento do ϵ , modificando esse valor mais rapidamente conforme a qualidade alvo aumenta. Assim, havia a necessidade de se obter uma função resultante que fosse monotonicamente crescente assintótica com crescimento não exponencial. Foi feita, portanto, uma análise empírica dos dados, a qual culminou na definição de uma curva de compensação utilizando-se uma espécie de *data fitting* (técnicas de ajuste). A partir disso, constatou-se que a raiz quadrada do logaritmo foi a que mais se aproximou do resultado desejado.

A Tabela 4.1 exemplifica o funcionamento desse ϵ , mostrando o mapeamento feito entre a qualidade alvo L_{QLD} e o parâmetro x , de modo a obter o fator ϵ de tolerância. A ideia é que quanto maior o x , maior será a frequência de mudança.

Tabela 4.1: Exemplo do mapeamento feito entre a qualidade alvo L_{QLD} e o parâmetro x para obter o valor de ϵ . Os dados da primeira coluna estão quantificados em decibéis em termos da relação sinal ruído, ou seja, quanto maior o valor, maior a qualidade. Os colchetes indicam que o intervalo é fechado e os parênteses indicam que o intervalo é aberto. Conforme a qualidade desejada aumenta, o fator de tolerância (ϵ) muda mais cedo. As constantes usadas nesse exemplo foram $a_1 = 5.81$ e $a_2 = 32$.

| L_{QLD} | x | $\lfloor \epsilon \rfloor$ |
|-----------|---------|----------------------------|
| [10,37] | [1,10) | 0 |
| [38,53] | [10,16) | 1 |
| [54,66] | [16,21) | 2 |
| [67,77] | [21,25) | 3 |

Caches múltiplas

Com o intuito de evitar a propagação de erros de arredondamento, foram criadas múltiplas *caches*, copiadas a partir da *cache* já existente no decodificador, sem a necessidade de se transmitir mais dados. Trata-se de 4 versões alternativas das imagens de subabertura: a primeira contém as SAIs não deslocadas, na segunda estão as SAIs sempre deslocadas para o *slope* mais recente, a terceira tem as SAIs deslocadas para um *slope* negativo e a quarta salva a versão deslocada das SAIs adicionadas para o *slope* correspondente. A ideia é usar a *cache* de menor deslocamento possível à cada novo *slope*.

4.2.2 Algoritmo de Refinamento para seleção de SAIs

O PLFC anterior [17] selecionava cada SAI usando basicamente um algoritmo guloso, selecionando uma imagem por vez com base na Equação (4.7). Este trabalho propõe uma abordagem intermediária entre um algoritmo guloso e uma solução de programação dinâmica, dizendo respeito ao passo B de [18]). O objetivo disso é selecionar não só as melhores imagens para a sintetização, mas também tentar refinar essas imagens com o intuito de encontrar os valores ideais p_i da Equação (4.3). Esse passo é aplicado depois de o servidor selecionar uma imagem adicional (4.1.2) e antes de o decodificador aprender a SAI (4.1.3), toda vez que uma nova SAI for selecionada.

O Algoritmo 1 possibilita um refinamento do processo de seleção das SAIs e é usado para verificar se as imagens selecionadas configuram realmente o melhor conjunto de imagens. Como um exemplo, assumamos que, para um novo *slope* particular β , mais de uma imagem precise ser enviada para que se atinja a qualidade pedida pelo usuário. Considere que \mathcal{AC} (abreviação do inglês *Additional Cache*) representa esse conjunto de novas imagens que serão transmitidas para o receptor, no qual cada imagem pode ser denotada por \mathbf{z}_i , de modo que $\mathbf{z}_i \in \mathcal{AC}$ e $1 \leq i \leq |\mathcal{AC}|$.

Suponha que o receptor possua a *cache* \mathcal{C} atualmente. Logo, quando o codificador selecionou \mathbf{z}_1 , ele levou em consideração apenas que o usuário tinha acesso a \mathcal{C} . Apesar disso, se o servidor souber que mais imagens serão transmitidas, ou seja, outras \mathbf{z}_i com $i > 1$, então a seleção original para \mathbf{z}_1 não necessariamente será a melhor. Isso ocorre porque as imagens subsequentes são comparadas com um conjunto cada vez maior de imagens de subabertura.

Algoritmo 1: Refinar as SAIs adicionais escolhidas para transmissão.

Dados de Entrada: (i) A *cache* atual \mathcal{C} antes de o *slope* β ser requisitado e (ii) O conjunto \mathcal{AC} de imagens adicionais necessárias para sintetizar a imagem de ponto focal \mathbf{v}^β para o *slope* β .

```

1  $\mathcal{RC} = \emptyset$            #  $\mathcal{RC}$  é o conjunto das imagens que podem ser removidas
2 while  $\mathcal{AC} \neq \mathcal{RC}$  do
3    $\mathcal{RC} = \mathcal{AC}$ 
4   for  $i=1:|\mathcal{AC}|$  do
5      $\mathcal{AC} = \mathcal{AC} \setminus \{\mathbf{z}_i\}$       # Remova a i-ésima SAI  $\in \mathcal{AC}$ 
6     Use  $\mathcal{C} \cup \mathcal{AC}$  para sintetizar uma imagem  $\mathbf{v}^\beta$ 
7     if Qualidade de ( $\mathbf{v}^\beta$ ) > Qualidade Alvo +  $\omega$  then
8       break
9       # Saia do loop for e refine o conjunto menor  $\mathcal{AC}$ , pois agora
10       $|\mathcal{AC}| = |\mathcal{AC}| - 1$ 
11    else
12       $\mathbf{z}_i = \mathbf{z}'_i$            # Selecione  $\mathbf{z}'_i$  usando a Equação (4.7)
13       $\mathcal{AC} = \mathcal{AC} \cup \{\mathbf{z}_i\}$  # Insira essa SAI em  $\mathcal{AC}$ 
14    end
15  end
16 return  $\mathcal{C} = \mathcal{C} \cup \mathcal{AC}$       # cache completamente refinada

```

O algoritmo proposto removerá inicialmente uma imagem por vez de \mathcal{AC} e selecionará outra imagem (que inclusive pode ser a que acabou de ser retirada). Essa seleção considera todas as imagens previamente selecionadas bem como aquelas que ainda serão escolhidas, isto é, o benefício futuro da Equação (4.5) também é levado em conta a cada uma dessas ‘substituições’. Essa operação é repetida até que o conjunto completo \mathcal{AC} não se modifique, ou seja, quando nada mais pode ser refinado.

Nota-se que, toda vez que uma imagem \mathbf{z}_i é substituída por uma outra imagem \mathbf{z}'_i , os valores p_i da combinação linear da Equação (4.3) precisam ser recalculados. As imagens são enviadas uma por uma para o decodificador depois que esse algoritmo se encerra.

É válido mencionar que esse refinamento pode reduzir o número de imagens transmitidas, como mostra a linha 8 do Algoritmo 1. Portanto, não há a necessidade de se selecionar uma SAI adicional se a condição de qualidade já for satisfeita com um conjunto menor (sendo que ω é uma pequena constante que garante alguma tolerância). Este algoritmo de refinamento para seleção das SAIs também é aplicado durante a inicialização da *cache* (Subseção 4.2.1) para obter a já referida *cache* ótima de tamanho t e, por conseguinte, ela também é chamada de *cache* refinada.

4.2.3 Seleção dinâmica de QP / contribuição de qualidade

No PLFC [17], um QP fixo empiricamente escolhido era usado para codificar todas as imagens transmitidas para cada qualidade alvo. Portanto, em alguns casos, a L_{QLD} podia nunca ser alcançada, mesmo que o conjunto completo \mathcal{S}^o de SAIs fosse enviado. Para evitar esse problema, o Algoritmo 2 estipula uma contribuição mínima de qualidade requerida para a seleção do QP antes da codificação residual da imagem sintetizada, o que ocorre depois de o servidor selecionar uma SAI adicional (4.1.2). Essa etapa corresponde ao passo C de [18].

Algoritmo 2: Selecionar dinamicamente o QP, levando em consideração uma contribuição mínima de qualidade estabelecida.

```

1  $QP\_atual = QP\_MaximoEscolhido + 1$ 
2 while  $contribuicao < contribuicaoMinima$  do
3   |  $QP\_atual = QP\_atual - 1$ 
4   | Usando o  $QP\_atual$ , simule a codificação da imagem sintetizada
5 end
6 return  $O\ QP$  para codificar a imagem sintetizada

```

4.2.4 Eliminação da combinação linear esparsa para o cálculo do benefício futuro

Assuma que B^i seja a abreviação da notação $B_v^i(\mathcal{C}, z)$ na Equação (4.4) e que B^f seja a versão breve de $B^f(z, \mathcal{C}, \mathcal{V})$ na Equação (4.5). Visando selecionar mais SAIs para usar na Equação (4.3), o PLFC anterior propõe uma combinação linear esparsa [17] para obter a SAI que forneça o maior aumento do benefício imediato (B^i), ou seja, aquela que possibilita uma aproximação maior da qualidade alvo no instante da solicitação feita pelo

cliente. Isso é feito para todas as imagens de subabertura restantes, as que ainda não estão na *cache* do usuário obviamente.

Assim, este trabalho propõe uma ligeira modificação desse processo (Subseção 4.1.2 - Cálculo do B^f). O servidor seleciona a melhor imagem para os *slopes* futuros utilizando não uma combinação linear esparsa, mas ao invés disso ele atribui o valor 1 a todos os coeficientes p_i da Equação (4.3). O cômputo do p_i para cada uma das SAIs, com o intuito de obter seu respectivo B^f requer um grande esforço computacional.

Nota-se que a remoção desse cálculo pode alterar a SAI que será escolhida, mas a carga computacional foi significativamente reduzida. Por mais que esse ganho tenha sido verificado na prática no momento da realização das simulações, convém dizer que não foi feito nenhum teste formal de redução da complexidade e/ou tempo de execução.

É preciso ressaltar que, uma vez escolhida uma determinada SAI para o envio, a combinação linear esparsa é aplicada como de costume, antes do envio propriamente dito.

4.2.5 Cômputo automático do multiplicador de Lagrange, μ

Um grande desafio do processo de seleção em 4.1.2 é encontrar o μ ótimo na Equação (4.7) o qual pondera os benefícios imediato e futuro. No PLFC original, isso era também determinado empiricamente. Entretanto, este trabalho adota:

$$\mu' = \frac{\max(B^i)}{\max(B^f)} \times \frac{\min(B^i)}{\min(B^f)} \quad (4.15)$$

onde $\min(B^{indice})$ é a mínima entrada não nula de B^{indice} , para $indice = i, f$. A primeira parte da Equação (4.15) funciona como um fator de normalização entre os dois custos e a segunda parte garante que B^f seja significativo na seleção das SAIs.

Adicionalmente, o μ final proposto é ajustado como:

$$\mu = \frac{\mu'}{((\lfloor L_{QLD}/d_1 \rfloor * d_1) * d_2)} \quad (4.16)$$

onde d_1 e d_2 são constantes que afetam a dependência do μ em termos de L_{QLD} , a qualidade alvo. A ideia é que, à medida que essa qualidade cresce, o μ deve diminuir. Isso porque, se o benefício futuro tiver uma influência grande demais no processo de seleção para altas qualidades alvo, a qualidade imediata requisitada pode nunca ser atingida. Essa maneira de estipular o multiplicador de Lagrange corresponde ao passo D de [18].

Capítulo 5

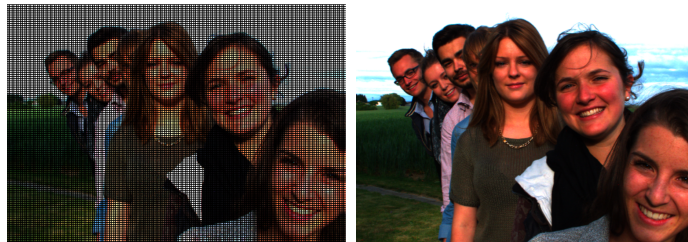
Resultados

Inicialmente este capítulo mostra as imagens *light field* que foram utilizadas para a validação do método proposto, bem como apresenta as configurações gerais do sistema. Em seguida constam os resultados obtidos pelo método proposto, referentes à taxa acumulada e à relação entre taxa e distorção ao final da transmissão. É feita também uma discussão acerca dos ganhos obtidos, comparando-os com os resultados do PLFC anterior [17].

5.1 Banco de imagens



(a) Friends 1



(b) Friends 3

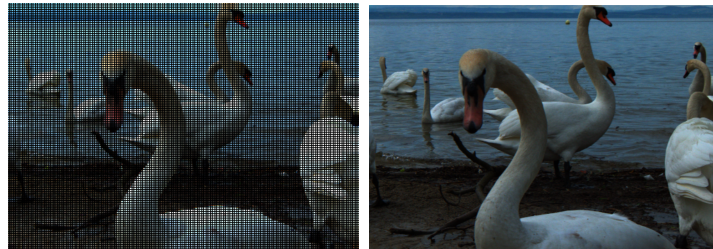
Figura 5.1: À esquerda: LF completo. À direita: Uma das SAIs obtidas.

As imagens usadas foram capturadas com uma câmera Lytro Illum, Figura 2.4 (b), sendo que 225 imagens de subabertura foram adquiridas para cada LF. Todos os 11 LFs (10 exibidos neste capítulo, além de Sophie & Vincent já apresentado no Capítulo 2) são do banco de imagens [67] e têm uma profundidade de 10 *bits* por *pixel*.

Essas SAIs foram quantizadas para 8 *bits* (ao invés dos 14 *bits* do PLFC anterior) e então 2 *pixels* foram removidos das bordas, por conta do ruído de aquisição. Objetivando um melhor deslocamento das SAIs, foi feito um *padding* de 10 *pixels* em cada direção.



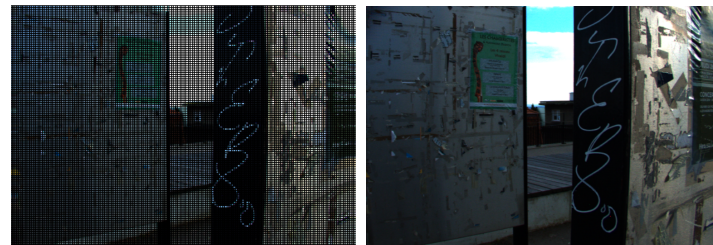
(a) Sophie Krios & Vincent



(b) Swans 1

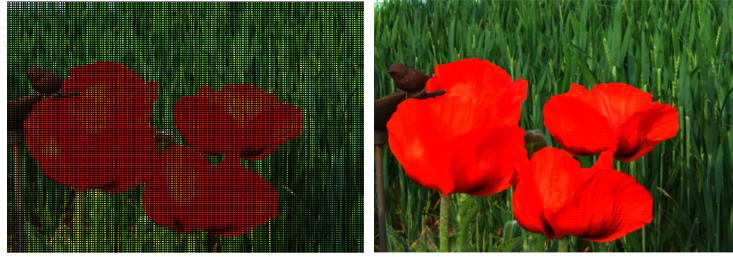


(c) Chain-link Fence 1

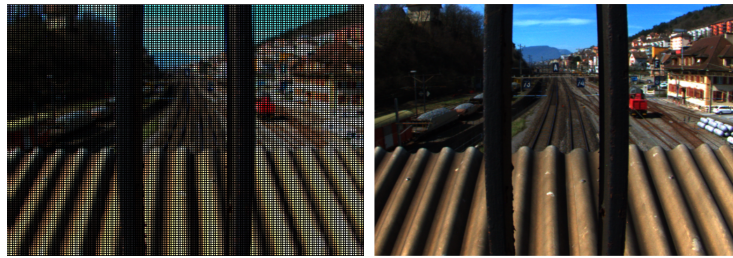


(d) Billboards

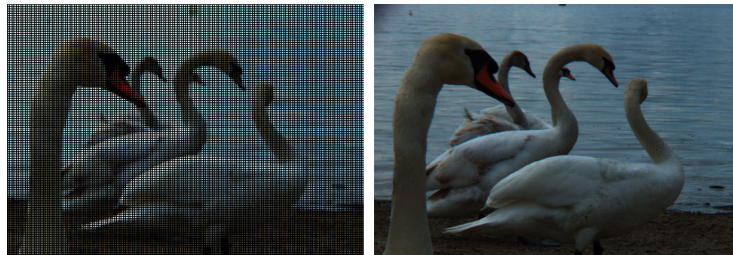
Figura 5.2: Lado esquerdo: LF completo, demonstrando os *macropixels* capturados. Lado direito: Ilustração de uma das 225 imagens de subabertura obtidas.



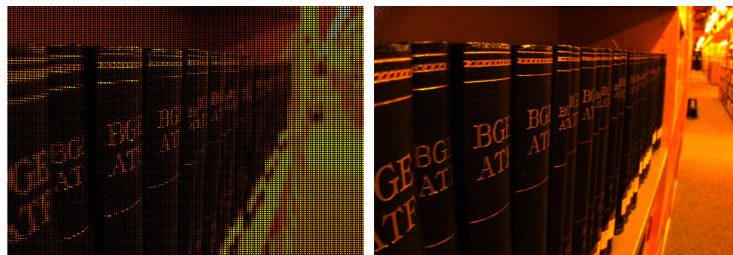
(a) Poppies



(b) Railway Lines 1



(c) Swans 2



(d) Books

Figura 5.3: Lado esquerdo: *raw* LF com resolução completa de 9615×6750 , sendo possível verificar os *macropixels* capturados. Lado Direito: Ilustração de uma das 225 SAIs obtidas.

5.2 Métrica de avaliação de resultados

Este trabalho utiliza a PSNR (*Peak-Signal-to-Noise-Ratio*) da luminosidade como critério objetivo para mensurar a qualidade das imagens transmitidas. Definindo R como o maior valor que um *pixel* de uma determinada imagem possa assumir e supondo que se deseje comparar duas imagens A_1 e A_2 , ambas de resolução $I \times J$, pode-se calcular [68]:

$$PSNR = 10 \log_{10} \left(\frac{R^2}{MSE} \right) \quad (5.1)$$

sendo o erro médio quadrático, do inglês *mean squared error* (MSE), definido como:

$$MSE = \sum_{i=0}^{I-1} \sum_{j=0}^{J-1} [A_1(i, j) - A_2(i, j)]^2 \quad (5.2)$$

Dito isso, é importante esclarecer que o objetivo deste trabalho é atingir a melhor qualidade, utilizando a menor taxa possível. Assim, quanto maior for a PSNR alcançada e menor for a quantidade de *bits* transmitida, melhor será o resultado.

5.3 Configurações dos testes

Os resultados obtidos foram comparados com um método baseado em ILFS [20–24], chamado de ILFS de Referência, no qual cada novo ponto focal sintetizado é codificado como um *frame-P* (*P-slice*), usando as imagens previamente transmitidas como predição. Para uma comparação adequada, o codificador usado neste trabalho foi o mesmo H.264/AVC encoder JM 19.0 [53] usado no ILFS de referência e também no PLFC anterior, operando no modo RGB.

Dessa maneira a codificação usada neste trabalho, assim como a utilizada pelo PLFC, é do tipo DPCM (Modulação por Codificação Diferencial de Pulso, do inglês *Differential Pulse-Code Modulation*), valendo destacar que foram usados macroblocos apenas do tipo $P16 \times 16$ e *PSKIP*, com todos os vetores de movimento iguais a zero. Para o ILFS de referência, todos os modos de codificação de macroblocos foram testados, além de ter sido feita uma estimação de movimento completa.

