# A method of generating artificial 2D images with different depth of fields

Fabio Pantano de Luca[1,2] and Carlos Eduardo Thomaz[1]
[1]Department of Eletrical Engineering, FEI, São Paulo, Brazil
[2]Mercedes-Benz do Brasil, São Paulo, Brazil

***Abstract*—Robust segmentation of objects in images is still a challenge. Recent methods of segmentation based on two images have been proposed and shown promising results. However, due to the novelty of these methods, image databases based on pair of images with different depth of fields are still not available. One way to tackle this lack of images for testing is to generate artificial ones, simulating all the aspects of the high and low depth of fields, as required by the segmentation algorithms. In this work we present a method of generating such pair of images, which brings the benefits of artificially building a large and extensive database as well as being able to automatically generate precisely the ground truth image, required for the comparative evaluation of the segmentation.**

***Index Terms*—low depth of field; artificial images; depth of field simulation**

## I. Introdução

A inovação no uso de duas imagens com profundidades de campos desiguais pode gerar resultados de acurácia de segmentação promissores se comparados com os métodos tradicionais [1], [2]. No entanto, há ainda uma falta de bancos de imagens disponíveis na literatura para uma avaliação mais detalhada do método e obtenção de resultados mais extensivos. A criação de fotografias reais aos pares pode ser trabalhosa para um número elevado de imagens, e assim, pouco eficiente, ainda mais quando se almeja uma avaliação numa etapa preliminar. Ainda, como se trata de uma investigação comparativa de parâmetros de operação de algoritmos, há necessidade de que mais versões de imagens controladas estejam disponíveis para avaliação do desempenho do método frente à abertura ótica. Adicionalmente, pode ser uma tarefa excessivamente árdua criar variações possíveis ou estatisticamente relevantes de composições de cenas possíveis. Na maioria das situações estudadas na prática, imagens reais precisam ser segmentadas por seres humanos para criação do *ground truth*, o que é uma tarefa demorada, e que se multiplica quando diversas versões de profundidade de campo existem.

Com base nessas limitações pelo uso de imagens reais, descrevemos e implementamos um método de criação de imagens artificiais em pares (ou múltiplos), valendo-se de simplificações conceituais, mas que produzem pouco prejuízo nas questões pertinentes aos possíveis algoritmos de segmentação aos quais serão submetidas. A vantagem dessa metodologia é que algumas limitações supramencionadas são anuladas, tais como: rápida criação de um número grande de imagens, representatividade de diversas variações de composição e criação automática de um *ground truth* preciso. A diferença das demais técnicas conhecidas [3] é que não há necessidade de um conhecimento detalhado da cena artificial, tal como dados precisos das formas e posições dos objetos. Também diverge por não ser uma técnica de simples introdução de desfoque numa cena já estabelecida.

O artigo está organizado conforme segue: na próxima seção são apresentados os princípios óticos e matemáticos em que a criação de imagens artificiais está baseada. Na sequencia o método de geração de imagens é detalhado. Na seção seguinte é apresentada uma forma de avaliação do método via comparação com imagens reais. Por fim, há as conclusões referentes aos resultados obtidos.

## II. Fundamentação

### A. Princípios óticos

A profundidade de campo é uma limitação de todo sistema ótico, descrita, de forma simplificada, como a região dentro de uma cena em que os objetos possuem foco dentro do limite aceitável para a aplicação submetida. Na cena, apenas um plano possui todos seus pontos em foco, e a medida que objetos se afastam de tal plano, tanto no sentido de se aproximarem da lente, quanto de se afastarem, o desfoque aumenta gradualmente. A análise desse comportamento pode ser feita pela relação que um ponto-objeto é representado como um ponto-imagem sobre o sensor (considerada uma câmera digital, sem prejuízos de conceituação) apenas se estiver sobre o plano focal. Deslocado de tal plano, o ponto-objeto tem uma imagem como uma mancha desfocada, normalmente caracterizada como circular. Essa imagem é usualmente denominada de círculo de confusão.

