

Uma abordagem híbrida e semi-automática para a remoção de nuvens e sombras de imagens de satélite

Danilo Frazão Sousa¹
Ana Carolina Quintão Siravenha¹
Damares Crystina Oliveira de Rezende¹
Reginaldo Cordeiro Dos Santos Filho¹
Evaldo Gonçalves Pelaes¹

¹ Universidade Federal do Pará – UFPA
Av. Augusto Correa, 01 – 66075-110 – Belém – Pa, Brasil
danilofrazao,siravenha,@ufpa.br, damares.resende@itec.ufpa.br, reginaldo.filho@icen.ufpa.br,
pelaes@ufpa.br

Abstract. The main goal of this work is to propose an efficient, and so automatic as possible process for replace clouds and shadows regions from satellite images. The motivation comes from the problems caused by these elements, among them: hinder the identification of image objects, prejudice urban and environmental monitoring, and disfavor crucial steps of the digital image processing to extract information from the user, such as segmentation and classification. Through a hybrid approach, is proposed a method for decomposing regions using a nonlinear median low-pass filter, in order to map structure (homogeneous) regions, like vegetation, and texture (heterogeneous) regions, like urban, in the image. In these areas, was applied the restoration method called inpainting by smoothing based on Discrete Cosine Transform (DCT), and an exemplar based texture synthesis. It is important to highlight that the techniques were modified to be able to work with images of peculiar characteristics as those obtained by satellite sensors, such as the large size and the high spectral variation. Finally, was verified the effectiveness of this methodology through a qualitative assessment, since is possible to check the trade-offs of each approach, and merge the advantages of each one, using the region mapping, to generate a more reliable result about the regions redefinition.

Keywords: remote sensing, image processing, inpainting, sensoriamento remoto, processamento de imagens, inpainting

1. Introdução

Em sensoriamento remoto é comum a presença de interferências atmosféricas, tais como nuvens e neblinas, ou mesmo sombras consequentes destes elementos, durante a captação de uma cena por um sensor. Essas ocorrências estão presentes em grande quantidade em regiões de clima equatorial e tropical quente e semi-úmido, e podem prejudicar a identificação de objetos da superfície da terra em uma imagem, afetando atividades como o monitoramento ambiental e urbano e as etapas subseqüentes do processamento digital de imagens, como a segmentação e classificação, grandes responsáveis pela extração de informações ao usuário.

Especialistas em todo o mundo têm realizado diversos estudos a fim de encontrar uma solução eficaz para reduzir ou eliminar o impacto causado pelas nuvens e sombras. É importante perceber que esse tipo de técnica voltada para o processamento de imagens de satélite deve ser, na mesma medida, robusta e computacionalmente viável, devido às grandes dimensões e variações espectrais desse tipo de imagem. Neste sentido, o incentivo para desenvolver novas alternativas e melhorar técnicas existentes nesta área é notável.

A remoção e a redefinição das regiões de nuvens e sombras são feitas basicamente através de duas abordagens: utilizando imagens de referência, em que se tem as informações reais das regiões cobertas, podendo substituí-las pelas áreas correspondentes; ou estimando as regiões

ocultadas por estes elementos, quando não se tem nenhuma informação *a priori* daquela cena. A primeira estratégia pode ser realizada fazendo uma análise multi-temporal (Zhang, Qin e Qin (2010)), ou usando uma cena captada por um sensor diferente, como em Hoan e Tateishi (2008), que usou uma imagem SAR (*Synthetic-aperture radar*). Já para estimar e redefinir as regiões, pode ser usado um método de interpolação chamado *inpainting* (Bertalmio et al. (2000)), que tem por objetivo preencher (ou restaurar) regiões danificadas e remover objetos, sem que a operação seja perceptível, funcionando bem para regiões homogêneas. Uma variação desta técnica, usando a transformada *Bandelet*, para remover nuvens em imagens de satélite foi proposta por Maalouf et al. (2009).

Outro método utilizado para redefinir regiões é a síntese de textura, que geralmente trabalha em blocos e alcança bons resultados para áreas que contém algum padrão textural ou que apresentam heterogeneidade. Em Liu, Wong e Fieguth (2010), é usada uma abordagem de campos hierárquicos de árvores estruturadas (*Tree-Structured Hierarchical Fields*) para modelar o comportamento complexo de dados operacionais de imagens *RADARSAT SAR* de mares e geleiras.