A mudança de *slope*, isto é, o próximo ponto focal requisitado é gerado aleatoriamente por uma distribuição gaussiana, definida pela função densidade de probabilidade [69]:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2} \left(\frac{x - m}{\sigma} \right)^2 \right) \quad (5.3)$$

na qual m representa a média e σ o desvio padrão. Para os testes deste trabalho, 100 possíveis *slopes* foram gerados, usando-se $\sigma = 0, 3$.

Como já dito, a taxa foi medida em *bits* e a qualidade foi medida em termos da PSNR entre a imagem sintetizada para o *slope* em questão com todas as SAIs disponíveis no codificador e a imagem sintetizada somente com as SAIs da *cache* do decodificador.

Sabendo-se que os pesos w_i , Equações (4.8), (4.1) e (4.3), são intrínsecos à câmera, é razoável considerar que a mesma câmera é usada para diversas transmissões. Assim, assume-se que tais pesos são conhecidos tanto pelo servidor quanto pelo usuário. A

respeito da inicialização da *cache* (Subseção 4.2.1) $k = 1,5$, $c_1 = 1/450$ e $c_2 = 3$. Para a Subseção 4.2.3 (seleção dinâmica do QP), o $QP_MaximoEscolhido$ é igual a 18 e $contribuicaoMinima = 0,5$ dB, garantindo que SAIs de boa qualidade sejam adicionadas na *cache* \mathcal{C} . Com relação ao cômputo do μ (Subseção 4.2.5), tem-se $\eta = 10$ e $\xi = 10/35$.

Convém ressaltar que a ferramenta computacional MATLAB [70] foi utilizada para a obtenção dos resultados deste trabalho. Os testes foram feitos em um computador com sistema operacional *Windows 10 Home Single 64 bits* e o processador usado foi um *Intel Core i7-6700HQ CPU* (Unidade Central de Processamento, do inglês *Central Processing Unit*) 2,60 *Gigahertz* com memória de 16,00 *Gigabytes*.

É importante mencionar que também foram feitas algumas otimizações com relação à execução. Antes, para obter o B^f da Equação (4.5), o sistema precisava deslocar todas as SAIs em \mathcal{C} para um número fixo de *slopes* discretos, com o intuito de calcular a importância de cada imagem com relação às sintetizações futuras. Esses deslocamentos são computacionalmente caros e eram feitos todas as vezes que uma nova SAI era considerada para envio (dentre as 225 no total, exceto aquelas que já estavam na *cache*). No entanto, agora o servidor faz isso apenas uma vez e armazena os dados em memória para reusar.

Ademais, usou-se a programação paralela disponível pelo próprio MATLAB, objetivando reduzir o tempo de cômputo desse benefício futuro (B^f), que é também muito custoso. Com essa finalidade, inicialmente são usados 4 núcleos (*workers*) para o processamento, mas essa quantidade depende do tanto de memória disponível. Desse modo, pode haver uma diminuição do número de núcleos conforme a execução do sistema avance. Com essas mudanças, tempos de execução total que chegavam a ser maiores que uma semana, agora não ultrapassam 2 dias.

5.4 Resultados Obtidos

Cada um dos resultados a seguir é referente a uma execução do sistema na qual o cliente inicialmente escolhe um LF e define uma PSNR mínima (baixa: 36 dB, média: 38 dB ou alta: 40 dB) para os pontos focais (*slopes*) que serão requisitados por ele. Posteriormente, o servidor sintetiza a imagem correspondente, envia os coeficientes da Equação (4.3) e, se necessário, transmite mais SAIs. Como o sistema recebe uma restrição de qualidade mínima a ser atingida, pode ser que a PSNR alcançada de fato seja maior que esse limite.

Os resultados obtidos essenciais para o entendimento deste capítulo estão listados abaixo (Swans 2, Sophie, Krios & Vincent e Sophie & Vincent 1). Entretanto, todos os demais resultados para os LFs Friends 1, Friends 3, Swans 1, Chain-link Fence 1, Billboards, Poopies, Railway Lines 1 e Books constam no Apêndice A.

5.4.1 Taxa acumulada

Esta subseção exibe os resultados referentes à taxa acumulada transmitida desde o início da execução do sistema, incluindo os *bits* da *cache* inicial. Esses resultados mostram como a taxa cresce à medida que mais *slopes* são pedidos pelo usuário.

Exemplos nos quais as taxas acumuladas finais desta proposta foram menores que as taxas obtidas pelo PLFC anterior [17] podem ser vistos nas Figuras 5.4, 5.5 e 5.6, que dizem respeito ao LF Swans 2, e nas Figuras 5.7, 5.8 e 5.9, que são do LF Sophie & Vincent 1.

As Figuras 5.10, 5.11 e 5.12 referem-se ao LF Sophie Krios & Vincent, sendo que o resultado correspondente à qualidade mínima exigida de 36 dB (Figura 5.10) é um exemplo em que o PLFC proposto apresentou uma taxa acumulada final maior que [17].

Além desse caso, apenas uma outra vez o resultado do método proposto foi desfavorável em termos de taxa transmitida com relação a [17]: LF Friends 3, também com PSNR mínima de 36 dB, que pode ser consultado no Apêndice A (Figura A.4). Mesmo assim, as taxas acumuladas ainda foram menores que as obtidas pela abordagem ILFS, cabendo destacar que a PSNR atingida supera inclusive a do PLFC anterior em ambos os casos.

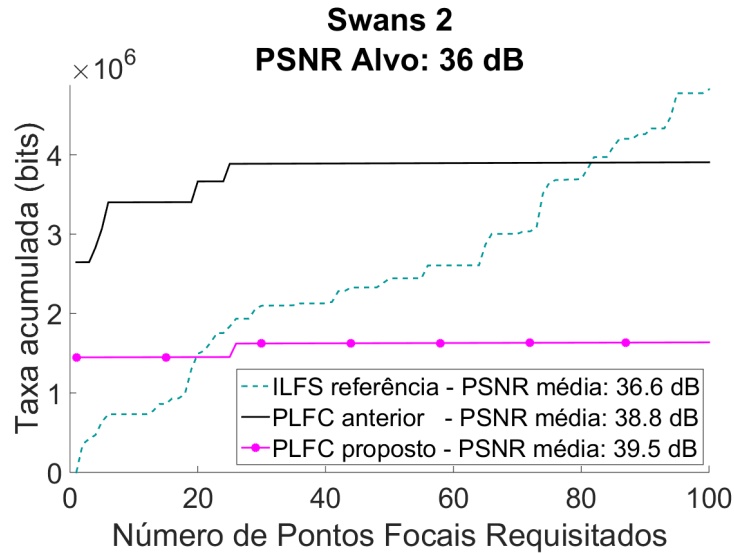


Figura 5.4: Resultado da transmissão de pontos focais sintetizados referentes ao *light field* Swans 2, com PSNR alvo de 36 dB. Apresenta-se os resultados das transmissões triviais ILFS, do PLFC original [17] e do PLFC proposto por este trabalho.

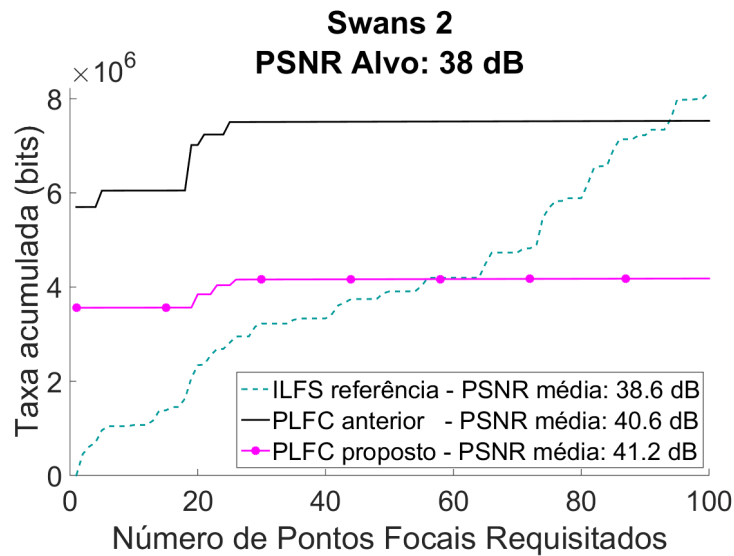


Figura 5.5: Resultado da transmissão de pontos focais sintetizados referentes ao LF Swans 2, PSNR alvo de 38 dB. Compara-se os resultados do ILFS, do PLFC [17] e do novo PLFC.

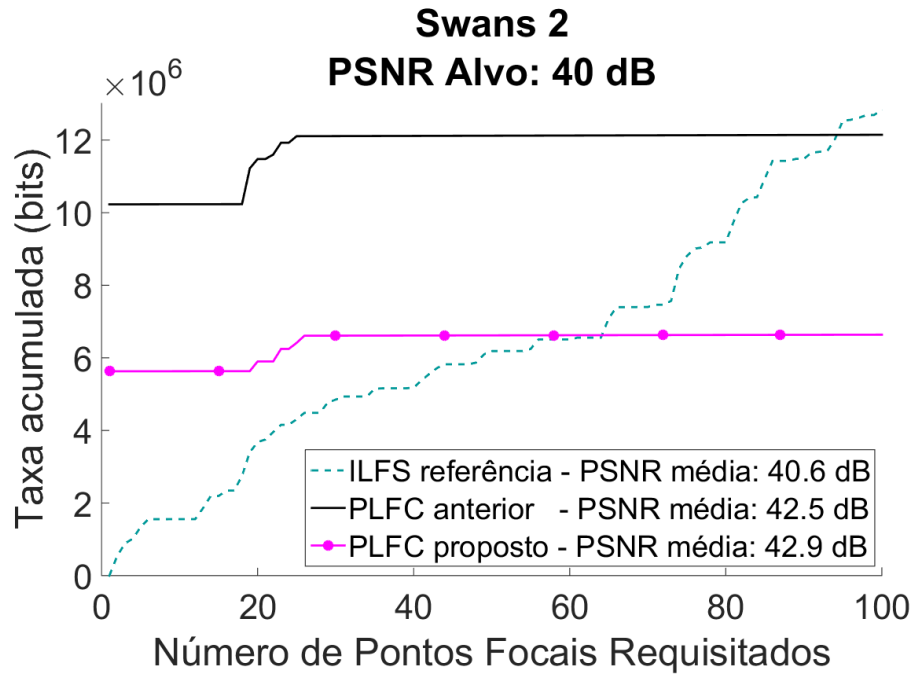


Figura 5.6: Resultado da transmissão de pontos focais sintetizados referentes ao LF Swans 2, PSNR alvo de 40 dB. Nota-se os resultados do ILFS, do PLFC [17] e do PLFC proposto.

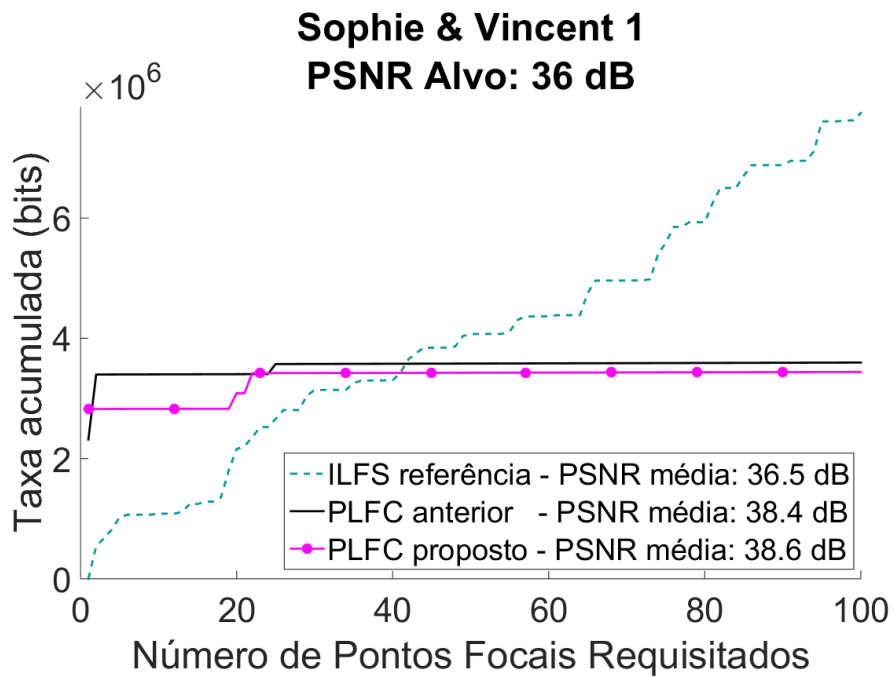


Figura 5.7: Resultado da transmissão de 100 *slopes* para PLFC [17], ILFS e PLFC proposto para o LF Sophie & Vincent 1, alvo 36 dB. A taxa acumulada, em *bits*, encontra-se no eixo vertical e o número de cada imagem requisitada (de 1 a 100) está no eixo horizontal.

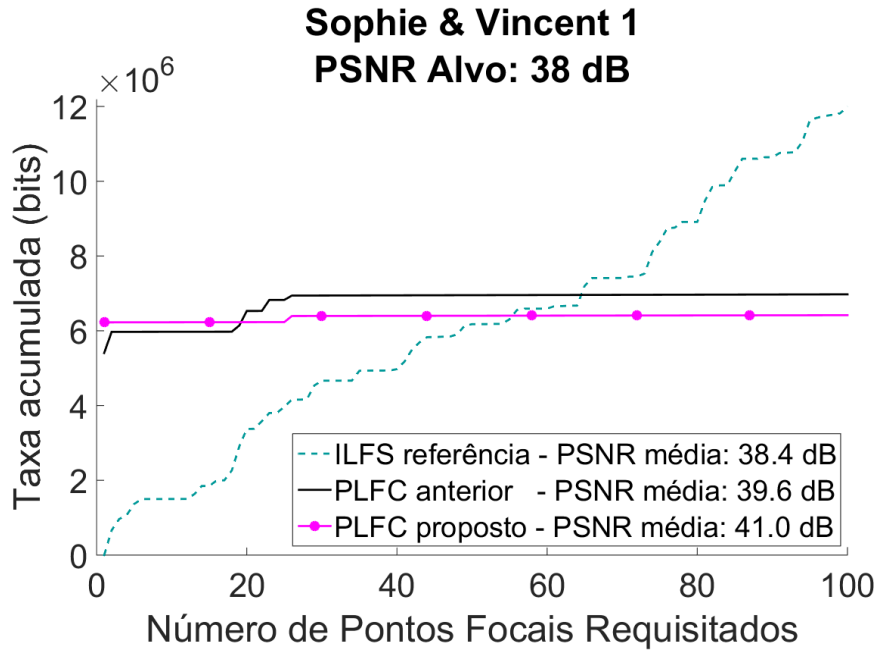


Figura 5.8: Resultado da transmissão de pontos focais sintetizados referentes ao *light field* Sophie & Vincent 1, com PSNR alvo de 38 dB. A taxa acumulada, em *bits*, encontra-se no eixo vertical e o número de cada imagem requisitada (de 1 a 100) está no eixo horizontal. Os resultados do ILFS, do PLFC original e do PLFC proposto são apresentados.

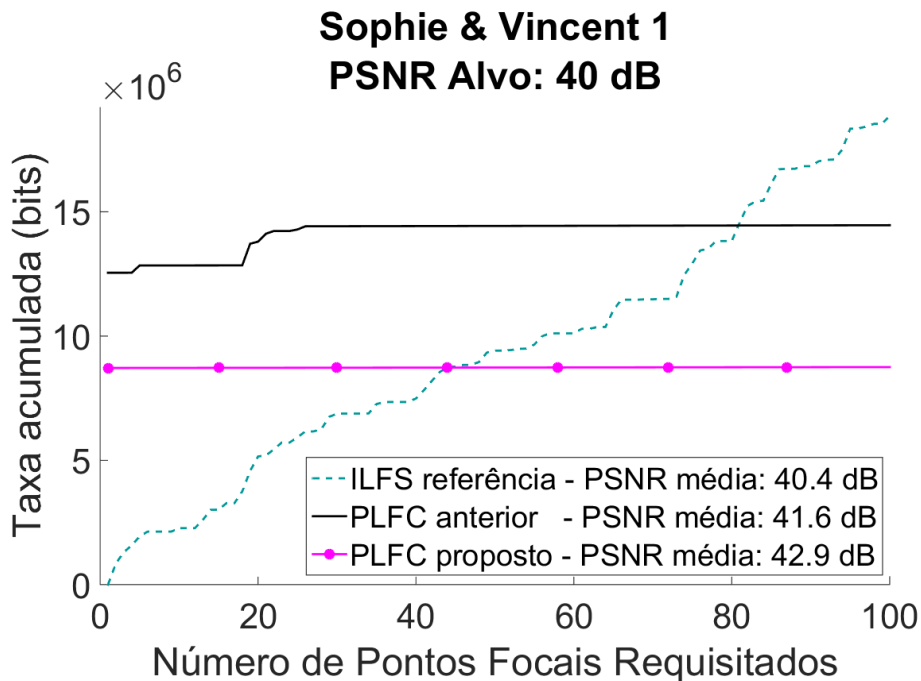


Figura 5.9: Resultado da transmissão de pontos focais sintetizados referentes ao LF Sophie & Vincent 1, PSNR alvo de 40 dB. Eixo vertical: taxa acumulada, em *bits*. Compara-se os resultados do ILFS, do PLFC [17] e do novo PLFC.

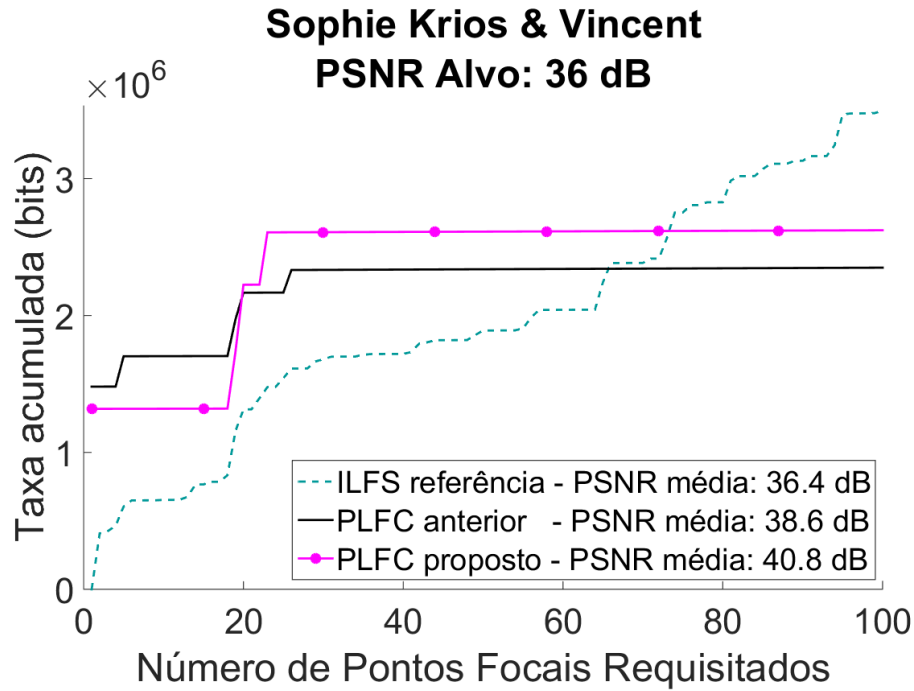


Figura 5.10: Resultado da transmissão de pontos focais sintetizados referentes ao LF Sophie Krios & Vincent, com PSNR alvo de 36 dB. Comparação entre PLFC original, PLFC proposto e ILFS.