A base da criação de imagens artificiais que simulem as características de profundidade de campo de um sistema ótico é o entendimento da caracterização do círculo de confusão, ou seja, o comportamento do desfoque dos pontos-objetos (e consequentemente do objeto como um todo) em cena em função de sua distância ao sistema ótico, bem como de demais características do mesmo sistema (como a abertura ótica, por exemplo). A medida do diâmetro do círculo de confusão é oriunda da seguinte equação [4]

$$c = \frac{f^2}{N.(S_{1O} - f)} . \left| 1 - \frac{S_{1O}}{S_{1NF}} \right| , \qquad (1)$$

onde $f$ é a distância focal da lente usada para capturar as imagens, $N$ é a abertura focal da lente, $S_{1O}$ a distância do

plano de foco à lente e $S_{1NF}$ é a distância do objeto em cena à lente.

A segunda característica para simulação da cena trata-se da representação do tamanho do objeto dentro da imagem, relativa ao tamanho total da imagem, representada por $i_{Rel}$ e determinada por

$$i_{Rel} = \frac{o.f}{h_S}.\left(\frac{1}{S_{1O} - f}\right) , \quad (2)$$

onde $h_S$ é o tamanho vertical do sensor fotográfico e o parâmetro $o$ o tamanho vertical do objeto real. Vale ressaltar que tanto (1) quanto (2) são aproximações do uso da ótica geométrica e paraxial, e, portanto, contemplam casos em que o objeto não esteja a uma distância exageradamente próxima à lente, particularmente mantendo-se dentro da região aproximadamente linear do gráfico apresentado na Fig. 1, no qual consta um exemplo para lente de 50 mm de distancia focal e sensor de 35 mm (24 mm de altura) [5].

### *B. Caracterização do desfoque*

Apesar da equação (1) já apresentar um valor de dimensão do círculo de confusão, é necessário ainda caracterizar sua distribuição de energia. A função de distribuição é denominada de *point spread function* (PSF) e é bem definida matematicamente [6]. No entanto, há uma complexidade [7], não só pela função em si, como pela realidade dos objetos em cena, que possui cada um de seus pontos-objetos dentro de um determinado plano. A simplificação da abordagem será feita em duas frentes: primeiramente a PSF será aproximada por uma distribuição gaussiana como adotado em [8], [9], dada pela função

$$PSF = \frac{2}{\pi\left(g.b\right)^2}.\exp\left(-\frac{2.r^2}{\left(g.b\right)^2}\right) , \quad (3)$$

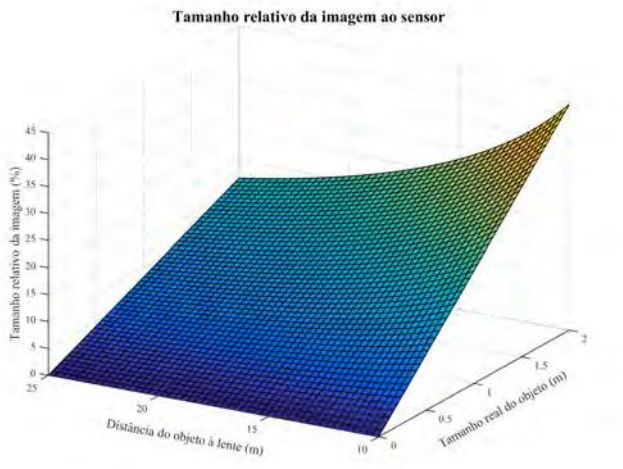


Figura 1: Exemplo de mapa de tamanho relativo de imagem ao sensor com distancia focal da lente de 50 mm e sensor de 35 mm (24 mm de altura) [5].

onde $r$ é o raio do ponto dentro do círculo de confusão, $g$ é uma constante e $b$ é dado pela função

$$b = \frac{1}{N}.\left|f - S_{2O}\right| ,$$

em que $S_{2O}$ é a distância entre a lente e o plano em que se forma a imagem.

A segunda simplificação aplicada é discretizar os planos dos objetos. Assim, cada objeto estará totalmente contido num único plano, ao invés de possuir um volume complexo, o que exigiria o conhecimento de todos os pontos do mesmo. Essa simplificação pode ser considerada menos prejudicial nas bordas do objeto, justamente na parte em que o algoritmo de segmentação ao qual a imagem será submetida apresenta os maiores desafios.

## III. Metodologia

A geração das imagens artificiais consiste na sobreposição de diversos elementos de cena sobre uma imagem de fundo. O número de elementos em cada imagem é aleatório, sendo definido pelo algoritmo de construção, assim como a posição de cada elemento em cena e a distância a que cada objeto real estará da lente. A aleatoriedade da formação da imagem é parte importante em seu conceito pois é do interesse que se explore o maior número de variações possíveis nas imagens geradas, vista a capacidade do algoritmo em gerar uma grande quantidade de amostras.

