

Multi-Frame Super-Resolução a Partir de uma Única Imagem por Estereoscopia

Felippe L. Nacif, Leandro A. F. Fernandes
Instituto de Computação, Universidade Federal Fluminense (UFF)
CEP 24210-346, Niterói, RJ, Brasil
Email: felippenacif@id.uff.br, laffernandes@ic.uff.br

Resumo—Super-resolução (SR) é um conjunto de técnicas que busca estimar uma imagem de alta resolução (HR) a partir de uma ou mais imagens de baixa resolução (LR). SR com o uso de uma imagem normalmente depende de métodos de aprendizado para melhorar uma LR. SR por múltiplas imagens, por outro lado, pode não assumir nenhum conhecimento prévio da cena e utilizar informação extraída das várias imagens LR capturadas por pontos de vistas diferentes para estimar uma imagem SR colorida. Neste trabalho, nós apresentamos uma abordagem de múltiplas imagens que utiliza uma lente estereoscópica para capturar duas imagens sobrepostas da cena ao mesmo tempo. Nós adaptamos o algoritmo clássico *Iterative Back Projection (IBP)* para estimar uma imagem HR colorida a partir de duas imagens LR coloridas. Nós modelamos virtualmente a lente proposta nesse artigo e demonstramos que utilizando apenas duas imagens sobrepostas, nossa técnica atinge resultados qualitativamente melhores que outras técnicas de SR por múltiplas imagens.

Palavras-chave: super-resolução; lentes estereoscópicas; iterative back projection

Abstract—Super-resolution (SR) is a set of techniques that estimates a high-resolution (HR) image from one or more low-resolution (LR) images. Single-frame SR usually relies on learn-based methods to enhance a LR image. Multi-frame SR, on the other hand, may not assume previous knowledge of the scene and use information extracted from multiple LR images taken from slightly different viewpoints to estimate a HR view. We present a multi-frame SR approach that uses a stereoscopic camera lens attachment to capture two overlapped views of the scene at the same time with the same image sensor. We adapt the classic *Iterative Back Projection (IBP)* SR algorithm to estimate a SR color image from two color LR views. We virtually modeled the proposed lens attachment and demonstrate that, by using only two overlapped views, our approach achieves qualitatively better results than other multi-frame SR techniques.

Keywords: super-resolution; stereoscopic lens attachment; multi-frame; iterative back projection

I. INTRODUÇÃO

A resolução de uma imagem depende principalmente do *hardware* da câmera que faz a captura. Existem duas maneiras de melhorar a qualidade de captura de uma imagem através do *hardware*. Uma é aumentando o tamanho do sensor da câmera, outra maneira seria reduzindo o tamanho de cada pixel no sensor, mas essa abordagem causa outros problemas devido à limitações físicas. De qualquer forma, a melhora no hardware leva naturalmente a equipamentos mais caros.

Nós propomos neste trabalho uma técnica de Super resolução (SR) que faz uso de uma lente que captura, ao mesmo tempo, duas visões da cena de forma sobreposta no mesmo sensor. Para isso, nós armazenamos em cada pixel da

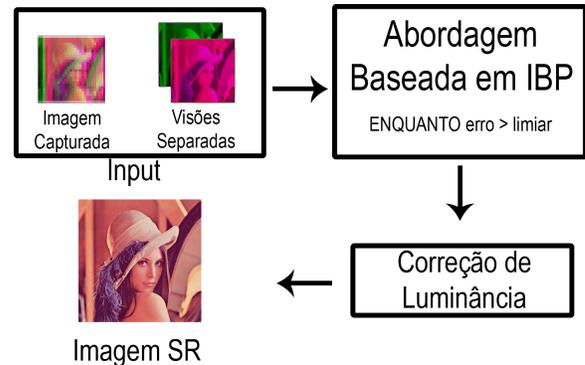


Fig. 1. Diagrama do processo proposto

imagem ambas as visões da cena utilizando cores ortogonais no espaço de cores RGB, permitindo que separemos essas visões. Posteriormente nós executamos uma versão modificada do método de SR conhecido como *Iterative Back Projection (IBP)* [1], que iterativamente reduz o erro de projeção da imagem de alta resolução estimada para a imagem de baixa resolução, em relação as versões em baixa resolução originais.