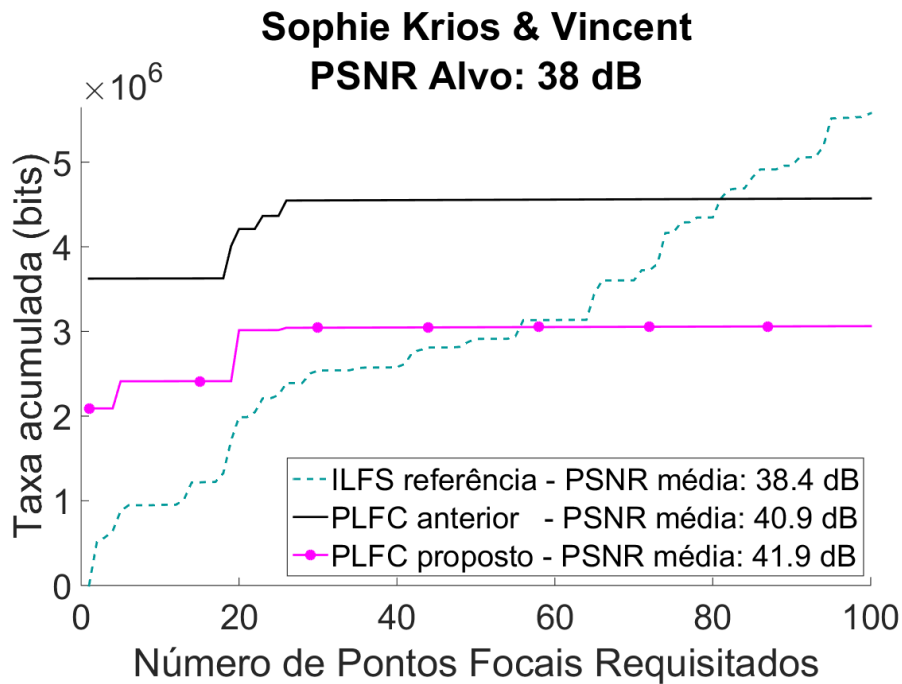


Figura 5.11: Resultado da transmissão de pontos focais sintetizados referentes ao LF Sophie Krios & Vincent, com PSNR alvo de 38 dB. Exibe-se a taxa acumulada conforme o número de imagens requisitadas pelo usuário.

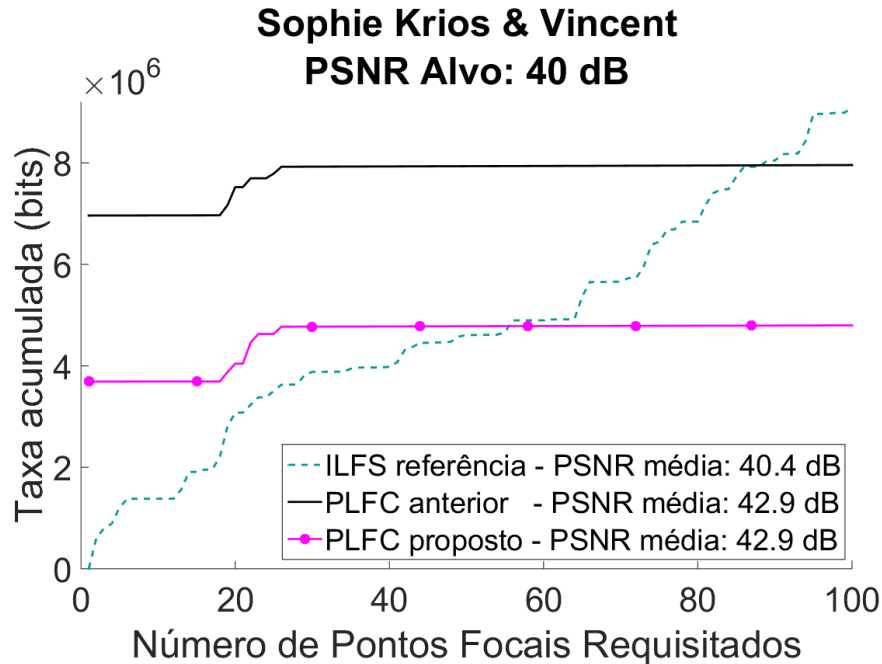


Figura 5.12: Resultado da transmissão de pontos focais sintetizados referentes ao *light field* Sophie Krios & Vincent, com PSNR alvo de 40 dB. A taxa acumulada, em *bits* está representada no eixo das ordenadas e o número do *slope* pedido (de 1 a 100) está no eixo das abcissas. Resultados do ILFS, do PLFC original e do PLFC proposto são apresentados.

5.4.2 Taxa distorção

Para medir a relação entre a quantidade de *bits* transmitidos e a qualidade obtida, foi utilizado o método definido por Bjøntegaard [19], o qual propõe uma maneira de interpolar pontos para medir diferenças de áreas entre duas curvas de taxa-distorção. Tal diferença de área pode ser estabelecida em termos da taxa associada à transmissão (BD-Taxa) ou em termos da medida de qualidade adotada (BD-PSNR, por exemplo). Em particular, este trabalho faz uso da BD-Taxa.

Dito isso, as Figuras 5.13, 5.14 e 5.15 apresentam o resultado taxa distorção para as imagens *light field* Swans 2, Sophie & Vincent 1 e Sophie Krios & Vincent, respectivamente. Nessa figuras, um resultado melhor é aquele que se encontra mais acima e mais à esquerda do gráfico, significando que nesse caso foi possível enviar menos taxa para o usuário, atingindo uma melhor qualidade.

Nota-se que na Figura 5.15, o primeiro ponto destacado da curva que corresponde ao resultado deste trabalho está mais acima com relação ao primeiro ponto da curva do PLFC anterior [17]. Isso significa que uma PSNR maior foi atingida pelo PLFC proposto e, portanto, uma qualidade melhor foi alcançada. Contudo, esse primeiro ponto de [17] está mais à esquerda, indicando que ele gastou uma taxa menor que o PLFC proposto.

Prosseguindo da esquerda para a direita nessas curvas, percebe-se que, conforme a qualidade atingida sobe, a taxa tende a ser menor no PLFC proposto, em comparação com o PLFC anterior, o que pode ser comprovado pela análise conjunta das Figuras 5.10, 5.11 e 5.12. Os resultados taxa distorção dos outros LFs usados estão no Apêndice A.

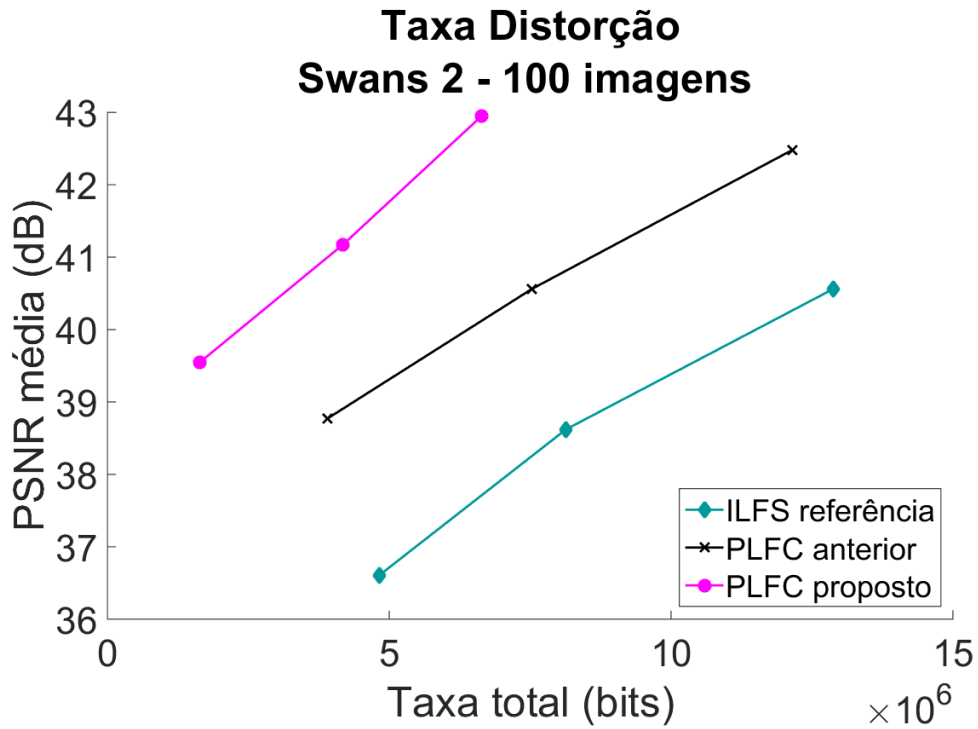


Figura 5.13: Resultado taxa distorção referente ao LF Swans 2 depois de 100 imagens transmitidas. Quanto mais acima e à esquerda uma curva estiver, melhor é o resultado. Exibe-se os resultados do ILFS, do PLFC e do PLFC proposto.

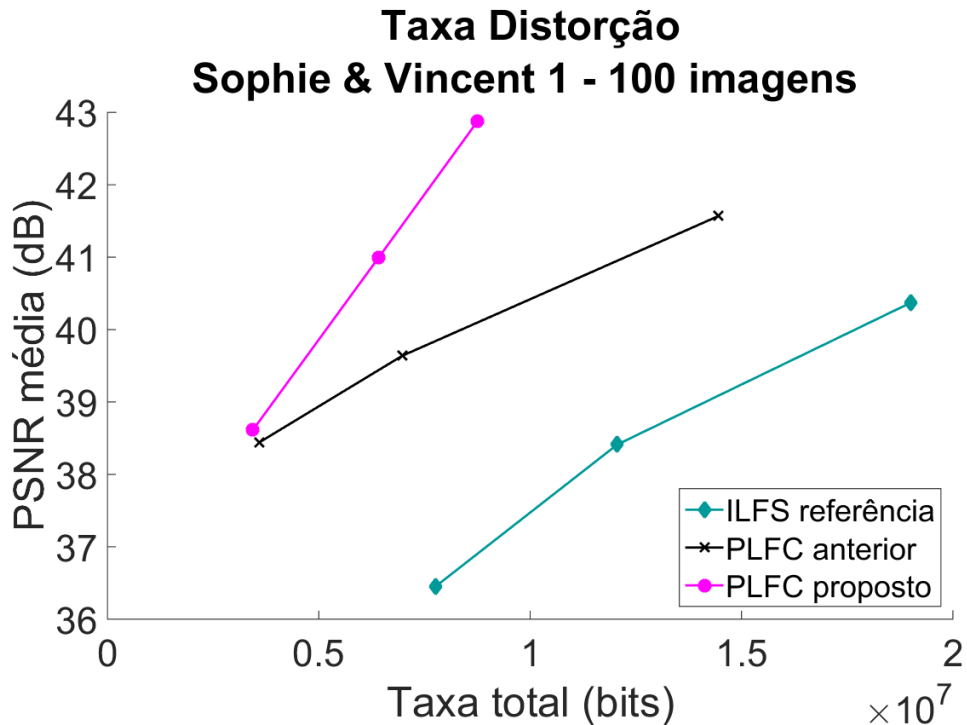


Figura 5.14: Resultado taxa distorção referente ao LF Sophie & Vincent 1.



Figura 5.15: Resultado taxa distorção referente ao LF Sophie Krios & Vincent.

5.5 Discussão de resultados

A Tabela 5.1 compara os resultados obtidos pelo método proposto com os resultados da versão anterior [17]. Ela se refere à BD-Taxa [19], sendo que cada valor exibido representa o quanto de taxa foi preciso transmitir a mais (valores positivos) ou a menos (valores negativos) com relação ao ILFS de referência [20–24] para conseguir uma mesma qualidade. Essa tabela proporciona uma análise progressiva, ou seja, é possível perceber como o sistema reage conforme mais pontos focais vão sendo solicitados. Porém, a última grande coluna é a mais importante, pois diz respeito ao resultado final depois de todos os *slopes* terem sido solicitados.

Assim, a taxa final acumulada pode ser reduzida em até 85,8% com relação à abordagem ILFS [20–24] e em até 32,8% com relação ao PLFC [17]. Em termos de ganho médio, o método proposto consegue uma redução de 75,5% frente ao ILFS e de 18,2% em comparação com o PLFC anterior. Além disso, verifica-se que o ganho final mínimo desta proposta (71,4%) é muito próximo do ganho máximo de [17] (73,7%), se o resultado da transmissão do LF Chain-link Fence 1 for desconsiderado.

Percebe-se que os maiores ganhos são obtidos conforme mais imagens são transmitidas, todavia esses ganhos são ainda mais significativos e acontecem mais cedo com as modificações propostas. Por exemplo, na imagem *Poopies* depois de 20 pontos focais, os resultados deste trabalho já apresentam um ganho de 25,6% sobre a abordagem ILFS, enquanto que no PLFC anterior um ganho de 15,7% aparece apenas depois de 60 imagens requisitadas. Situação semelhante ocorre com *Friends 1* e *Friends 3*.

Tabela 5.1: BD-Taxa [19] depois de um determinado número de imagens transmitidas, usando ILFS [20–24] como base. O resultado do PLFC anterior [17] aparece nas subcolunas da esquerda, ao passo que o resultado desta proposta encontra-se nas subcolunas da direita (prop). Valores positivos referem-se ao quão pior o método em questão é em comparação com o método base, sendo que valores negativos representam os ganhos obtidos em porcentagem.

| Imagem LF | Número de Imagens Transmitidas | | | | | | | | | |
|---------------------------|--------------------------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 20 | | 40 | | 60 | | 80 | | 100 | |
| | [17] | prop | [17] | prop | [17] | prop | [17] | prop | [17] | prop |
| Friends 1 | +20.7 | -23.5 | -9.7 | -46.4 | -37.7 | -64.8 | -58.9 | -76.0 | -70.2 | -82.3 |
| Friends 3 | +4.4 | -29.5 | -18.7 | -47.8 | -43.5 | -67.3 | -63.3 | -77.2 | -73.7 | -83.3 |
| Sophie and Vincent | +70.8 | +11.7 | +18.7 | -21.6 | -18.3 | -45.7 | -45.2 | -62.1 | -59.8 | -71.4 |
| Sophie, Krios and Vincent | +33.3 | +12.2 | +19.0 | -17.8 | -10.1 | -45.4 | -43.9 | -64.0 | -57.7 | -73.1 |
| Swans 1 | +28.6 | -28.9 | +16.8 | -49.9 | -15.8 | -65.6 | -51.6 | -77.5 | -65.6 | -83.2 |
| Chain-link Fence 1 | — | +155.0 | — | +95.0 | — | +32.3 | — | -21.6 | — | -42.9 |
| Billboards | -9.2 | -38.0 | -17.5 | -51.2 | -40.3 | -67.0 | -65.7 | -79.9 | -73.5 | -85.8 |
| Poopies | -32.8 | -25.6 | +12.1 | -34.8 | -15.7 | -50.7 | -50.6 | -68.7 | -66.1 | -78.0 |
| Railway Lines 1 | 54.6 | -10.5 | 8.1 | -30.7 | -23.2 | -54.9 | -54.3 | -70.0 | -66.3 | -77.1 |
| Swans 2 | +147.4 | -13.0 | +86.6 | -31.1 | +7.1 | -59.1 | -22.8 | -71.3 | -47.3 | -80.1 |
| Books | +125.1 | +0.3 | +81.0 | -26.4 | +29.5 | -40.7 | -27.7 | -63.3 | -50.6 | -72.9 |

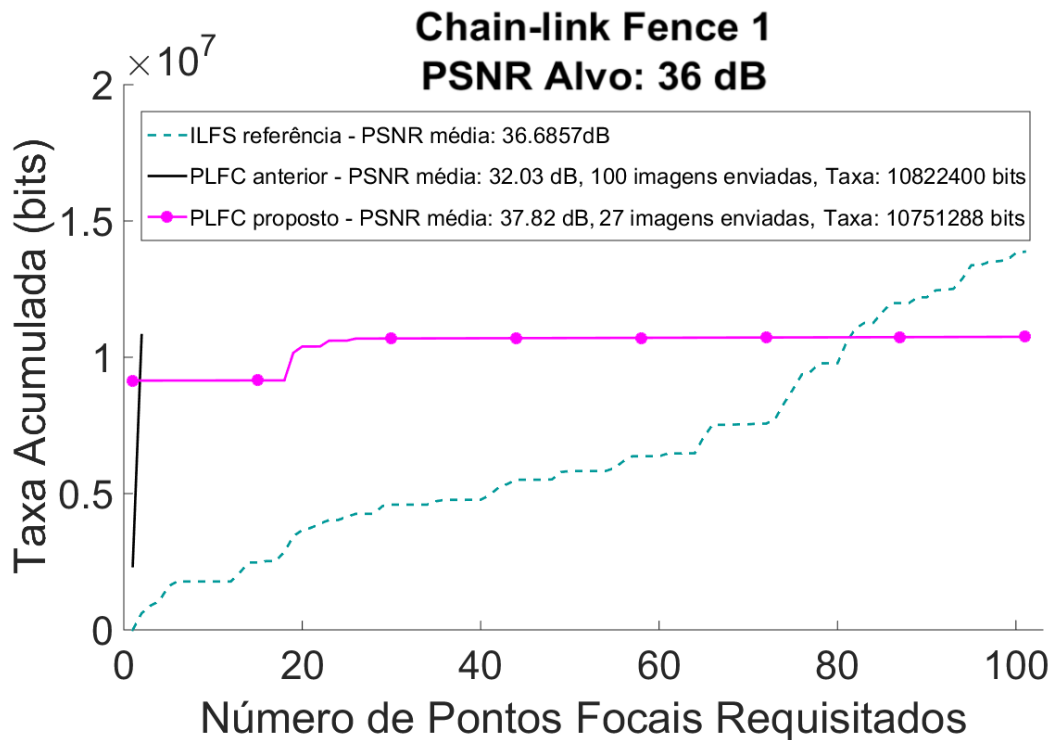


Figura 5.16: Caso particular do PLFC anterior [17]. Nota-se que com 100 imagens enviadas a qualidade mínima não foi atingida. Nessa situação o sistema chegava a mandar todas as 225 SAIs disponíveis no servidor e ainda assim a PSNR mínima não era alcançada. Por outro lado, o PLFC proposto envia menos taxa com relação ao momento retratado de [17] e muito menos imagens, suficientes para todos os 100 slopes requisitados.

Quanto ao resultado referente ao LF Chain-link Fence 1, nota-se que o método proposto apresenta uma perda de 155,0% em comparação com o ILFS após os primeiros 20 *slopes*, haja vista que uma *cache* inicial maior foi estipulada. No entanto, os ganhos aparecem conforme a transmissão avança, já que poucas imagens adicionais precisam ser enviadas para o usuário. Ao final, economizam-se 42,9% dos *bits* transmitidos pelo ILFS.

Ainda a respeito da Tabela 5.1, percebe-se que na linha do LF Chain-link Fence 1 existem algumas entradas não preenchidas (-). Trata-se de um caso particular do PLFC anterior [17], no qual o LF inteiro (ou seja, todas as imagens de subabertura) era transmitido e, mesmo assim, a qualidade requerida pelo usuário não era atingida. Nos resultados deste trabalho, isso não mais acontece, por conta da contribuição estabelecida na Subseção 4.2.3. A Figura 5.16 ilustra essa situação.