A preferência de escolha foi por objetos que possuem um tamanho real limitado para representação em cena, especialmente para não violar restrições das simplificações previamente impostas. Além disso, há um cuidado na escolha dos elementos e fundos para que sejam geradas preferencialmente imagens com uma plausibilidade contextual. Essa plausibilidade não é necessária do ponto de vista do algoritmo de segmentação, visto que não há uma avaliação da semântica visual da imagem, mas é interessante para os primeiros experimentos, para que seres humanos possam avaliá-las sem prejuízos. Exemplos de elementos e fundos estão representados nas Fig. 2 e Fig. 3. Para a operação é necessário que haja previamente disponível um repertorio de elementos segmentados, com fundo transparente (como exibido na Fig. 2), além de informação da altura real do objeto (tamanho na dimensão vertical) e um segundo conjunto de imagens de fundos. Ainda, para cada composição, realiza-se a mesma construção alterando-se o parâmetro de abertura ótica $N_A$ vezes. O algoritmo de geração é descrito brevemente a seguir:

- Sorteio de parâmetros de cena: Número de imagens (sendo no mínimo um); definição aleatória das imagens; posição de cada elemento em cena, sendo possível que os mesmos se localizem parcialmente fora da mesma; distância do elemento à lente simulada;
- Ordenação dos elementos para que o processo seja feito do elemento mais afastado em direção ao mais próximo;
- Ajuste do tamanho de cada elemento conforme (2);
- Realização do deslocamento do objeto em cena para cada elemento;

- Cálculo do desfoque e aplicação do mesmo conforme (1), e empregando convolução de uma função gaussiana conforme (3). Antes de ser feita a aplicação da convolução, no entanto, parte do fundo da imagem que circundaria a mesma é agregado à borda, de modo a ter sua participação no desfoque;
- Mescla das camadas sobrepostas para formação de uma única imagem;
- Criação do *ground truth*, dada um limite do diâmetro do círculo de confusão, para cada possibilidade de parâmetro ótico usado.

Sendo possível que a operação ocorra em *loop* $N$ vezes, então o resultado do algoritmo de geração é um conjunto de $N.N_A$ imagens para o banco de testes, somado aos $N.N_A$ imagens de *ground truth*. A Fig. 4 exemplifica o resultado de uma imagem.

## IV. Resultados Experimentais

### A. Criação de um padrão de comparação

Como validação do método de geração de imagens artificiais, fez-se a comparação direta entre uma imagem real (fotografia feita com todos seus parâmetros controlados) e uma imagem artificial simulando as mesmas condições, repetindo a comparação para diversas aberturas óticas. O aparato montado para criação das imagens reais e artificias está representado na Fig. 5, onde dois objetos, *A* (lente objetiva) e *B* (*timer*), foram colocados em cena com distâncias diferentes para que o grau de desfoque fosse igualmente único. Um terceiro elemento pode ser ainda considerado que é o fundo da cena, por possuir uma distância diferente das demais. Tem-se assim elementos distantes *440 mm* (objeto *A*), *580 mm* (objeto *B*) e *1590 mm* (fundo) da lente da câmera. O foco da lente está sobre o objeto *B*. As imagens foram feitas com 5 aberturas distintas, a saber $f/1.8$, $f/2.5$, $f/5$, $f/10$ e $f/22$, e duas delas, a título de exemplo, estão representadas na Fig. 6. As fotografias foram

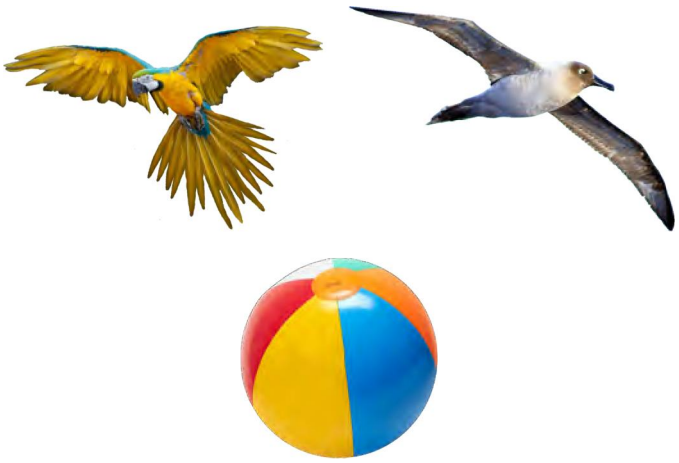

Figura 2: Exemplos de elementos de cena usados na geração de imagens.