A Figura 1 ilustra os passos propostos na nossa abordagem. Uma adaptação de lente com dois filtros, um verde e um violeta (Seção III-A) capturam duas visões levemente diferentes da cena alvo. Após isso nós modificamos o algoritmo de IBP atribuindo peso para cada uma das visões da cena (Seção III-C). Por fim, nós executamos um balanceamento da luminância na imagem final em relação ao par de imagens de entrada, para que assim ambas as imagens tenham a mesma luminância média (Seção III-D).

As principais contribuições desse trabalho são:

- A proposta de uma técnica de SR, que faz uso de uma lente estereoscópica com filtro de cores, que captura duas versões simultâneas da cena usando o mesmo sensor.
- Uma modificação do algoritmo IBP existente, para funcionar com as imagens de cores ortogonais capturadas pela lente proposta.

Nós realizamos experimentos de validação tanto sobre um conjunto de imagens sintéticas (Seção V), quanto em imagens que simulam o efeito esperado das nossas lentes. Nossos resultados foram avaliados quantitativamente utilizando as métricas PSNR e SSIM, e qualitativamente. Nossa técnica

atingiu resultados competitivos mesmo utilizando apenas duas visões capturadas em uma única exposição da câmera à cena.

A. Trabalhos relacionados

A técnica de SR por múltiplas imagens foi introduzida em [2] e algumas abordagens diferentes foram propostas desde então. Irani e Peleg [1] é um exemplo de algoritmo IBP utilizado em um conjunto de imagens de baixa resolução (LR) com movimento entre as imagens desconhecido.

Pelo que sabemos, a primeira tentativa de capturar duas visões de uma cena ao mesmo tempo foi proposta por Nene e Nayar [3]. Eles utilizaram um conjunto de espelhos para obter um par de imagens estéreas, porém cada imagem ocupava apenas metade do sensor. Outra abordagem similar para a captura das imagens foi proposta por Bando et al. [4], porém o objetivo desse trabalho era recuperar informações de profundidade.

B. Visão geral da técnica proposta

Nossa técnica objetiva reconstruir uma imagem de alta resolução (HR) a partir de um par de imagens LR. O par de imagens é capturado utilizando duas cores ortogonais, permitindo que nós as separemos facilmente. Para isso, nós desenvolvemos um aparato composto por uma lente estereoscópica, um espelho e um semi-espelho. Esse aparato é acoplado a câmera e captura duas vistas da cena, registrando uma no canal verde e outra nos canais vermelho e azul, tendo então duas projeções diferentes da cena no mesmo sensor.

Para registrar as imagens nós primeiro utilizamos o algoritmo de *minimum eigenvalue* para encontrar pontos de interesse, então extraímos as características e registramos as imagens utilizando descritores SURF. No processo iterativo baseado em IBP, nós melhoramos a estimativa da imagem HR através de uma medida de erro entre as imagens LR e a estimativa atual, ponderadas pela importância do canal de cor para a percepção humana. Quando o processo iterativo converge, nós executamos uma equalização da luminância. A Figura 1 mostra uma visão geral da técnica.

II. TÉCNICA IBP

Algumas técnicas de super-resolução utilizam o seguinte algoritmo: Um conjunto de imagens de baixa resolução g são capturadas pela mesma câmera durante um período de tempo e são registradas em relação a uma imagem de referência g_0 , depois um processo iterativo tenta minimizar uma função de erro e encontrar a imagem de alta resolução f que através de um modelo de degradação, melhor explica as imagens de entrada. Normalmente esse modelo é usado como:

$$g_i = D(B_i(W_i(f))), \quad (1)$$

onde g_i é a i -ésima imagem LR, f é a imagem HR objetivo, W_i é a transformação geométrica que a i -ésima imagem sofreu, B_i é o operador de *blur* da i -ésima imagem, normalmente assumido como o PSF da câmera, e D é a operação de redução por um fator de escala.