É pertinente tratar de um cenário mostrado anteriormente na Figura 5.9, também a respeito da inicialização da *cache*. Essa figura mostra que para o LF Sophie & Vincent 1, mais especificamente quando a PSNR mínima é de 40 dB, apenas a *cache* inicialmente transmitida pelo PLFC proposto já é suficiente para todos os *slopes* solicitados. O detalhe é que, mesmo com uma taxa menor, a qualidade atingida é superior a [17], evidenciando a efetividade do processo de inicialização da *cache* (Subseção 4.2.1) combinado com o refinamento da Subseção 4.2.2. Assim, a escolha de quais devem ser as imagens transmitidas é tão importante quanto a definição da quantidade inicial.

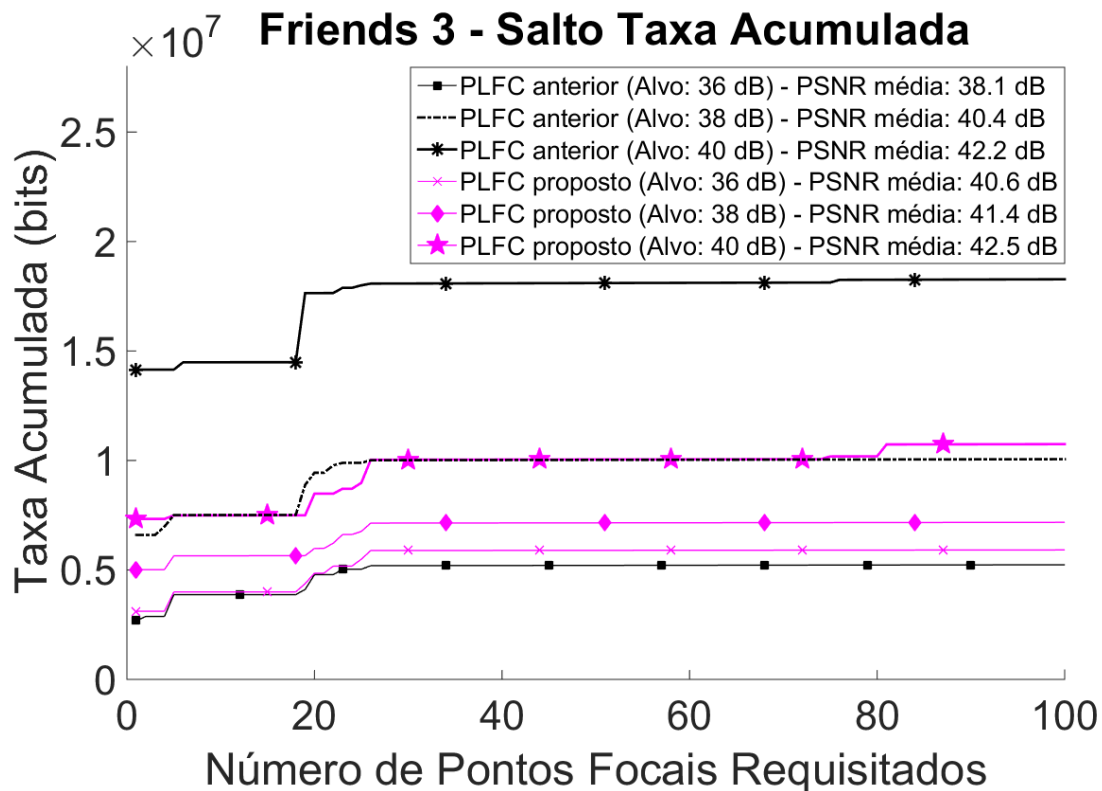


Figura 5.17: Ilustração do comportamento do método proposto em comparação com PLFC original [17]. Nota-se que conforme a qualidade requisitada aumenta, o PLFC original precisa mandar uma taxa significativamente maior em comparação com a qualidade imediatamente anterior. Por outro lado, o método proposto apresenta um salto menor em termos de taxa à medida que a qualidade aumenta.

Outra análise que merece destaque diz respeito ao salto de taxa necessária para atingir determinada qualidade (PSNR), como mostra a Figura 5.17. Nesse sentido, em [17] era preciso enviar quase o dobro de *bits* quando a PSNR aumentava de 36 dB para 38 dB por exemplo. Em contrapartida, os resultados deste trabalho demonstram um salto menor em termos de taxa, quando se aumenta a qualidade alvo. Diante disso, no caso do LF *Friends 3*, nota-se que o resultado do PLFC anterior para 40 dB de alvo é muito parecido com o resultado obtido pelo PLFC anterior [17] quando a PSNR requerida é de 38 dB

Embora essas qualidades alvo sejam diferentes, este método envia quase a mesma taxa acumulada que [17] nesse caso específico, mas atinge uma qualidade superior. Essa situação ilustra bem o ganho obtido em termos de taxa e pode ser vista pelas duas linhas quase sobrepostas, mais ao meio da Figura 5.17. Os resultados referentes ao LF *Friends 3* podem ser observados separadamente no Apêndice A.

Em geral, os resultados referentes à taxa acumulada demonstram ganhos ainda mais evidentes conforme a qualidade alvo aumenta, como ilustram as figuras anteriores relativas aos LFs *Sophie & Vincent 1* e *Sophie Krios & Vincent* por exemplo. Isso se deve tanto pela definição dinâmica do QP utilizado para a codificação das imagens transmitidas, quanto pela melhor definição da *cache* inicial, uma vez que o usuário aprende imagens realmente úteis para as sintetizações subsequentes.

Finalmente, o cálculo proposto para obter o μ da Equação (4.7) mostrou-se eficaz, uma vez que o seu uso gerou resultados melhores ou iguais a quando ele foi ignorado. Tal parâmetro reduziu a taxa acumulada na maioria dos casos e, via de regra, aumentou a PSNR atingida. A Figura 5.18 e a Figura 5.19 comparam o resultado do PLFC proposto com uma versão alternativa que considera apenas o benefício imediato, não utilizando o μ , para dois LFs diferentes.

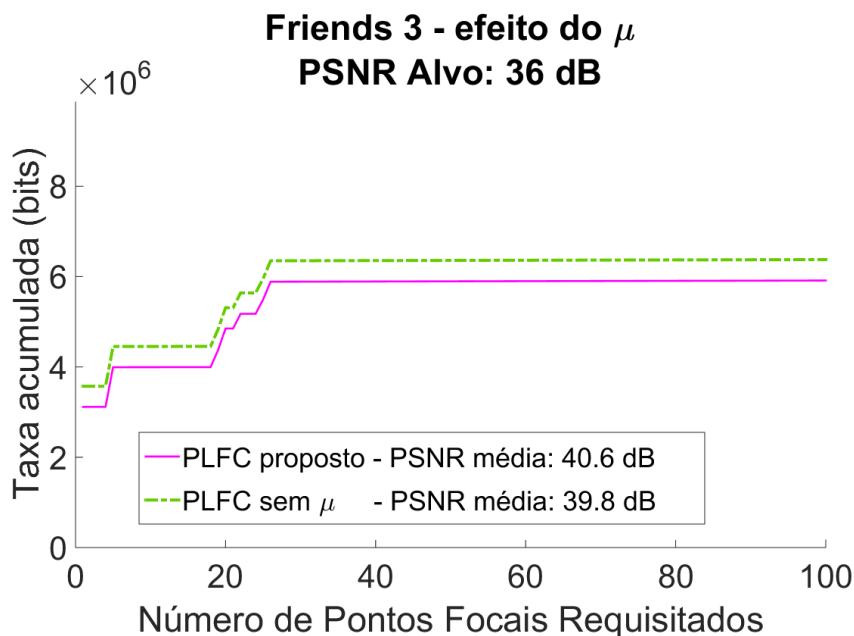


Figura 5.18: Análise do efeito do benefício futuro, o qual é controlado pelo parâmetro μ da Equação (4.7), na transmissão do *light field* *Friends 3*. No caso que considera o benefício futuro, a taxa acumulada final foi menor e, mesmo assim, a qualidade atingida (PSNR) foi maior.

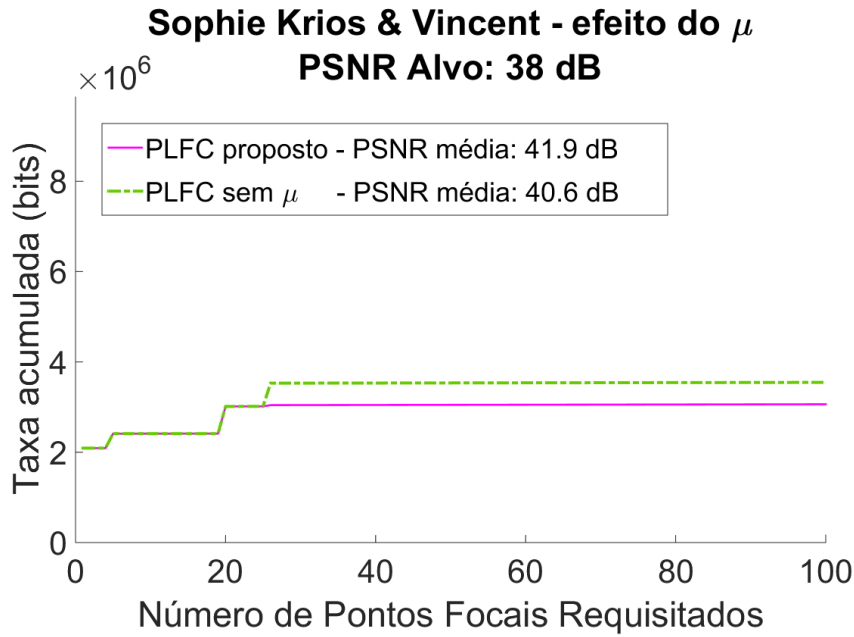


Figura 5.19: Análise do efeito do benefício futuro, controlado pelo μ da Equação (4.7), na transmissão do LF Sophie Krios & Vincent. Nota-se a diminuição da taxa acumulada e o aumento da qualidade atingida (PSNR) quando se considera também o benefício futuro.

Diante disso, percebe-se que esse μ é significativo o bastante para fazer com que a escolha das imagens de subabertura transmitidas leve em consideração os pontos focais que ainda serão requisitados. De fato, o benefício futuro cumpre o que foi estabelecido no segundo item da Seção 4.2.

No próximo capítulo, encontra-se a conclusão deste trabalho, bem como sugestões de trabalhos futuros.

Capítulo 6

Conclusões

Este trabalho propôs um método de compressão e transmissão de imagens *light field* sintetizadas. As imagens LF em questão são requisitadas por um observador e se referem a diferentes pontos focais, mas adaptar o sistema para gerar diferentes pontos de vista é relativamente simples, bastando que se modifique a maneira de sintetizar as imagens a serem transmitidas.

Sendo um aprimoramento do PLFC anterior [17], este método corrige algumas pontualidades do trabalho citado bem como propõe um novo modo para refinar as imagens de subabertura selecionadas para transmissão. Tal algoritmo de refinamento consegue enviar menos imagens, portanto menos taxa, e ainda assim atinge uma qualidade consideravelmente maior que [17] para a grande maioria dos LFs considerados. Esse processo é favorecido pela contribuição mínima de qualidade estabelecida para cada SAI, o que aumenta a qualidade obtida (medida pela PSNR) durante a codificação e evita a transmissão de imagens pouco importantes para a sintetização.

Além disso, deve-se mencionar que o refinamento também é usado para estipular o tamanho ótimo da *cache* inicial, diminuindo a taxa necessária para a inicialização do sistema e selecionando as melhores imagens disponíveis. Dessa forma, tem-se um conjunto de SAIs de qualidade o qual pode ser usado para sintetizar as próximas imagens de ponto focal. Ademais, foi proposto um cômputo automático do multiplicador de Lagrange que combina os benefícios futuro e imediato.

Os experimentos realizados apontam ganhos cada vez maiores conforme mais imagens vão sendo requisitadas pelo usuário, os quais acontecem mais cedo em comparação aos ganhos de [17]. Assim, os resultados obtidos foram satisfatórios, uma vez que apontaram uma redução em termos da taxa acumulada de até 85,8% com relação a abordagem ILFS [20–24] e de até 32,8% se comparado ao PLFC anterior [17]. O ganho médio foi também significativo, sendo de 75,5% em comparação com transmissões triviais ILFS e de 18,2% frente a [17].

Esses resultados refletem bem a efetividade das modificações propostas, ressaltando-se a importância da aplicação do algoritmo de refinamento tanto na inicialização da *cache* quanto na transmissão de novas SAIs. Em boa parte dos casos, a transmissão de imagens adicionais foi bem menos frequente do que acontece em [17], sendo que em algumas situações ela nem foi necessária, pois a *cache* inicial já era boa o bastante para sintetizar todas as imagens referentes à qualidade mínima desejada (LF Sophie & Vincent e LF Books).

A seleção dinâmica do QP para codificar as SAIs corrigiu um caso peculiar de [17] no qual todo o LF era transmitido, mas o limiar mínimo de qualidade desejada não era alcançado. Nota-se ainda que o controle da influência do benefício futuro tem um papel fundamental no processo de escolha das novas imagens. Isso porque o refinamento reavalia cada uma das SAIs selecionadas para a transmissão, não só pensando no momento imediato, mas também considerando as sintetizações posteriores.

Trabalhos futuros podem ter como foco o tratamento de imagens LF específicas, nas quais existam duas ou mais profundidades muito distantes, havendo um plano de fundo com muita textura, como em **Chain-link Fence 1**. Nesse caso em particular, este método superou transmissões triviais apenas depois que o 80º ponto focal foi requisitado. Adicionalmente pode-se analisar a situação em que apenas uma imagem adicional é necessária para atingir a qualidade alvo. Talvez seja melhor o servidor transmitir duas imagens ao invés de uma, pois além de poder usar um QP maior para cada imagem, logo menos taxa total, ele ainda pode aplicar o refinamento e selecionar as imagens ideais a serem enviadas para o cliente.

Por fim, seria pertinente usar outros codificadores, como o HEVC, e testar o método proposto em cenários reais, isto é, com usuários reais solicitando pontos focais ou pontos de vista, com o intuito de validar o sistema pragmaticamente.

Referências

- [1] G. Lippmann, “Epreuves reversibles donnant la sensation du relief,” *J. Phys. Theor. Appl.*, pp. 821–825, 1908. [viii, 1, 2](#)
- [2] R. Ng, “Digital light field photography,” *A dissertation submitted to the department of computer science and the committee on graduate studies of Stanford University*, 2006. [viii, 3, 7, 8, 9, 11, 12, 13, 16, 17](#)
- [3] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, “Interactive digital photomontage,” *Transactions on Graphics (Proceedings of SIGGRAPH)*, vol. 23, no. 3, pp. 292–300, 2004. [viii, 3](#)
- [4] M. Lingenauber, K. Strobl, H. Klaus, W. Oumer, and S. Kriegel, “Benefits of plenoptic cameras for robot vision during close range on-orbit servicing maneuvers,” *2017 IEEE Aerospace Conference*, pp. 1–18, março 2017. [viii, 6, 7, 10](#)
- [5] Lytro. Disponível em <https://support.lytro.com/hc/en-us>, 2006. Acessado em 13 de maio de 2018. [viii, 9](#)
- [6] S. Zhao, Z. Chen, K. Yang, and H. Huang, “Light field image coding with hybrid scan order,” *Visual Communications and Image Processing - VCIP*, pp. 1–4, 2016. [viii, 2, 9](#)
- [7] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, “Light field photography with a hand-held plenoptic camera,” *Stanford Tech Report CTSR*, 2005. [ix, 1, 2, 9, 10, 13, 16, 17, 18, 31](#)
- [8] G. Wetzstein, “Light field photography - stanford university.” Disponível em <http://stanford.edu/class/ee367/slides/lecture8.pdf>, 111 Slides, 2015. Acessado em 06 de junho de 2018. [ix, 11](#)
- [9] D. Dansereau, “Mathworks light field toolbox v0.4 (lftoolbox.pdf).” Disponível em <https://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4>, 2015. Acessado em 10 de setembro de 2016. [ix, 15, 16, 23](#)
- [10] M. Levoy, “Light fields and computational imaging,” *IEEE Computer*, vol. 39, no. 8, pp. 46–55, 2006. [x, 14, 17, 18](#)
- [11] C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch, “Refocusing distance of a standard plenoptic camera,” *Opt. Express*, vol. 24, pp. 21521–21540, Sep 2016. [x, 20, 21](#)

- [12] C. Hahne, “The standard plenoptic camera: applications of a geometrical light field model,” *A thesis submitted to the University of Bedfordshire*, 2016. x, 3, 20, 21
- [13] V. Vaish, B. Wiblurn, N. Joshi, and M. Levoy, “Using plane + parallax for calibrating dense camera arrays,” *Proceedings of CVPR*, 2004. x, 18, 22, 31
- [14] Canon, “Eos movie compression options: All-i and ipb.” Disponível em http://www.canon.com.hk/cpx/en/technical/va_EOS_Movie_Compression_Options_All_I_and_IPB.html, 2014. Acessado em 19 de fevereiro de 2017. x, 24
- [15] Lytro. Disponível em <https://www.lytro.com/>, 2006. Acessado em 17 de fevereiro de 2017. x, 2, 6, 26, 27
- [16] G. Alves, F. Pereira, and E. da Silva, “Light field imaging coding: Performance assessment methodology and standards benchmarking,” *IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, pp. 1–6, 2016. x, 26, 27
- [17] E. Peixoto, B. Macchiavello, E. Mintsu, C. Dorea, and G. Cheung, “Progressive communication for interactive light field image data streaming,” *International Conference on Image Processing - ICIP*, 2017. x, xi, xii, xiv, 28, 29, 30, 31, 32, 33, 34, 37, 38, 40, 44, 45, 46, 47, 49, 51, 52, 53, 54, 56, 57, 68
- [18] W. B. de Souza, B. Macchiavello, E. Peixoto, E. M. Hung, and G. Cheung, “A sub-aperture image selection refinement method for progressive light field transmission,” *IEEE International Workshop on Multimedia Signal Processing - MMSP*, 2018. x, 30, 33, 35, 37, 38, 39
- [19] G. Bjøntegaard, “Improvement of the BD-PSNR model,” *ITU-T SG16/Q6 35th VCEG Meeting VCEG-A111*, 2008. xiv, 26, 28, 49, 51, 52
- [20] P. Ramanathan and B. Girod, “Random access for compressed light fields using multiple representations,” *IEEE International Workshop on Multimedia Signal Processing, Siena*, 2004. xiv, 28, 43, 51, 52, 56
- [21] A. Aaron, P. Ramanathan, and B. Girod, “Wyner-ziv coding of light fields for random access,” *IEEE International Workshop on Multimedia Signal Processing*, 2004. xiv, 26, 27, 28, 43, 51, 52, 56
- [22] W. Cai, G. Cheung, T. Kwon, and S.-J. Lee, “Optimized frame structure for interactive light field streaming with cooperative cache,” *IEEE International Conference on Multimedia and Expo, Barcelona*, 2011. xiv, 28, 43, 51, 52, 56
- [23] W. Cai, G. Cheung, S. J. Lee, and T. Kwon, “Optimal frame structure design using landmarks for interactive light field streaming,” *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE*, pp. 1445—1448, 2012. xiv, 28, 43, 51, 52, 56
- [24] B. Motz, G. Cheung, and A. Ortega, “Redundant frame structure using m-frame for interactive light field streaming,” *Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, 2016.*, pp. 1369–1373, 2016. xiv, 28, 43, 51, 52, 56