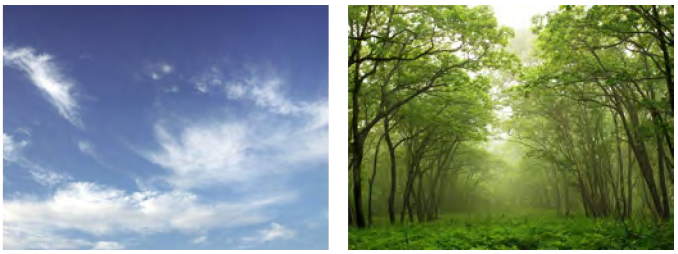

Figura 3: Exemplos de fundos de cena usados na geração de imagens.

feitas em sequência para que as condições de luminosidade fossem as mais próximas possíveis entre si.

Para as imagens artificiais, utilizou-se a fotografia feita com $f/22$ por ser a menor abertura possível para a câmera/lente utilizada nos testes, e aquela que produz, portanto, a maior profundidade de campo (todo o campo fotografado em foco). Assim, segmentou-se manualmente os elementos da cena, separando-os. Uma nova foto, mantendo todos as características de distância e luminosidade, foi feita apenas do fundo. Os elementos e fundo resultantes estão demonstrados na Fig. 7.

Os elementos individualizados mais o fundo foram aplicados ao algoritmo de geração de imagens, no entanto, a etapa de sorteio de elementos e posicionamento foram suprimidos para que a imagem pudesse ser semelhante e comparável a real. Esse processo foi repetido quatro vezes, para as aberturas óticas $f/1.8$, $f/2.5$, $f/5$ e $f/10$. Para $f/22$, por ter sido a base dos elementos, a comparação foi descartada, além de não prover um desfoque sensível, o qual é o interesse da análise. As imagens foram analisadas após conversão para cinza para redução de efeitos de diferença de tonalidade que são irrelevantes ao método, mas que poderiam trazer discrepância nos resultados, além disso, passa-se a trabalhar com uma matriz de dados por imagem, ao contrário das 3 necessárias para cores. A conversão seguiu conforme a função nativa do Matlab

$$P_C = 0,2989.P_R + 0,5870.P_G + 0,1140.P_B$$

onde $P_C$ é a intensidade de um pixel na matriz de cinzas e $P_R$, $P_G$ e $P_B$ as intensidades dos pixels nas matrizes vermelha, verde e azul, respectivamente.

### B. Quantificação dos resultados

As imagens reais e artificiais foram comparadas aos pares de aberturas óticas empregando duas métricas independentes, com os resultados das medições apresentados na Tabela I. Primeiramente calculou-se o valor de *structural similarity index* (índice de similaridade estrutural, $SSIM$) [10] e posteriormente fez-se a mesma análise por correlação matemática $R$. Ambas as métricas foram calculadas em três momentos para cada imagem: imagens completas; ampliação em região de médio contraste e ampliação em região de alto contraste. Essas ampliações são interessantes para que haja avaliação em imagens em que a quantidade de pixels que fazem parte do