Esse processo iterativo utiliza uma estimativa inicial f_0 , e em cada iteração calcula o erro entre a imagem HR atual \hat{f} após o modelo de degradação e as imagens LR originais g_i .

III. ABORDAGEM PROPOSTA

A abordagem proposta é composta por quatro passos principais. Primeiro os pontos de vista sobrepostos são capturados utilizando nosso aparato (Seção III-A). Então as imagens LR são separadas e registradas através de uma transformação projetiva (Seção III-B). Quando as imagens estiverem alinhadas pode-se executar o algoritmo IBP modificado proposto (Seção III-C). A imagem final do processo iterativo não possui tanto contraste quanto o esperado, e é por isso que executamos o balanceamento da luminância (Seção III-D) para corrigir a escala de luminância da imagem final em relação às de entrada.

A. Lentes estereoscópicas

Nós modelamos nosso aparato da seguinte forma: um espelho e um semi-espelho foram posicionados lado a lado a 45° em frente a um filtro de cor cada. O semi-espelho posicionado em frente ao sensor da câmera permite que a imagem de ambos os filtros sejam projetadas na lente. Nós escolhemos em separar o canal verde dos demais pois as câmeras normalmente são desenvolvidas seguindo o padrão de Bayer que possui duas vezes mais sensores para a esta cor do que para as outras.

Utilizar duas visões da cena armazenadas em diferentes canais de cor pode ser interpretado como a cena capturada por dois sensores diferentes. Zomet e Peleg [5] estudaram essa abordagem de SR de multi sensores e declararam que duas visões de uma cena capturadas por diferentes sensores, quando alinhadas, estão estatisticamente relacionadas.

B. Alinhamento e transformação

A lente proposta induz uma transformação entre as imagens capturadas. Algoritmos de registro bem conhecidos em SR [6], [7] falharam em alguns dos nossos casos de testes, provavelmente por causa da diferença de luminâncias nas imagens de entrada. Para resolver isso, nós estimamos a transformação projetiva entre as imagens e buscamos pontos de interesse por *minimum eigenvalue*. Para extrair as características da imagem nós utilizamos o descritor SURF. Então realizamos o casamento dos pontos de interesse e calculamos a transformação que melhor os explica. Um alinhamento sub-pixel pode ser realizado a fim de melhorar o registro das imagens.

C. Super resolução

Nossa abordagem é baseada na técnica proposta por Irani e Peleg [1]. O algoritmo começa extraindo a luminância de ambas as imagens de entrada, e então as alinha. Nossa estimativa inicial e cada versão melhorada subsequente simula o processo de degradação de (1) para obter as imagens LR. Assumindo que o nosso modelo de degradação é perfeito, se f_0 é a imagem HR correta, então a diferença entre as imagens LR simuladas e as imagens originais de entrada seria zero. Como isso não é verdade, qualquer solução terá um erro, nosso objetivo é encontrar a solução que minimize esse erro.

O erro entre a imagem LR estimada a partir de HR após o modelo (1) e as imagens de entrada pode ser calculada utilizando:

$$e = \sum_{x,y,i} ((g_i(x,y) - \hat{g}_i(x,y))p_i)^2, \quad (2)$$

onde i é o índice da imagem LR, g é a imagem de entrada, \hat{g} é a imagem simulada, e p_i é o peso associado a cada uma das visões para a imagem final.

Experimentalmente observamos que os valores $p_1 = 0.57$ para a imagem originalmente verde e $p_2 = 0.43$ para a violeta produzem o melhor resultado na maioria dos casos.