- [25] E. H. Adelson and J. R. Bergen, “The plenoptic function and the elements of early vision,” *Computational Models of Visual Processing*, pp. 3–20, 1991. 1
- [26] M. Levoy and P. Hanrahan, “Light field rendering,” *Proc. of ACM SIGGRAPH 96*, p. 31–42, 1996. 1, 2, 31
- [27] H. Ives, “Parallax panoramagrams made with a large diameter lens,” *Journal of the Optical Society of America*, vol. 20, no. 6, pp. 332–342, 1930. 1
- [28] T. Georgiev, A. Lumsdaine, and G. Chunev, “Spatial analysis of discrete plenoptic sampling,” *Proc. SPIE 8299, Digital Photography VIII*, vol. 8299, 2012. 2
- [29] E. H. Adelson and J. Y. A. Wang, “Single lens stereo with plenoptic camera,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, p. 99–106, 1992. 2
- [30] A. Isaksen, L. McMillan, and S. Gortler, “Dynamically reparametrized light fields,” *Proc. of ACM SIGGRAPH*, 2000. 2
- [31] R. Szeliski, “Shape and appearance models from multiple images,” *Workshop on Image Based Rendering and Modelling*, 1998. 2
- [32] T. Bishop and P. Favaro, “The light field camera: Extended aliasing, and super-resolution,” *Pattern Analysis and Machine Intelligence, IEEE Trans*, vol. 34, no. 5, p. 972–986, 2012. 2
- [33] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, and E. Antunez, “High performance imaging using large camera arrays,” *ACM Trans. on Graphics (TOG)*, vol. 24, no. 3, pp. 765 – 776, 2005. 2
- [34] Z. Xu, J. Ke, and E. Lam, “High-resolution lightfield photography using two masks,” *Optics Express*, vol. 20, no. 10, p. 10971–13983, 2012. 2
- [35] T. Svoboda, D. Martinec, and T. Pajdla, “A convenient multicamera self-calibration for virtual environments,” *Presence: Teleoperators and Virtual Environments*, vol. 14, no. 4, p. 407–422, 2005. 2
- [36] Raytrix. Disponível em <https://www.raytrix.de/>, 2008. Acessado em 19 de fevereiro de 2017. 2
- [37] JPEG, “Jpeg pleno abstract and executive summary.” https://jpeg.org/items/20150320_pleno_summary.html, 2015. Acessado em 20 de dezembro de 2016. 2, 25
- [38] J. Smith, *Modern Optical Engineering*. McGraw Hill Professional, 4^a ed., 2007. 11
- [39] J. A. King and R. Correll, *Canon EOS Rebel T6i / 750D For Dummies*. For Dummies, 1^a ed., 2015. 11
- [40] R. White, *How Digital Photography Works*. Que, 2nd ed., 2007. 12
- [41] L. Stroebel, J. Compton, I. Current, and R. Zakia, *Photographic Materials and Processes*. Focal Press, 1986. 16

- [42] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas, “Synthetic aperture confocal imaging,” *IACM Transactions on Graphics (Proceedings of SIGGRAPH 2004)*, vol. 23, no. 3, pp. 822–831, 2004. 18
- [43] J. A. Simpson and E. S. Weiner, *The Oxford English Dictionary*. Oxford University Press, 2^a ed., 1989. 18
- [44] C. A. L. Bailer-Jones, “Estimating distances from parallaxes coryn a.l. bailer-jones,” *Publications of the Astronomical Society of the Pacific*, vol. 127, no. 956, 2015. 18
- [45] C. Hahne, “The standard plenoptic camera.” Disponível em <http://www.plenoptic.info/pages/refocusing.html>. Acessado em 22 de junho de 2018. 21
- [46] R. Gonzalez and R. Woods, *Digital Image Processing*. Prentice Hall, 3rd ed., 2010. 23
- [47] R. Gonzalez and R. Woods, *Digital Image Processing*. Prentice Hall, 2nd ed., 2002. 23
- [48] JPEG, “Jpeg.” Disponível em <https://jpeg.org/jpeg/index.html>, 1994. Acessado em 15 de fevereiro de 2017. 23, 25
- [49] K.Sayood, *Introduction to Data Compression*. Elsevier, 3^a ed., 2006. 24, 25, 35
- [50] T. Wiegand, G. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the h.264/avc video coding standard,” *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003. 25
- [51] JPEG, “Jpeg 2000.” Disponível em <https://jpeg.org/jpeg2000/>, 1999. Acessado em 15 de fevereiro de 2017. 25
- [52] ITU-T Recommendation H.265 | ISO/IEC 23008-2:2013, *High efficiency video coding*, 2013. Int’l Telecomm. Union | Int’l Organization for Standardization, Geneva, Switzerland, 2013. 26
- [53] ITU-T Recommendation H.264 | ISO/IEC 14496-10:2005 (MPEG-4), *Advanced Video Coding for Generic Audiovisual Services*, 2005. Int’l Telecomm. Union | Int’l Organization for Standardization, Geneva, Switzerland, 2005. 26, 43
- [54] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Transactions on Information Theory*, vol. IT-22, no. 1, pp. 1–10, 1976. 26
- [55] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973. 27
- [56] C. L. Chang, X. Zhu, P. Ramanathan, and B. Girod, “Light field compression using disparity-compensated lifting and shape adaptation,” *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 793–806, 2006. 27

- [57] M. Magnor and B. Girod, "Model-based coding of multi-viewpoint imagery," *Proc. SPIE Visual Communications Image Processing, Australia*, vol. 1, pp. 14–22, 2000. 27
- [58] J. M. Santos, P. A. Assuncao, L. A. Cruz, L. Távora, R. P. Pinto, and S. M. Faria, "Lossless light-field compression using reversible colour transformations," *Image Processing Theory, Tools and Applications (IPTA), 2017*, pp. 1–6, 2017. 27
- [59] T. H. Tran, Y. Baroud, Z. Wang, S. Simon, and D. Taubman, "Light-field image compression based on variational disparity estimation and motion-compensated wavelet decomposition," *International Conference on Image Processing - ICIP, 2017*. 27
- [60] Y. H. Chao, G. Cheung, and A. Ortega, "Pre-demosaic light field image compression using graph lifting transform," *International Conference on Image Processing - ICIP, 2017*. 27
- [61] S. Zhao and Z. Chen, "Light field image coding via linear approximation prior," *International Conference on Image Processing - ICIP, 2017*. 28
- [62] Y. Yuan, G. Cheung, and P. Frossard, "Optimizing landmark insertions for interactive light field streaming," *International Conference on Image Processing - ICIP, 2017*. 28
- [63] Y. Liu, Q. Dai, and W. Xu, "A real time interactive dynamic light field transmission system," *2006 IEEE International Conference on Multimedia and Expo*, pp. 2173–2176, 2006. 28
- [64] P. A. Kara, M. G. Martini, Z. Nagy, and A. Barsi, "Cinema as large as life: Large-scale light field cinema system," *2017 International Conference on 3D Immersion (IC3D)*, pp. 1–8, 2017. 28
- [65] A. S. Moghaddam, P. L. Correia, and F. Pereira, "Light field local binary patterns description for face recognition," *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3815–3819, 2017. 29
- [66] T. C. Wang, J. Y. Zhu, H. Ebi, M. Chandraker, A. A. Efros, and R. Ramamoorthi, "A 4d light-field dataset and CNN architectures for material recognition," *CoRR*, 2016. 29
- [67] M. Rerabek and T. Ebrahimi, "New light field image dataset," *8th International Conference on Quality of Multimedia Experience (QoMEX)*, 2016. 40
- [68] J. Moreno, B. Jaime, and S. Saucedo, "Towards no-reference of peak signal to noise ratio," *International Journal of Advanced Computer Science and Applications - IJACSA*, vol. 4, no. 1, pp. 123–130, 2013. 43
- [69] W. Hines, D. Montgomery, D. Goldsman, and C. Borrer, *Probability and statistics in engineering*. John Wiley & Sons, Inc, 4^a ed., 2003. 43
- [70] Mathworks, "Matlab software." <http://www.mathworks.com/products/matlab/>, 1994-2018. Acessado em 01 de abril de 2018. 44

Apêndice A

Resultados Adicionais

Neste apêndice está o restante dos resultados obtidos. Assim como no Capítulo 5, eles se encontram agrupados por LF e divididos em duas seções, uma para a taxa acumulada e outra para os resultados referentes à taxa distorção.

A.1 Taxa acumulada

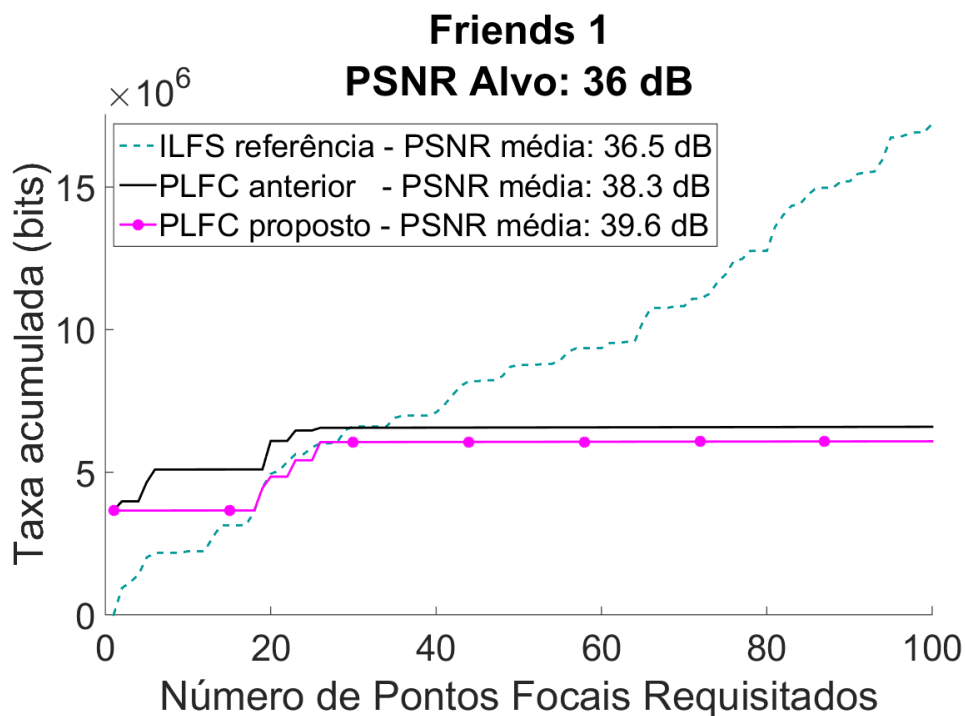


Figura A.1: Resultado da transmissão de pontos focais sintetizados referentes ao *light field Friends 1*, com PSNR alvo de 36 dB. A taxa acumulada, em *bits*, encontra-se no eixo vertical e o número de cada imagem requisitada (de 1 a 100) está no eixo horizontal. São comparados os resultados obtidos pelo ILFS de referência, os resultados do PLFC anterior bem como os resultados do PLFC proposto.

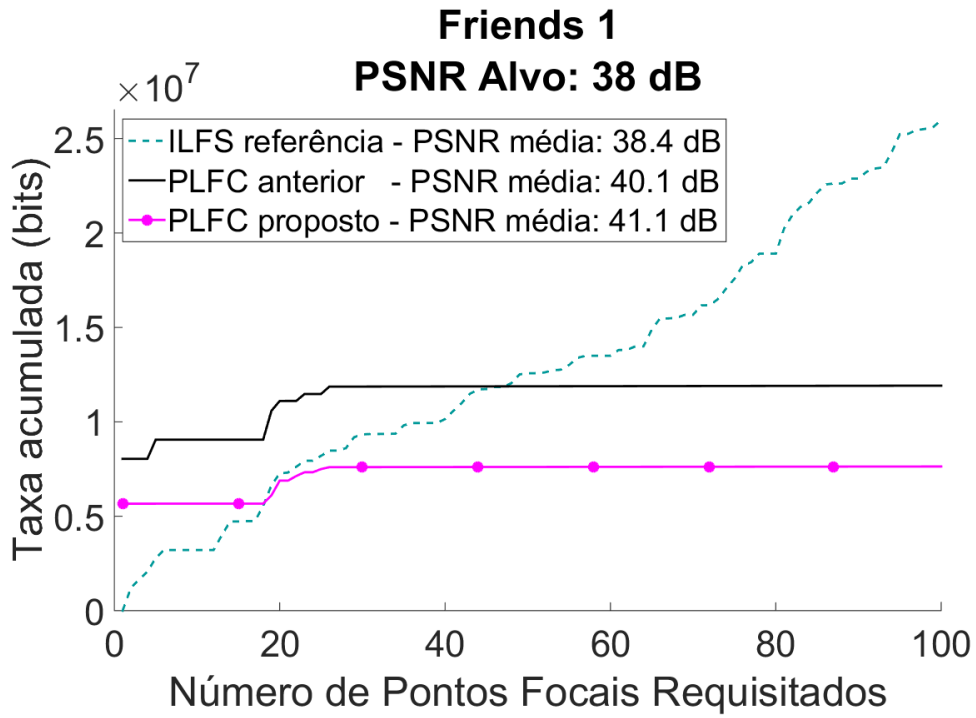


Figura A.2: Resultado da transmissão de pontos focais sintetizados para o *light field* Friends 1, com PSNR alvo de 38 dB.

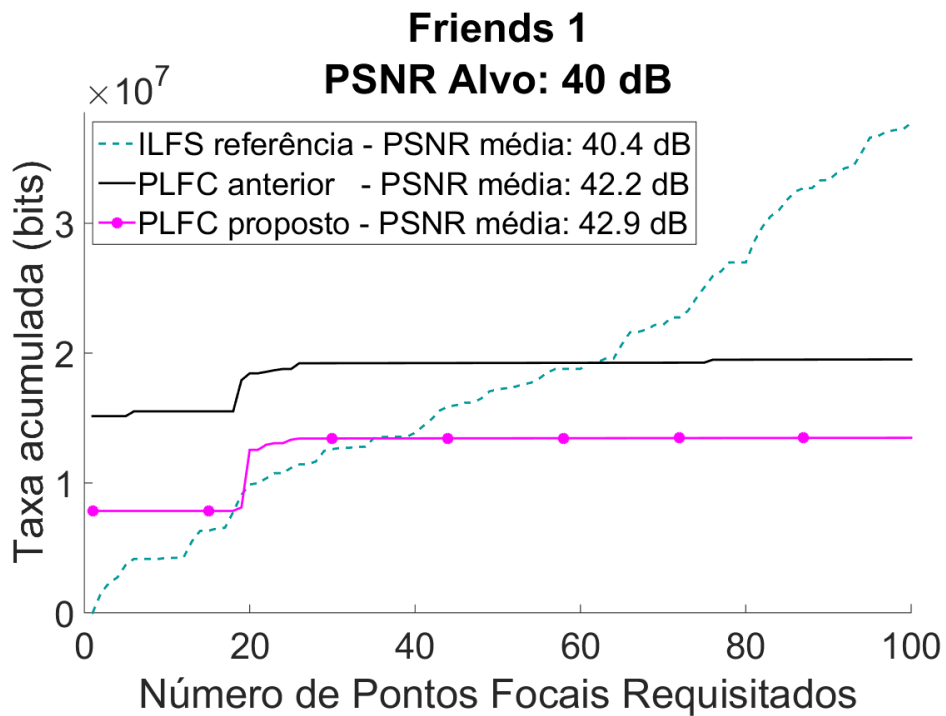


Figura A.3: Resultado da transmissão de pontos focais sintetizados referentes ao *light field* Friends 1, PSNR alvo de 40 dB.

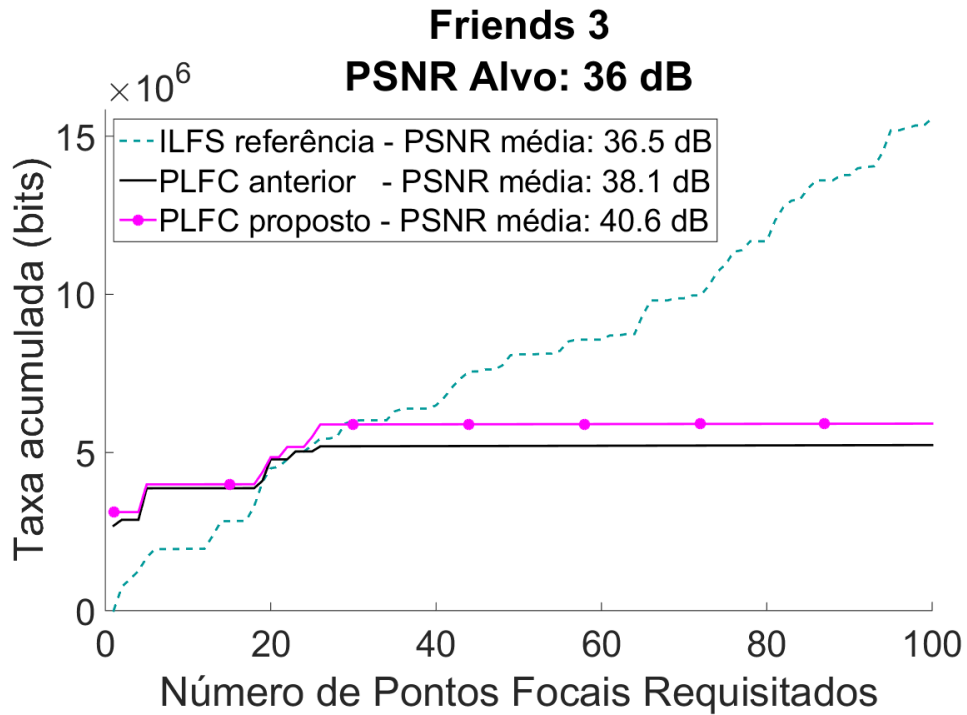


Figura A.4: Resultado da transmissão de pontos focais sintetizados referentes ao LF Friends 3, PSNR alvo de 36 dB.

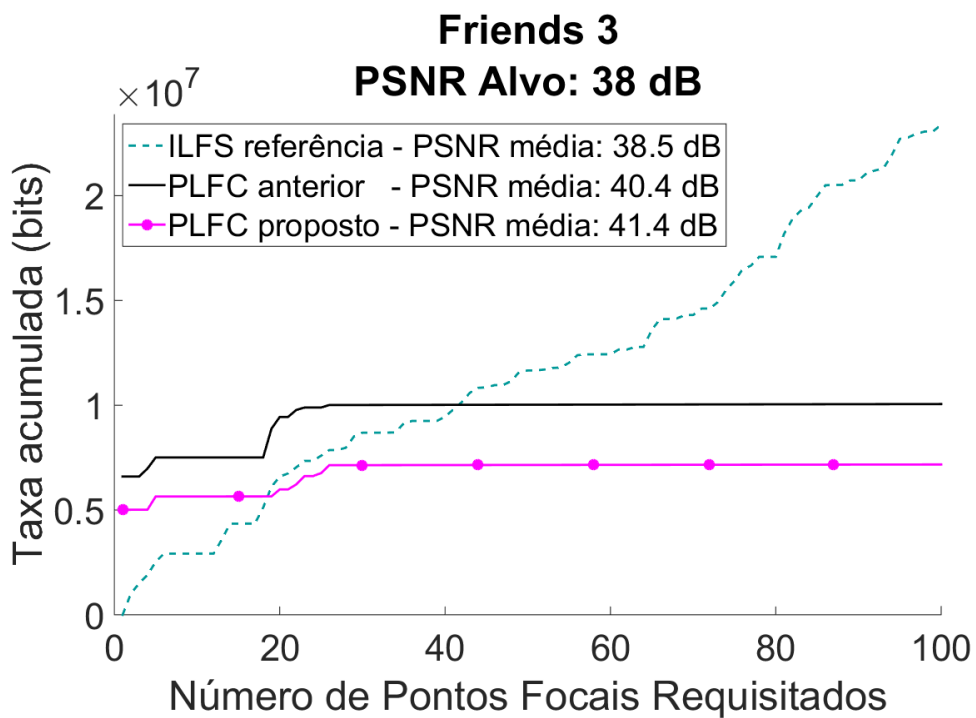


Figura A.5: Resultado da transmissão de pontos focais sintetizados referentes ao LF Friends 3, PSNR alvo de 38 dB.