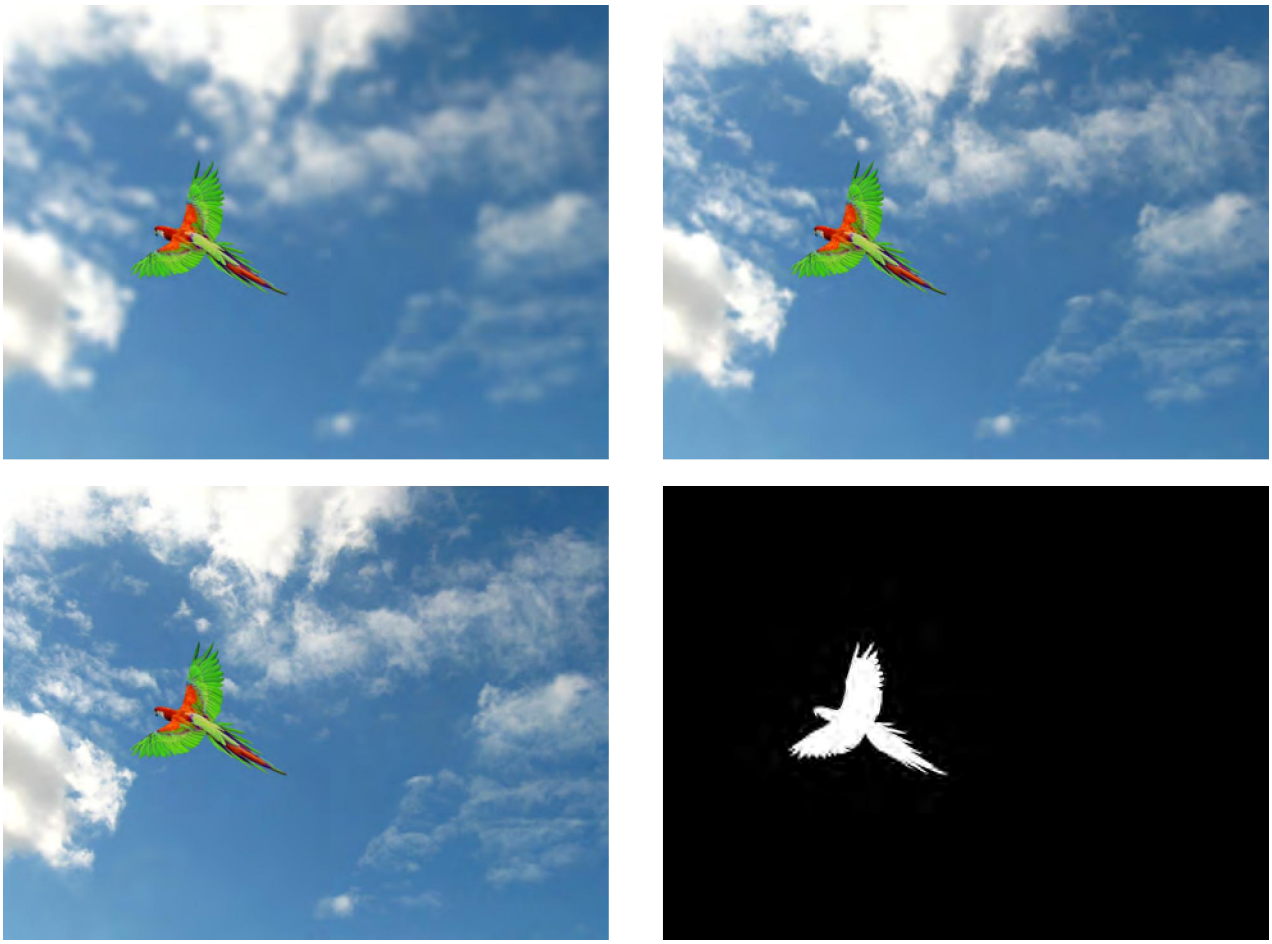

Figura 4: Exemplos de imagens geradas pelo algoritmo. Superior esquerdo: Imagem com $f/1.8$; superior direito: imagem com $f/5.6$; inferior esquerdo: imagem com $f/22$; inferior direito: *ground truth* gerado para a imagem com $f/1.8$.

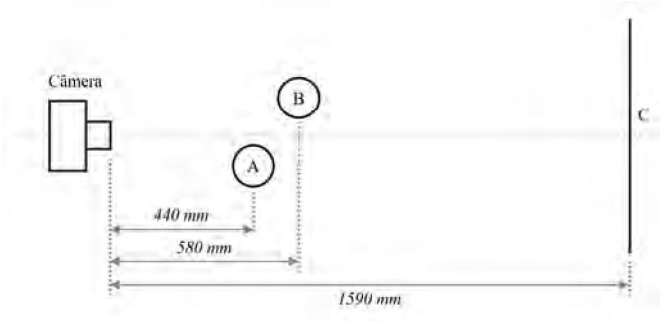


Figura 5: Aparato usado para a construção de imagens reais e artificiais comparáveis entre si. Foco da lente sobre o objeto *B*.

desfoque seja maior. Os resultados visuais estão presentes na Fig. 8 para a imagem completa (representada em cores, para uma melhor avaliação por seres humanos). As Fig. 9 e Fig. 10 mostram analogamente os resultados visuais para a ampliação de médio contraste e de alto contraste.