Em cada iteração o processo atualiza a estimativa da imagem HR em um pixel $x = (x, y)$ da seguinte forma:

$$f'^n(x) = f'^{n-1}(x) + \sum_{y,i} (g_i(y) - g_i^n(y)) p_i \frac{(H_{bp}(x))^2}{c \sum_y (H_{bp}(x))}, \quad (3)$$

onde \hat{f}^n é a estimativa de HR após n iterações, \hat{g}_i^n é o i -ésima visão simulada da n -ésima iteração usando (4), H_{bp} é $\hat{f}^{(n-1)}$ convoluído com o núcleo de *back-projection* (Seção V), c é a constante para ajustar o tamanho do passo, x é um pixel de alta resolução e y é um pixel de baixa resolução influenciado por x , isto é, um pixel em g que está no raio de ação de x .

As versões de baixa resolução da estimativa de alta resolução f^n é derivada utilizando a seguinte expressão:

$$\hat{g}_i^n(y) = p_i \sum_x (f^n(x) H_{psf}(x)), \quad (4)$$

onde H_{psf} é \hat{f}^{n-1} convoluído com o núcleo de PSF da câmera ou uma melhor aproximação para o *blur* presente no processo de captura. Veja a seção V para mais informações.

O algoritmo primeiro extrai as luminâncias das imagens de entrada. O processo iterativo começa com uma estimativa da imagem HR e a cada passo atualiza essa estimativa utilizando (3), projeta as duas imagens com o modelo de degradação (4) e calcula o erro usando (2). As iterações terminam quando um dos seguintes critérios de parada é atingido: a variação do erro é pequena o suficiente, ou um limite de iterações definido tenha sido atingido. O processo é executado na componente Y da imagem no espaço de cores YIQ, pois nossa separação dos canais de cor não nos possibilita aproveitar informação redundante nas componentes em RGB. Para a imagem final as componentes IQ são interpoladas.

D. Correção de luminância

Devido ao fato de estarmos fundindo informação de canais de cores diferentes, nossa imagem final é normalmente mais escura do que o desejado. Assim, nós melhoramos o resultado, realizando um balanceamento da luminância da imagem final. Nós assumimos que a imagem final desejada deve ter a mesma luminância média que a de entrada. Para isso nós escalamos as intensidades da imagem de saída proporcionalmente. Este procedimento melhorou a imagem final.

IV. IMPLEMENTAÇÃO

Nós implementamos a abordagem proposta utilizando MATLAB [®] devido a extensa quantidade de ferramentas disponíveis. Para estimar o *blur* que utilizamos como o PSF do algoritmo, nós utilizamos a implementação disponibilizada por Goldstein et al. [8].

Desenvolvemos o protótipo em 3D das lentes propostas neste trabalho utilizando o Blender para testar sua viabilidade. Nós também geramos algumas imagens de teste nessa ferramenta utilizando a configuração de *ray-tracing* disponível.

V. EXPERIMENTOS E RESULTADOS

Testamos nossa técnica sob duas condições diferentes. Em uma nós renderizamos a imagem usando nossas lentes modeladas no Blender usando *ray-tracing*, no segundo nós geramos dois conjuntos de imagens usando o processo de captura que nossas lentes seguiriam em condições perfeitas. Para os demais algoritmos presentes na Tabela I, as imagens LR continham todos os canais de cor. Isso permitiu que comparássemos os resultados de maneira correta, visto que os demais algoritmos não foram projetados para receber imagens de canais de cor diferentes. Depois disso, nós aplicamos um filtro gaussiano com desvio padrão igual a 1, e reduzimos a imagem por um fator de 2 para obter as imagens de entrada.

Nos nossos testes o núcleo de PSF (H_{psf}) foi estimado usando a implementação de Goldstein et al. [8], o núcleo de *back projection* (H_{bp}) foi definido como a raiz quadrada do (H_{psf}) e a constante c foi um valor entre 0.1 e 1.8 dependendo do conjunto de dados. O critério de parada foi definido como variação do erro menor que 0.1 ou 30 iterações. As implementações utilizadas dos demais algoritmos de SR foram disponibilizados por Vandewalle et al. [7].