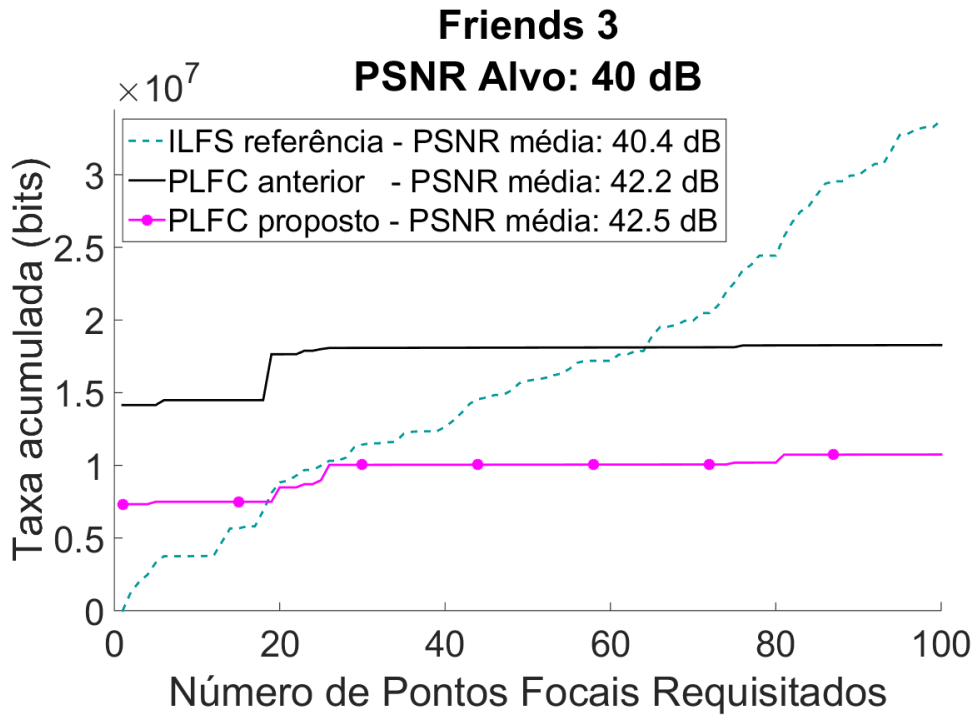


Figura A.6: Resultado da transmissão de pontos focais sintetizados referentes ao LF Friends 3, PSNR alvo de 40 dB.

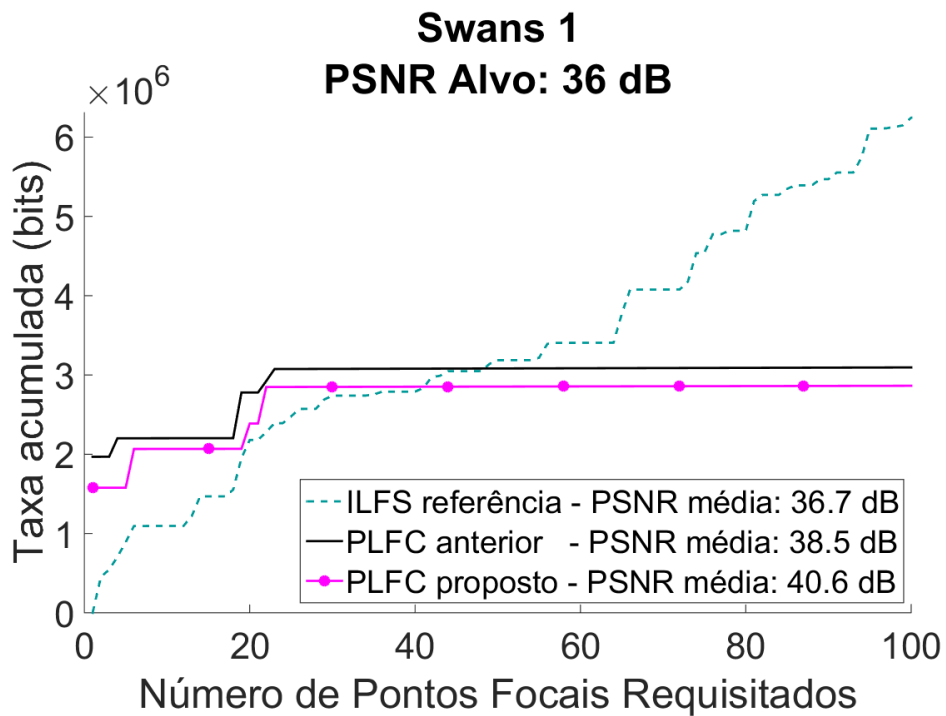


Figura A.7: Resultado da transmissão de pontos focais sintetizados referentes ao LF Swans 1, PSNR alvo de 36 dB.

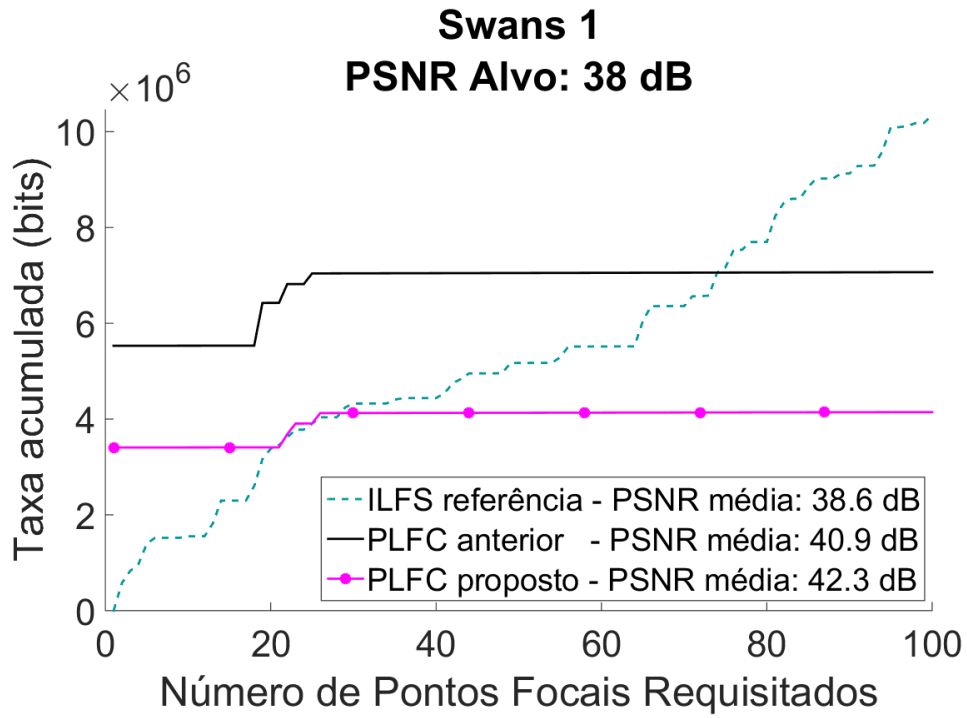


Figura A.8: Resultado da transmissão de pontos focais sintetizados referentes ao LF Swans 1, PSNR alvo de 38 dB.

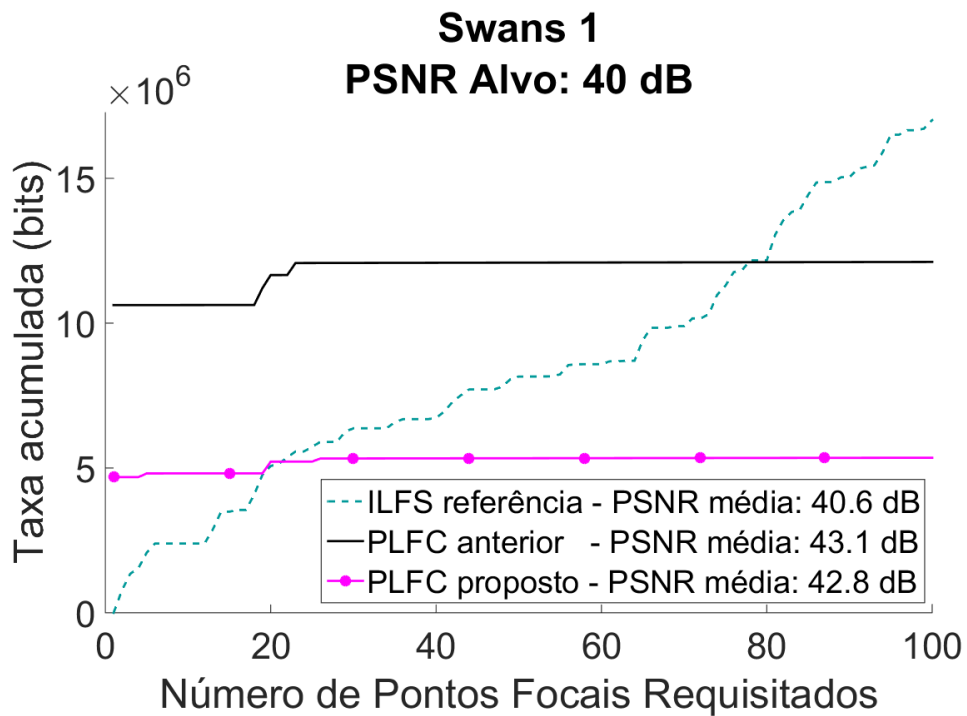


Figura A.9: Resultado da transmissão de pontos focais sintetizados referentes ao LF Swans1, PSNR alvo de 40 dB.

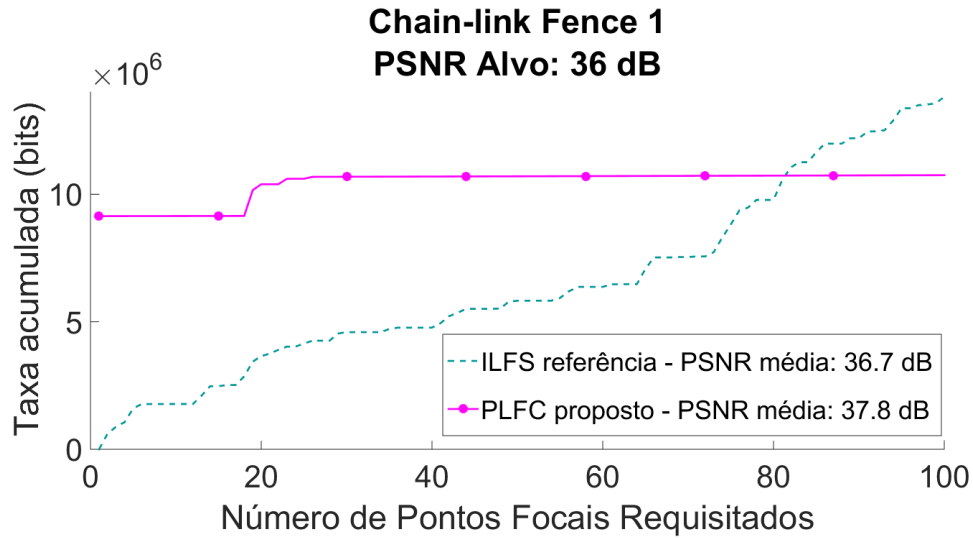


Figura A.10: Resultado da transmissão de pontos focais sintetizados referentes ao *light field* Chain-link Fence 1, com PSNR alvo de 36 dB. A taxa acumulada desde o início da execução do sistema, em *bits*, encontra-se no eixo vertical e o número de cada imagem requisitada (de 1 a 100) está no eixo horizontal. Compara-se os resultados PLFC proposto com os resultados do método referência ILFS.

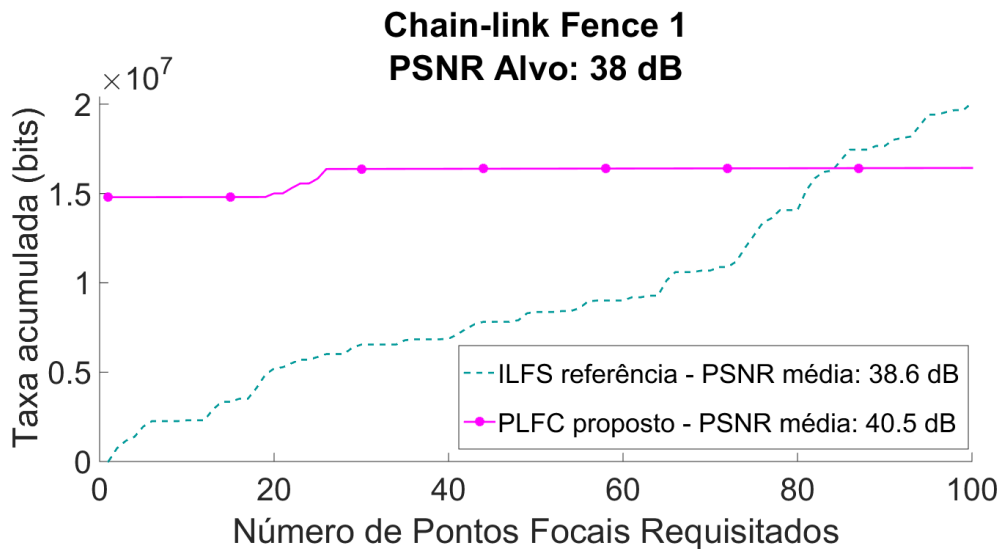


Figura A.11: Resultado da transmissão de pontos focais sintetizados referentes ao *light field* Chain-link Fence 1, com PSNR alvo de 38 dB. Compara-se o resultado desta proposta com o resultado do ILFS.

Nos resultados da imagem Chain-link Fence 1 (Figuras A.10, A.11 e A.12), nota-se a ausência do resultado referente ao PLFC anterior [17]. Isso ocorre pois nesse caso nem a transmissão do LF inteiro consegue atingir a PSNR requerida, como já explicado no Capítulo 5.

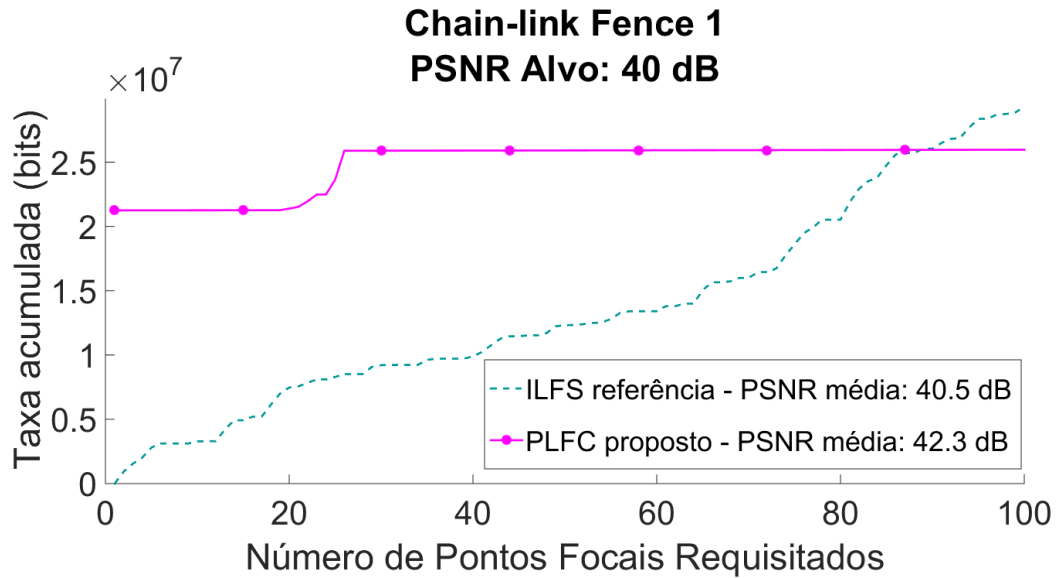


Figura A.12: Resultado da transmissão de pontos focais sintetizados referentes ao *light field* Chain-link Fence 1, com PSNR alvo de 40 dB. É feita a comparação entre este método e o ILFS.

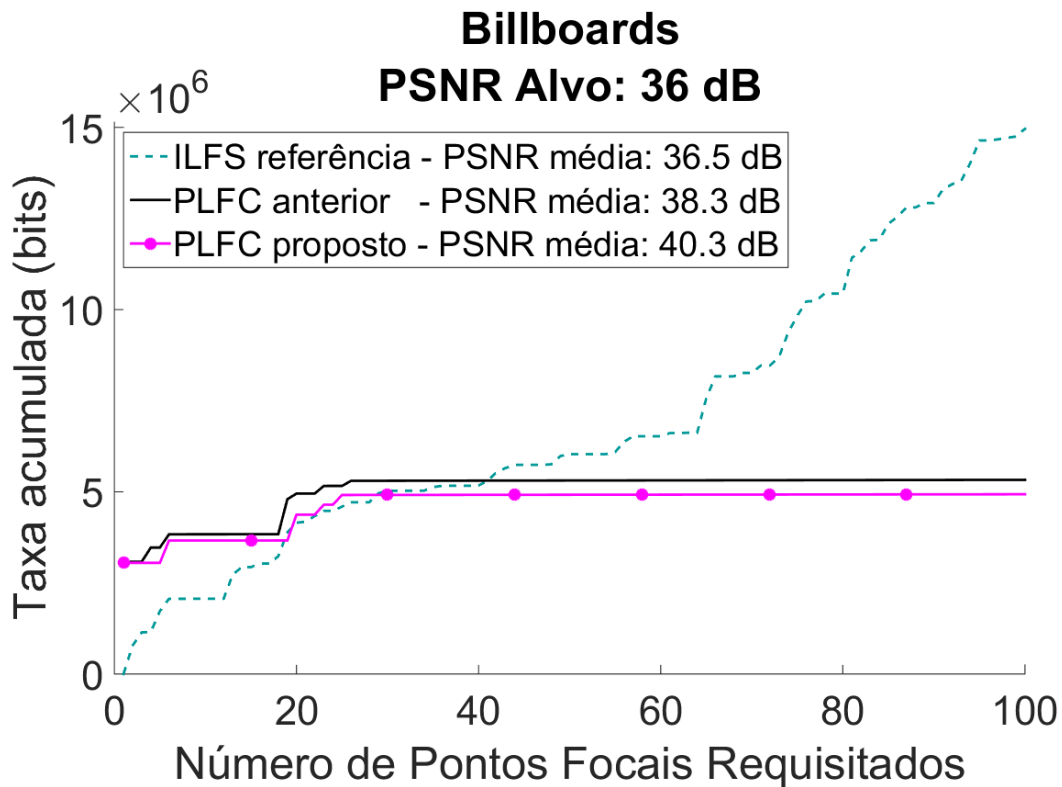


Figura A.13: Resultado da transmissão de pontos focais sintetizados referentes ao *light field* Billboards, com PSNR alvo de 36 dB. Exibe-se o resultado do ILFS, do PLFC e do PLFC proposto.