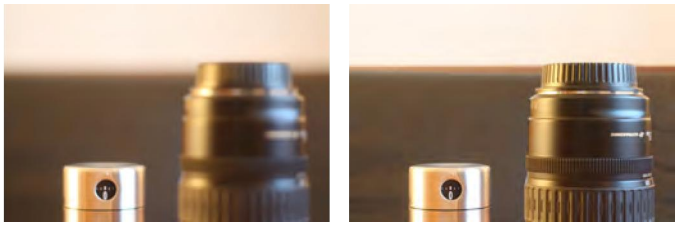

Figura 6: Exemplo de imagem real do mesmo cenário-aparato com aberturas óticas de $f/1.8$ (esquerda) e $f/10$ (direita).

## V. Conclusões

Os valores de $R$ e $SSIM$ são altos para todas as imagens, ficando acima de 99% para a correlação e acima de 97% para o $SSIM$. Esses resultados são satisfatórios, mas poderiam ser melhorados se fosse realizada a equalização de luminosidade entre imagens. Também há possível melhoria na extração manual (recorte) dos elementos da imagem de abertura $f/22$ que foram empregados na formação das imagens artificiais. Como pode ser visto pelos valores resultantes, há uma relação

Figura 7: Elementos da cena segmentados a partir da imagem de profundidade de abertura ótica $f/22$. A direita o fundo da cena também com abertura ótica $f/22$.

Tabela I: RESULTADOS QUALITATIVOS

| **Imagem** | | **Valores medidos** | |
|---|---|---|---|
| ***Dercrição da imagem*** | ***Abertura ótica*** | ***SSIM*** | ***R*** |
| Imagem completa (cinza) | $f/1.8$ | 98,091% | 99,243% |
| | $f/2.5$ | 98,094% | 99,485% |
| | $f/5$ | 97,967% | 99,734% |
| | $f/10$ | 97,433% | 99,803% |
| Imagem de contraste médio (cinza) | $f/1.8$ | 99,029% | 99,702% |
| | $f/2.5$ | 98,453% | 99,678% |
| | $f/5$ | 97,719% | 99,738% |
| | $f/10$ | 95,383% | 99,415% |
| Imagem de contraste alto (cinza) | $f/1.8$ | 97,954% | 99,694% |
| | $f/2.5$ | 98,064% | 99,758% |
| | $f/5$ | 98,046% | 99,819% |
| | $f/10$ | 96,479% | 99,701% |

satisfatória também em nível local (ampliações), obtendo valores acima de 99% para $R$ e valores acima de 95% para o $SSIM$. Numa análise subjetiva também se constata o alto nível de semelhança. Com base em tais resultados e tendo em vista que a aplicação dos mesmos é destinada para o uso de algoritmos de segmentação que não possui uma dependência com a forma do desfoque, ou seja, com o PSF, mas com a presença ou não do mesmo, a conclusão é que as imagens obtidas anteriormente são válidas para análise de algoritmos de segmentação que objetivam explorar tal modelo. Em especial, há grande potencial de exploração dessa técnica de segmentação em aplicações de tempo real com alta exigência de segmentação, como ocorre, por exemplo, na industria automotiva.

Como uma forma de provocar um descasamento entre as partes com foco e gerar uma adversidade maior, a proposta é que como trabalho futuro as imagens geradas artificialmente, tenham uma etapa adicional acrescida ao fim, onde um ruído (gaussiano, sal e pimenta, etc.) seja gerado.

Uma ação que merece destaque futuro é a repetição da técnica de comparação (imagem real comparada a uma artificial) para outros casos de cenas, a fim de haver uma maior validação estatística de resultados.