Nós validamos nossa técnica ao executar o algoritmo apresentado em um conjunto de imagens. Resumimos nossos resultados numéricos na Tabela I com as métricas PSNR e SSIM. Na Figura 2 comparamos nosso resultado em um caso de teste renderizado com o melhor resultado numérico apresentado na Tabela I.

Como pode ser visto na Figura 2c, as bordas no 'S' e na face do personagem estão melhor definidas do que as do melhor resultado numérico, apresentado na Figura 2b. Enquanto isso, é possível notar que as luminâncias no resultado dessa figura são mais semelhantes aos apresentados no gabarito (Figura 2a) do que na nossa proposta, o que explica melhores resultados numéricos obtidos por este algoritmo [9].

Os testes que simulam o comportamento esperado de nossas lentes (Lena e Eia), foram realizados para descobrir como os canais de cores poderiam realmente melhorar a qualidade da imagem. Portanto, é possível ver na Tabela I que o conjunto de dados Lena, originalmente colorido, teve um melhor desempenho comparando com o conjunto de dados Eia, que é originalmente em tons de cinza. Assim, é possível concluir a partir desses resultados que as informações complementares presentes nos diferentes canais de cores contribuíram de fato para o aprimoramento da imagem.

O PSNR é uma métrica geralmente usada para comparar técnicas de super-resolução. Entretanto é importante comentar que mesmo os casos em que nossos resultados são numericamente inferiores as outras técnicas, é possível observar, como na Figura 2, que a qualidade do nosso resultado é melhor do que outros algoritmos de SR testados, tendo menos *blur*, melhores bordas e menos serrilhamento.

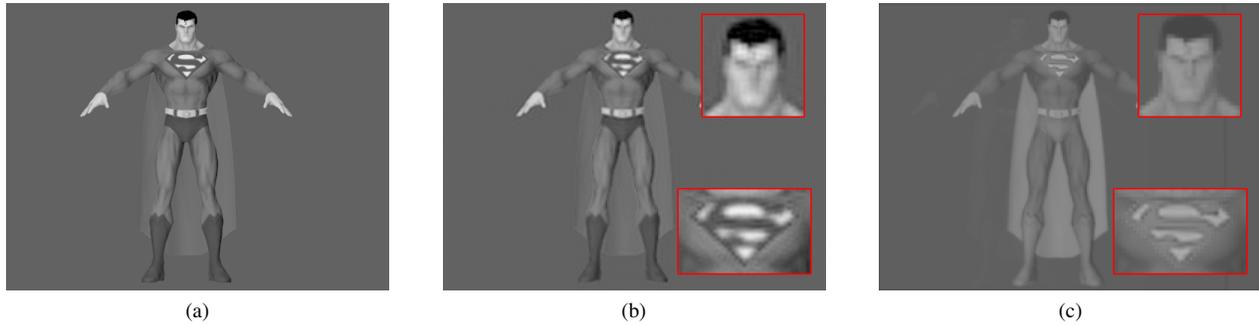


Fig. 2. A imagem de alta resolução do super-homem (a) comparada com os resultados obtidos usando o algoritmo [9] (b) e a nossa proposta (c). As partes destacadas mostram que nossa abordagem produz arestas melhores, mas o contraste da imagem produzida por [9] é maior.

Tabela I

CONJUNTO DE IMAGENS UTILIZADAS E OS RESULTADOS NUMÉRICOS COM PSNR E SSIM DA TÉCNICA APRESENTADA COMPARADA COM OUTROS CONHECIDAS (PSNR(dB) / SSIM). A ANOTAÇÃO (S) DENOTA OS TESTES FEITOS NAS IMAGENS GERADAS ATRAVÉS DO PROCESSO DE SIMULAÇÃO.