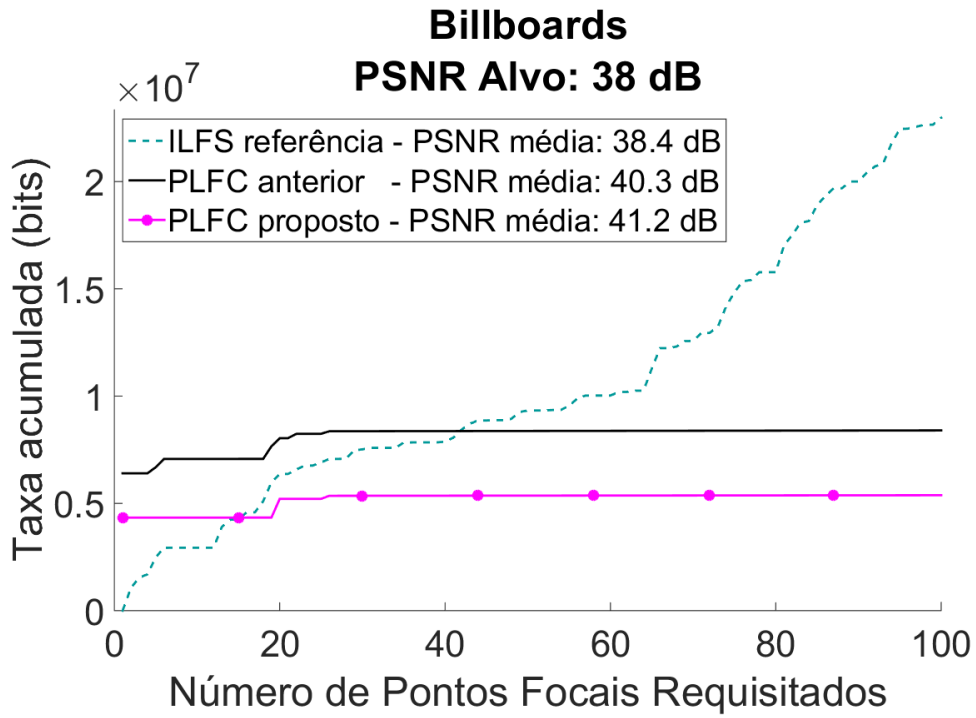


Figura A.14: Resultado da transmissão de pontos focais sintetizados referentes ao *light field* Billboards, com PSNR alvo de 38 dB.

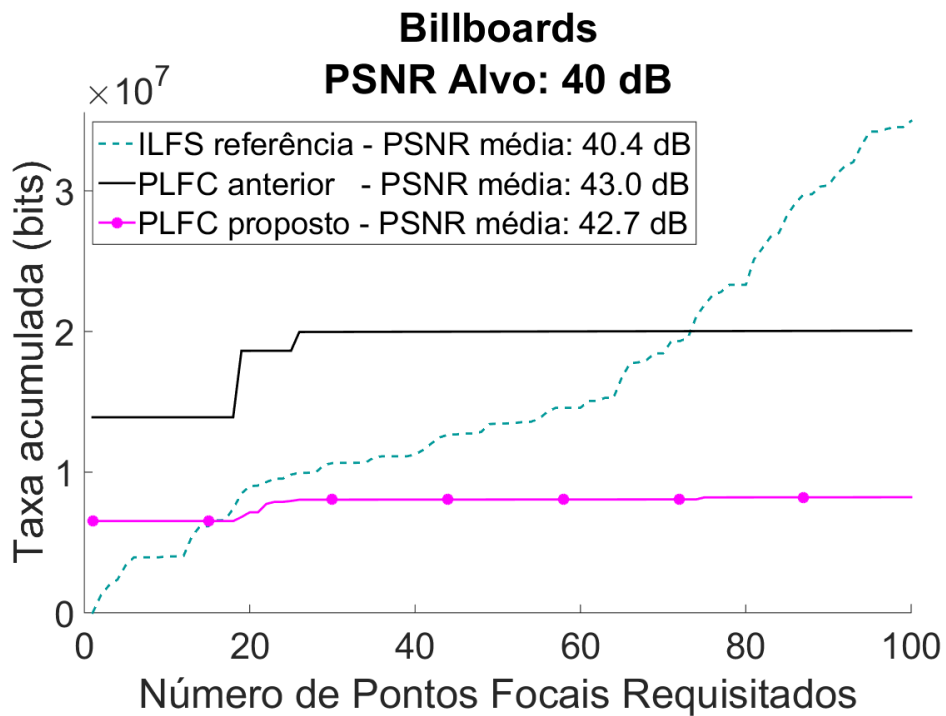


Figura A.15: Resultado da transmissão de pontos focais sintetizados referentes ao *light field* Billboards, PSNR alvo de 40 dB.

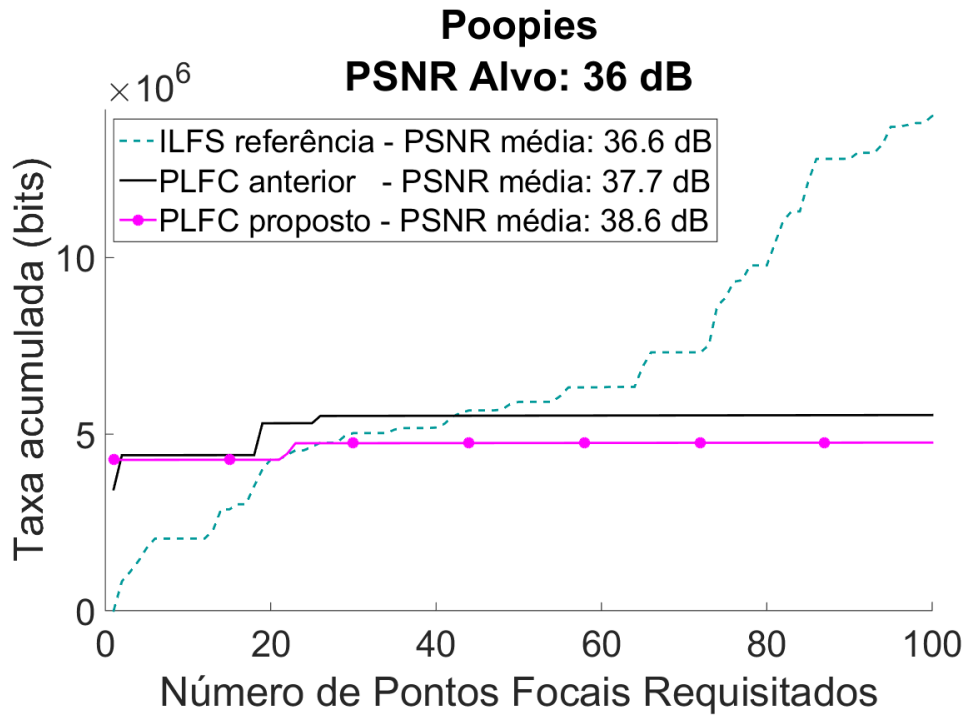


Figura A.16: Resultado da transmissão de pontos focais sintetizados referentes ao *light field* Poopies, PSNR alvo de 36 dB.

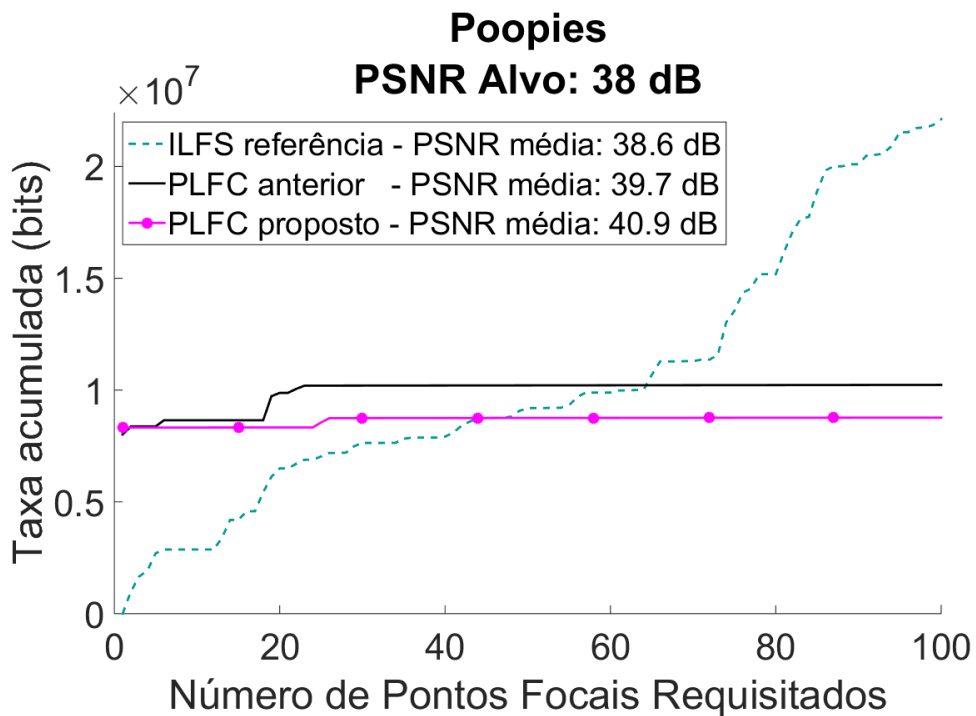


Figura A.17: Resultado da transmissão de pontos focais sintetizados referentes ao *light field* Poopies, PSNR alvo de 38 dB.

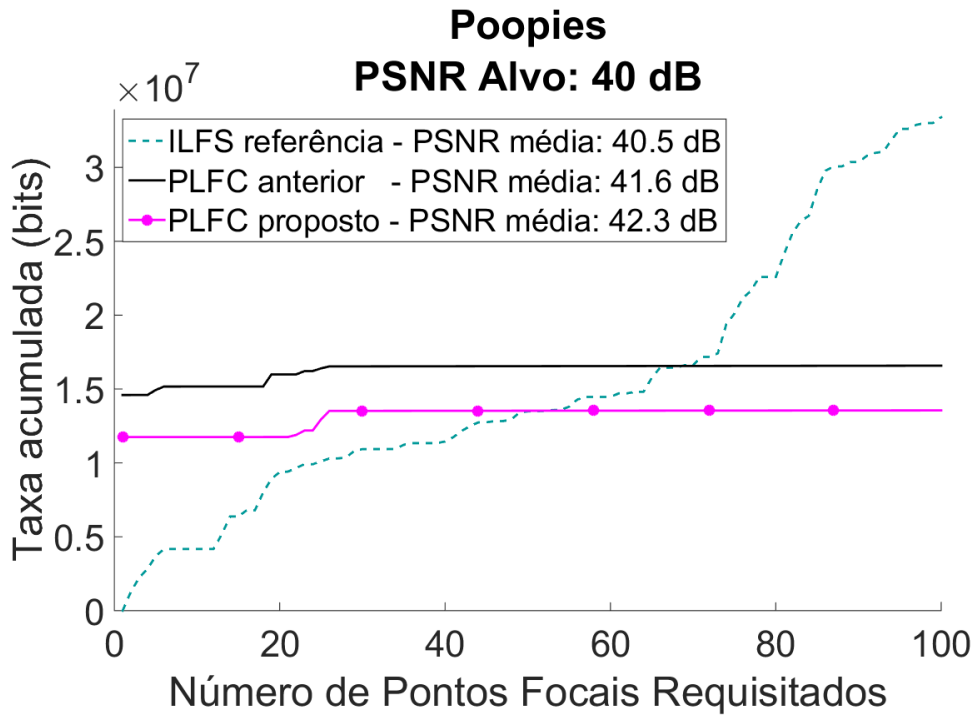


Figura A.18: Resultado da transmissão de pontos focais sintetizados referentes ao *light field* Poopies, PSNR alvo de 40 dB.

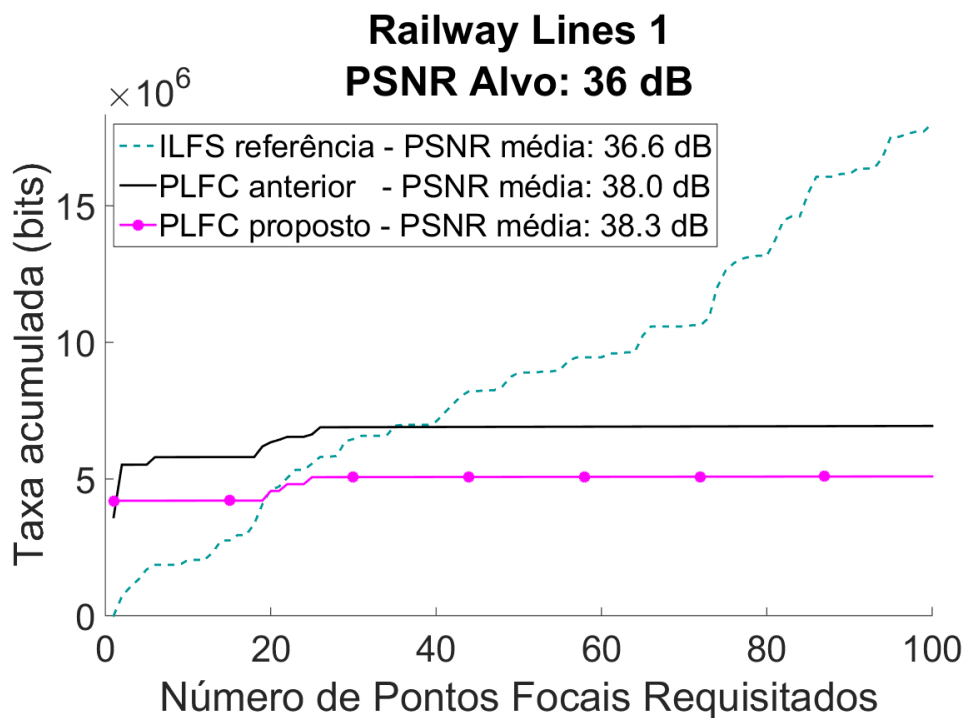


Figura A.19: Resultado da transmissão de pontos focais sintetizados referentes ao LF Railway Lines 1, PSNR alvo de 36 dB.

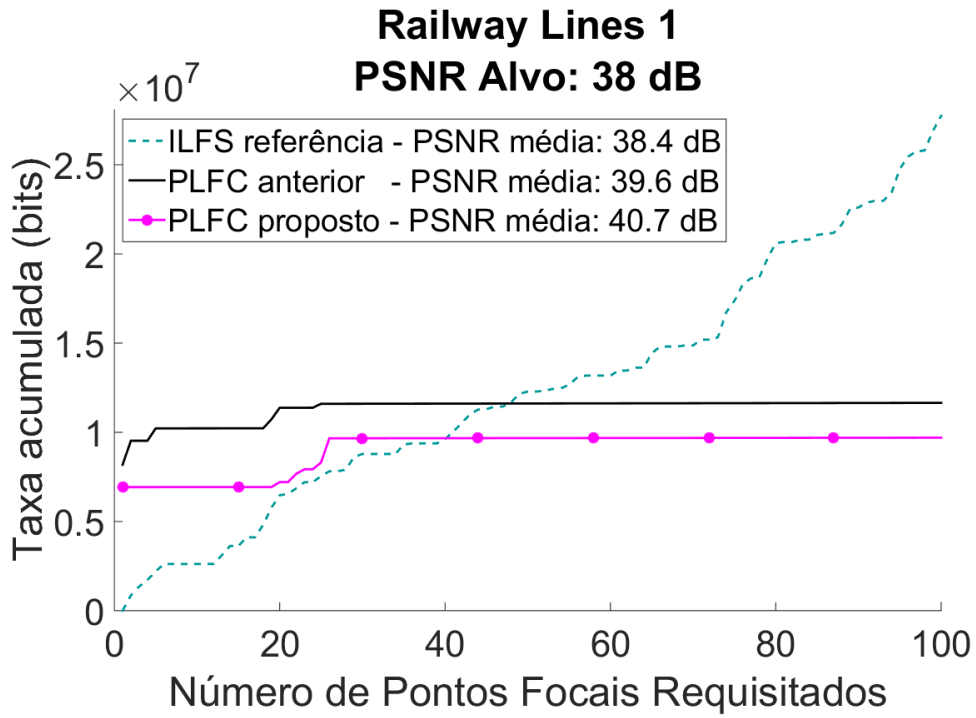


Figura A.20: Resultado da transmissão de pontos focais sintetizados referentes ao LF Railway Lines 1, PSNR alvo de 38 dB.

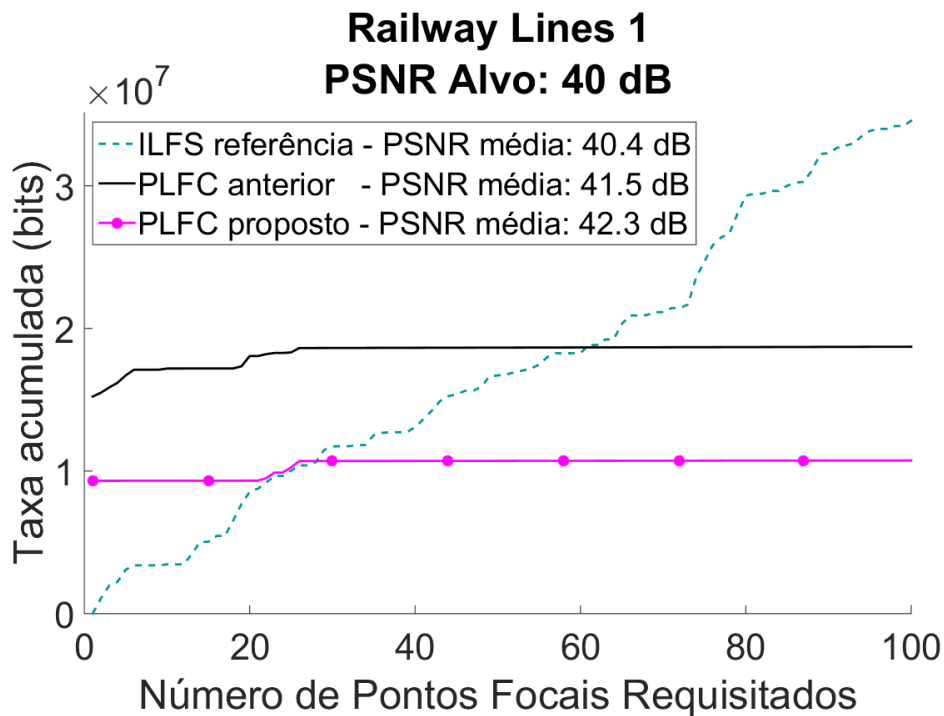


Figura A.21: Resultado da transmissão de pontos focais sintetizados referentes ao LF Railway Lines 1, PSNR alvo de 40 dB.