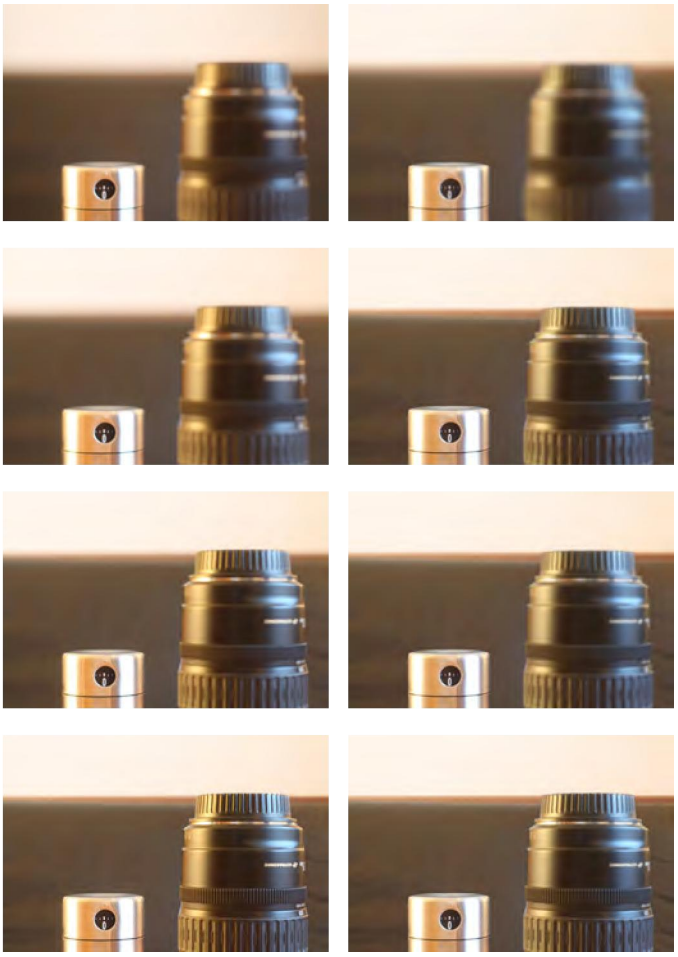

Figura 8: Resultados visuais das imagens completas reais (coluna da esquerda) e artificais (coluna da direita) para $f/1.8$ (primeira fileira), $f/2.5$ (segunda fileira), $f/5$ (terceira fileira) e $f/10$ (quarta fileira).

## AGRADECIMENTOS

Os autores deste artigo gostariam de agradecer o apoio financeiro da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) por meio de concessão de bolsa de estudos para o primeiro autor.

## REFERÊNCIAS

[1] F. P. de Luca, "Segmentação de imagens com baixa profundidade de campo para aplicações de tempo real," Master's thesis, Centro Universitário da FEI, São Bernardo do Campo, 2016.

[2] F. P. de Luca and C. E. Thomaz, "Analysis of methods for extraction of information on images with low-depth of field," in *Anais Congresso SAE Brasil 2016*, Congresso SAE Brasil. São Paulo: SAE, set. 2015.

[3] B. A. Barsky and T. J. Kosloff, "Algorithms for rendering depth of field effects in computer graphics," in *Proc. WSEAS International Conference Computers 2008*, ser. ICCOMP'08, WSEAS International Conference on Computers. Stevens Point, WI: World Scientific and Engineering Academy and Society (WSEAS), 2008, pp. 999–1010.

[4] R. Jacobson, S. Ray, G. G. Attridge, and N. Axford, *Manual of Photography (Media Manual)*, 9th ed. Oxford: Focal, August 2000.

[5] J. Tarrant, *Understanding Digital Cameras: Getting the Best Image from Capture to Output*. Oxford: Focal, February 2007.