	Superman	Can	Lena (S)	Eia (S)
[9]	36.05 / 0.97	24.54 / 0.56	11.81 / 0.50	11.71 / 0.69
[1]	35.57 / 0.97	25.61 / 0.89	11.46 / 0.54	10.15 / 0.63
[10]	35.90 / 0.97	25.81 / 0.89	11.49 / 0.56	10.16 / 0.62
[11]	32.64 / 0.94	24.42 / 0.56	11.75 / 0.48	11.25 / 0.64
[12]	35.98 / 0.97	23.91 / 0.79	11.86 / 0.60	10.10 / 0.54
Our	26.43 / 0.93	24.02 / 0.88	26.10 / 0.86	18.24 / 0.86

A. Limitações

O aparato proposto faz com que parte da imagem que aparece em uma visão não apareça na outra, principalmente em fotos tiradas de perto, devido ao campo de visão da câmera, ou por oclusão por paralaxe. Além disso, nossa abordagem impõe limitações à fidelidade de cores.

VI. CONCLUSÃO

Nós apresentamos neste artigo uma modificação no algoritmo de IBP para super-resolução para tirar vantagem das diferentes informações presentes nos diferentes canais de cor. Nós também apresentamos uma lente especial que pode ser projetada para se encaixar em qualquer câmera colorida para capturar duas visões da cena em apenas uma exposição;

Comparamos nossos resultados de forma quantitativa e qualitativa com outros algoritmos comuns de SR, e verificamos que apesar de nossos resultados numéricos serem piores que os apresentados por técnicas conhecidas em alguns casos, esse comportamento era esperado devido a técnica proposta não preservar a qualidade da cor como outras. Entretanto, nós mostramos qualitativamente que mesmo utilizando apenas duas imagens, nossos resultados têm menos serrilhamento e arestas mais fiéis às da imagem original que as outras técnicas comparadas presentes na literatura.

AGRADECIMENTOS

Felippe L. Nacif recebeu bolsa de Iniciação Científica da FAPERJ (Processo E-26/202.510/2015). O projeto fez

uso de recursos de projeto JNCE da FAPERJ (Processo E-26/202.832/2015).

REFERÊNCIAS

- [1] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical models and image processing*, vol. 53, no. 3, pp. 231–239, 1991.
- [2] T. S. Huang and R. Y. Tsay, "Multiple frame image restoration and registration," in *Advances in Computer Vision and Image Processing*, vol. 1. Greenwich: JAI, 1984, pp. 317–339.
- [3] S. A. Nene and S. K. Nayar, "Stereo with mirrors," in *Computer Vision, 1998. Sixth International Conference on*. IEEE, 1998, pp. 1087–1094.
- [4] Y. Bando, B.-Y. Chen, and T. Nishita, "Extracting depth and matte using a color-filtered aperture," in *ACM Transactions on Graphics (TOG)*, vol. 27, no. 5. ACM, 2008, p. 134.
- [5] A. Zomet and S. Peleg, "Multi-sensor super-resolution," in *Applications of Computer Vision, 2002.(WACV 2002). Proceedings. Sixth IEEE Workshop on*. IEEE, 2002, pp. 27–31.
- [6] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using sub-pixel displacements," in *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR'88., Computer Society Conference on*. IEEE, 1988, pp. 742–746.
- [7] P. Vandewalle, S. Süsstrunk, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution," *EURASIP Journal on applied signal processing*, vol. 2006, pp. 233–233, 2006.
- [8] A. Goldstein and R. Fattal, "Blur-kernel estimation from spectral irregularities," in *European Conference on Computer Vision (ECCV)*. Springer, 2012, pp. 622–635.
- [9] A. Papoulis, "Generalized sampling expansion," *IEEE transactions on circuits and systems*, vol. 24, no. 11, pp. 652–654, 1977.
- [10] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–I.
- [11] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1064–1076, 1997.
- [12] T. Q. Pham, L. J. Van Vliet, and K. Schutte, "Robust fusion of irregularly sampled data using adaptive normalized convolution," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, p. 083268, 2006.