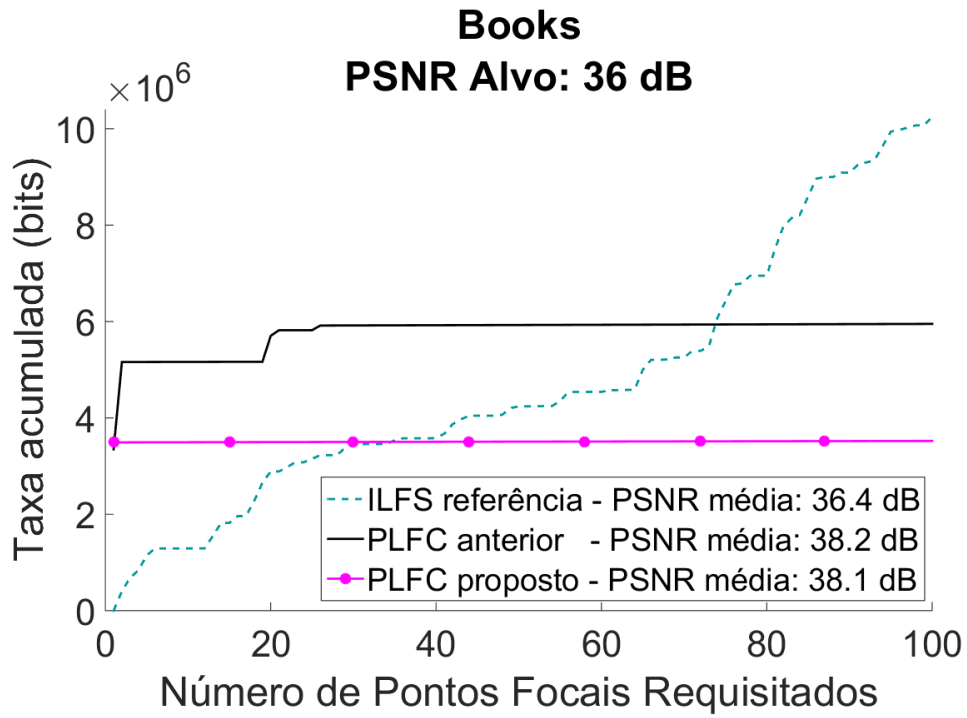


Figura A.22: Resultado da transmissão de pontos focais sintetizados referentes ao LF Books, PSNR alvo de 36 dB.

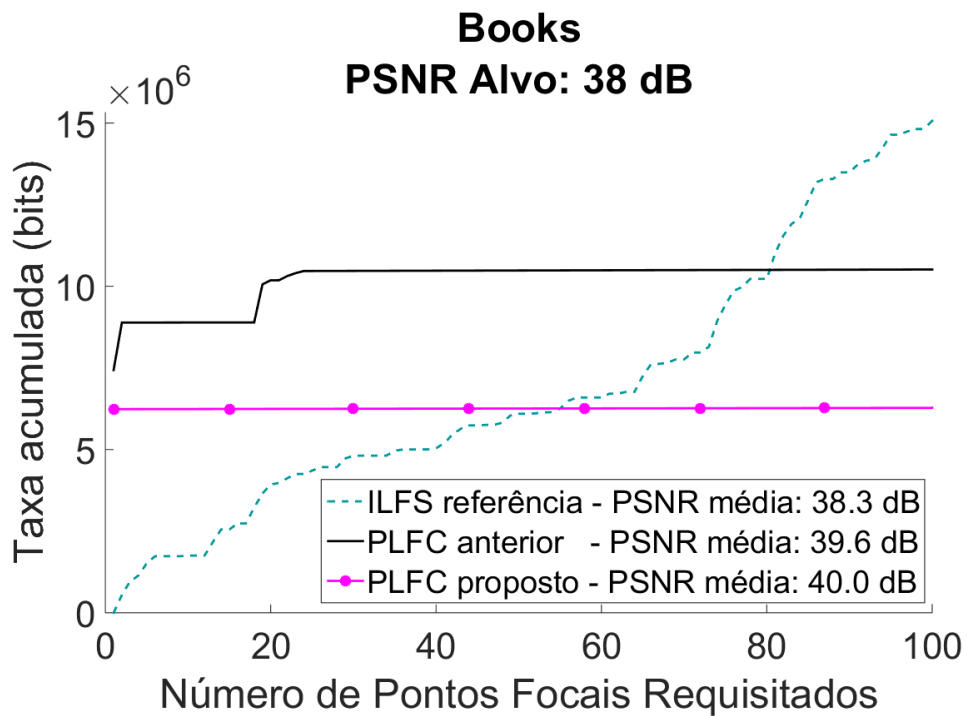


Figura A.23: Resultado da transmissão de pontos focais sintetizados referentes ao LF Books, PSNR alvo de 38 dB.

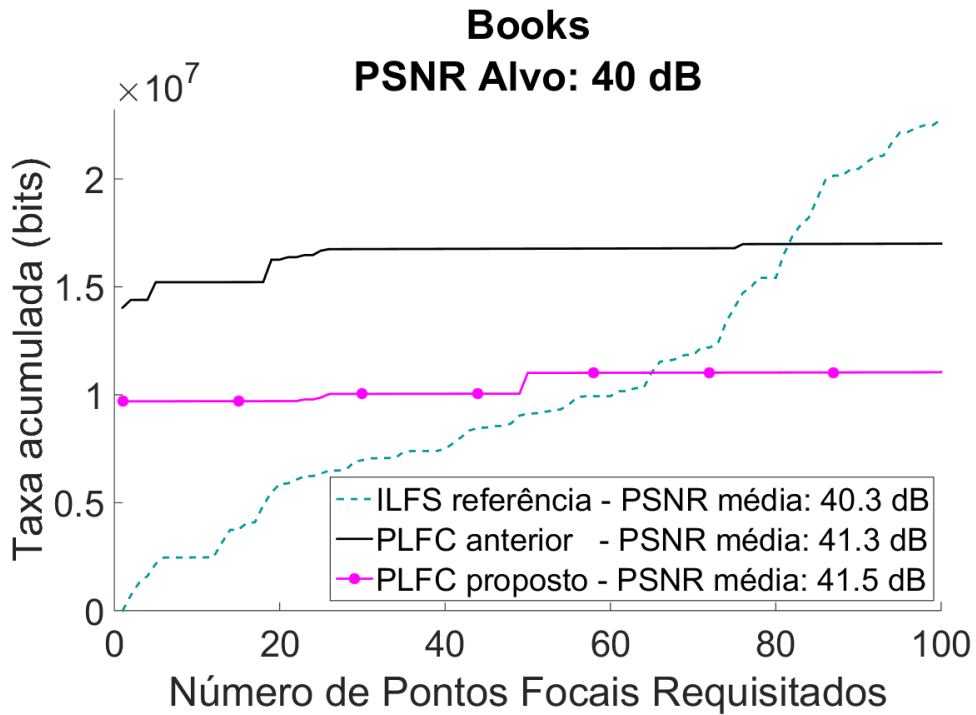


Figura A.24: Resultado da transmissão de pontos focais sintetizados referentes ao LF Books, PSNR alvo de 40 dB.

A.2 Taxa distorção

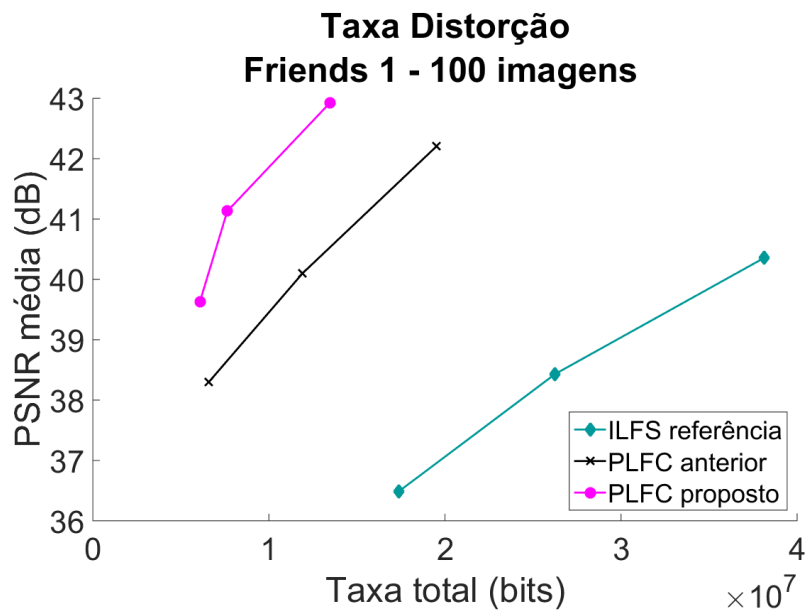


Figura A.25: Resultado taxa distorção referente ao LF Friends 1 após 100 imagens transmitidas. Quanto mais acima e à esquerda uma curva estiver, melhor é o resultado. Exibe-se os resultados do ILFS, do PLFC e do PLFC proposto.

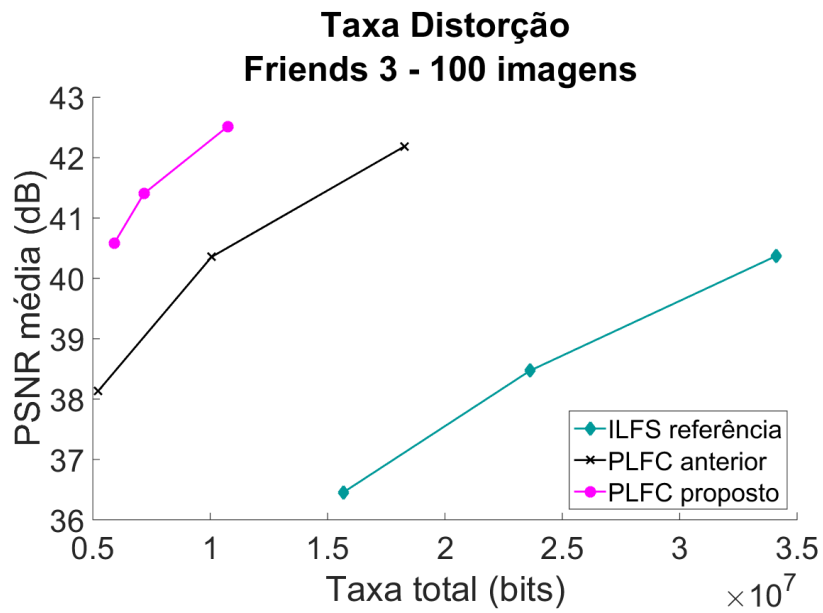


Figura A.26: Resultado taxa distorção referente ao LF Friends 3.

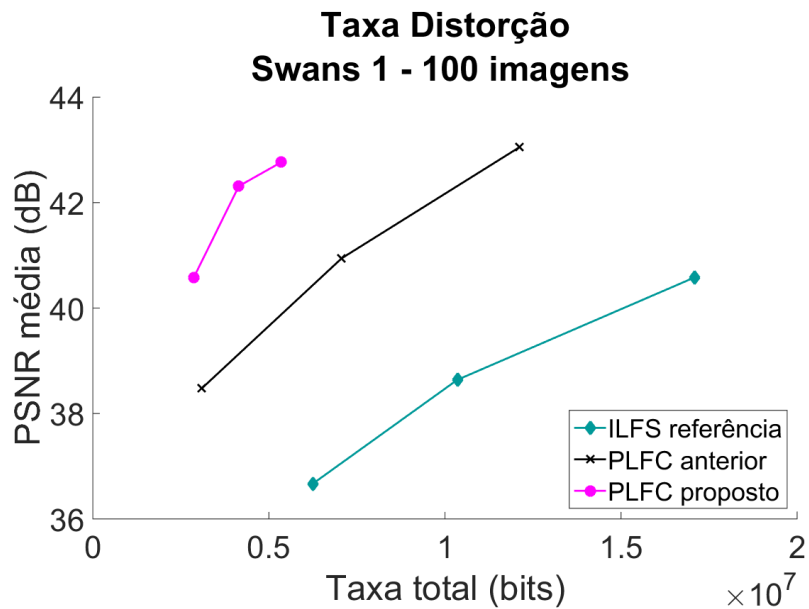


Figura A.27: Resultado taxa distorção referente ao LF Swans 1.

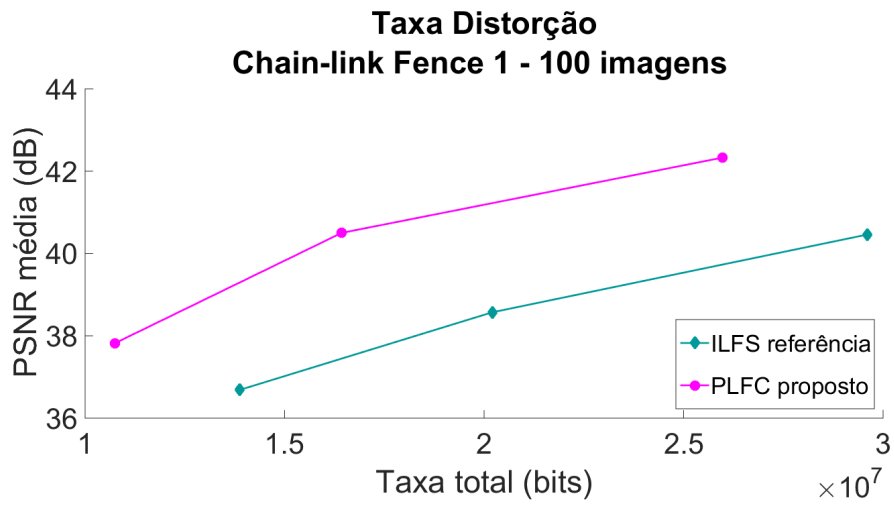


Figura A.28: Resultado taxa distorção referente ao LF Chain-link Fence 1.

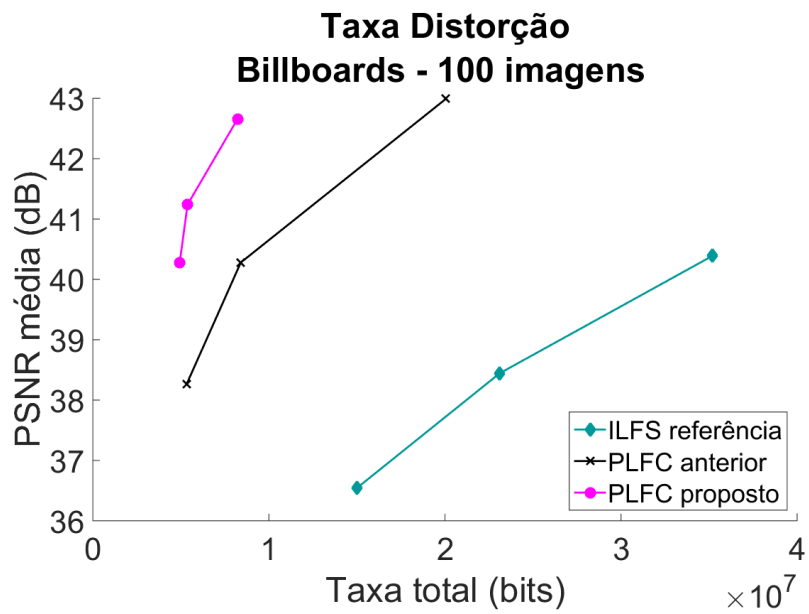


Figura A.29: Resultado taxa distorção referente ao LF Billboards.

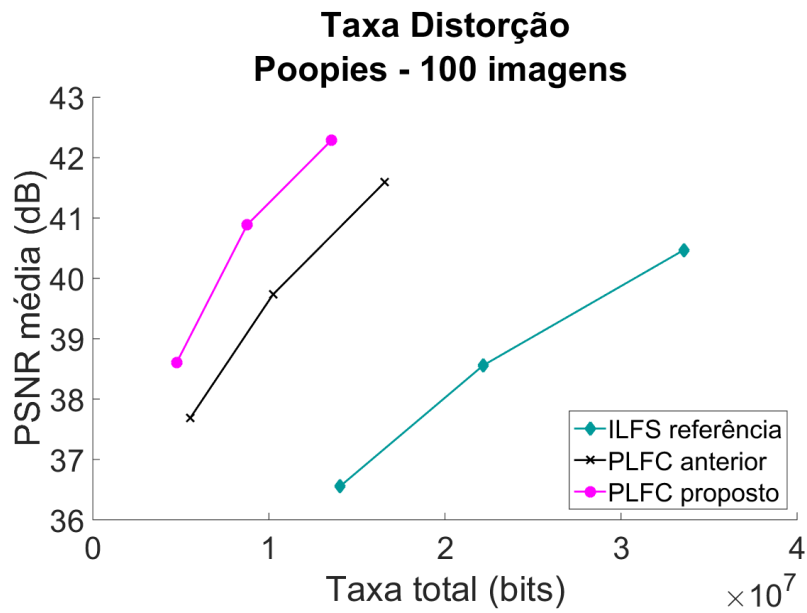


Figura A.30: Resultado taxa distorção referente ao LF Poopies.

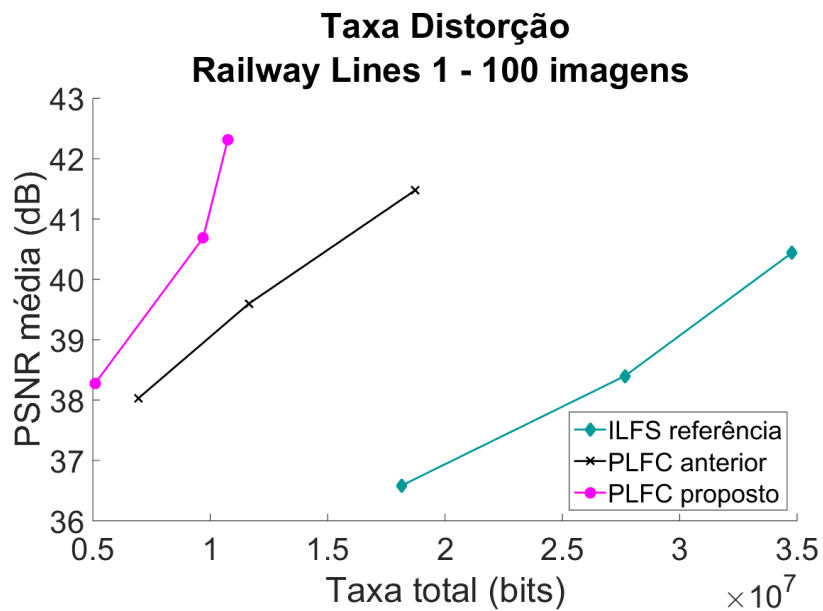


Figura A.31: Resultado taxa distorção referente ao LF Railway Lines 1.

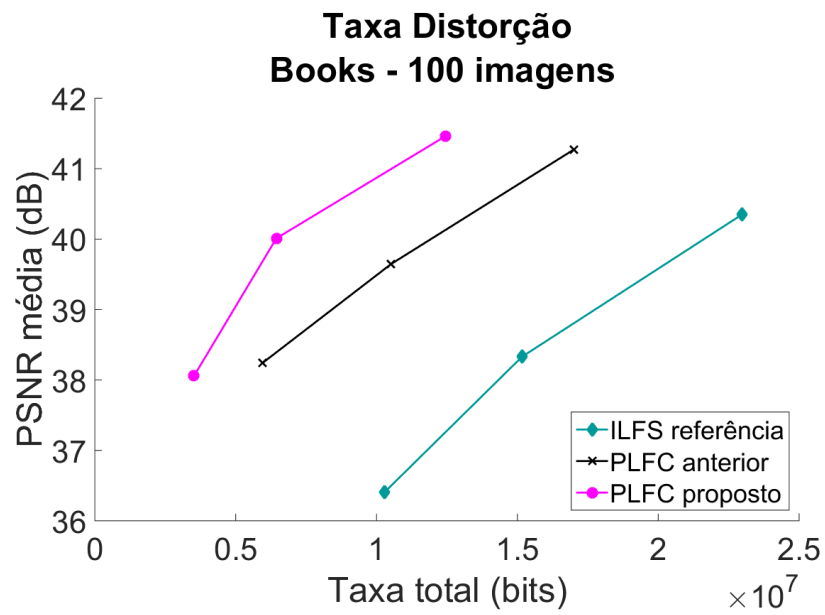


Figura A.32: Resultado taxa distorção referente ao LF Books.