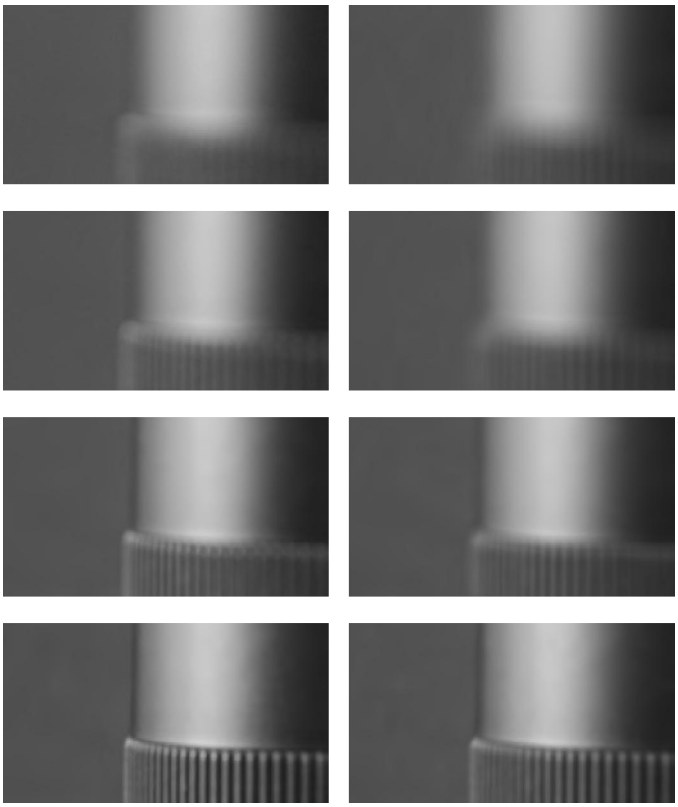

Figura 9: Resultados visuais das ampliações em região de médio contraste reais (coluna da esquerda) e artificais (coluna da direita) para $f/1.8$ (primeira fileira), $f/2.5$ (segunda fileira), $f/5$ (terceira fileira) e $f/10$ (quarta fileira).


[6] E. Allen, S. Triantaphillidou, and G. G. Attridge, *The manual of photography*. Oxford: Elsevier/Focal Press, 2010.

[7] M. Potmesil and I. Chakravarty, "A lens and aperture camera model for synthetic image generation," in *Proc. SIGGRAPH Annual Conference Computer Graphics and Interactive Techniques 1981*, SIGGRAPH Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM, ago. 1981, pp. 297–305.

[8] S. Kuthirummal, H. Nagahara, C. Zhou, and S. Nayar, "Flexible depth of field photography," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 58–71, jan. 2010.

[9] C. D. Claxton and R. C. Staunton, "Measurement of the point-spread function of a noisy imaging system," *J. Opt. Soc. Am. A*, vol. 25, no. 1, pp. 159–170, jan. 2008.

[10] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, abr. 2004.


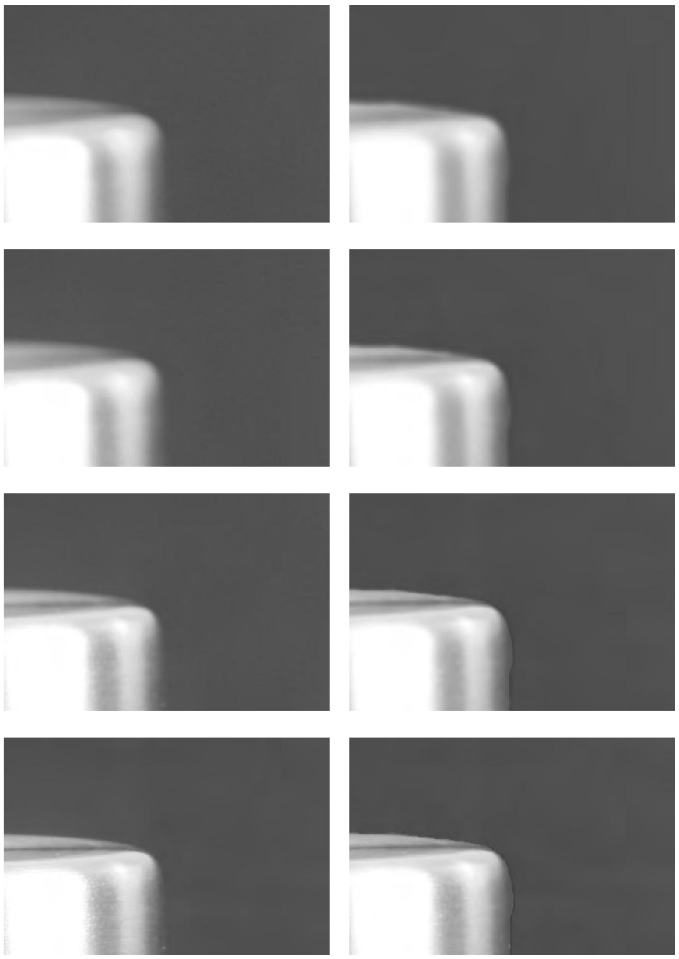

Figura 10: Resultados visuais das ampliações em região de alto contraste reais (coluna da esquerda) e artificais (coluna da direita) para $f/1.8$ (primeira fileira), $f/2.5$ (segunda fileira), $f/5$ (terceira fileira) e $f/10$ (quarta fileira).