Alguns estudos têm explorado soluções híbridas que preservam a eficácia da síntese de textura e do *inpainting* sobre o mesmo processamento. Nesse contexto, cada abordagem é responsável por redefinir as regiões de textura (heterogênea) e estrutura (homogênea) separadamente, garantindo imagens sem componentes atmosféricos e com texturas reais.

Este trabalho apresenta uma nova abordagem híbrida usando *inpainting* por suavização baseado em DCT, proposta por Garcia (2010), e síntese de textura baseada em exemplos, proposta por Criminisi, Perez e Toyama (2004), ambas com modificações propostas e necessárias para o processamento de imagens de satélite. A união das técnicas é permitida por um processo de mapeamento de textura e estrutura baseado em Buades et al. (2011), aliado a um filtro de mediana não-linear, que permite definir o contexto em que se encontram os elementos a serem removidos.

Este artigo está estruturado da seguinte maneira: Na Seção 2 são explicados os métodos que compõem a abordagem híbrida. Seguindo, a Seção 3 mostra a aplicação da abordagem proposta neste trabalho em detrimento de outras. Finalmente, a Seção 4 faz um resumo do que foi tratado e faz perspectivas futuras.

2. Metodologia

2.1. Detecção de Regiões

Em primeiro lugar, serão definidas as regiões da imagem que serão processadas, usando o algoritmo descrito em (SOUSA; SIRAVENHA; PELAES, 2012). Nesta etapa, as características da imagem são identificadas e separadas em quatro classes: nuvem densa, nuvem suave, sombra e área não afetada. Para isso, são calculadas medidas estatísticas da cena, como a média e o desvio padrão da distribuição dos valores digitais dos *pixels*. A equação 1 descreve essa operação, que é realizada para as bandas correspondentes à cor verdadeira (*true color*) do sensor.

$$m(x, y) = \begin{cases} f(x, y) < f_s \times f_{m-dp}, & f(x, y) \in 0, \\ f_{m-dp} < f(x, y) < f_m, & f(x, y) \in 1, \\ f_m < f(x, y) < f_n \times f_{m+dp}, & f(x, y) \in 2, \\ f(x, y) > f_n \times f_{m+dp}, & f(x, y) \in 3, \end{cases} \quad (1)$$

onde $f(x, y)$ é o valor digital do pixel, f_m é o valor médio dos pixels da imagem, f_{m+dp} e f_{m-dp} são a soma e a subtração do valor médio com o desvio padrão da imagem, respectivamente. Também há a presença dos fatores f_n (fator de nuvem) e f_s (fator de sombra) que servem para

regular a detecção das regiões, modificando os limiares de cada classe. Cada região foi rotulada com um respectivo número. Regiões definidas como 0 são as regiões de sombra da imagem, as definidas como 1 são as livres de qualquer tipo de ruído, já as definidas como 2 são rotuladas como sendo nuvens esparsas, e finalmente as definidas como 3 são detectadas como nuvens densas.

Para finalizar o processo, é aplicada uma operação de abertura morfológica para remover pequenos objetos, detectados por engano, que podem provocar confusões nos passos seguintes.

2.2. Inpainting por suavização baseado em DCT

Este método foi proposto por Garcia (2010), e assim como em Bertalmio et al. (2000), é baseado na propagação de informação através da suavização. A especificidade desta abordagem está relacionada com a utilização da Transformada Discreta do Cosseno (DCT - *Discrete Cosine Transform*) para simplificar e resolver sistemas de equações lineares, gerando resultados eficientes.

O processo é aplicado para dados multidimensionais conforme a Eq. 2.

$$\hat{y}_{k+1} = IDCTN(\Gamma^N \circ DCTN(y_k)). \quad (2)$$

Onde y é um *pixel*, $\hat{y}$ é o *pixel* suavizado, $DCTN$ e $IDCTN$ referem-se à DCT N-dimensional e sua inversa, respectivamente; k é o número de iterações, N é o número de dimensões, $\circ$ é o produto Schur e Γ^N representa um tensor de rank N . Para mais informações sobre a definição desses dois últimos, consultar Garcia (2010).

É importante observar que, quando há valores indefinidos na imagem, a suavização atua como um interpolador de dados, funcionando como um método de *inpainting*. A fim de acelerar a convergência, o processo inicia com uma iteração de interpolação do vizinho mais próximo sobre a imagem a ser restaurada.

2.3. Síntese de Textura

A síntese de textura tem sido um campo intenso de estudos, devido a sua variedade de propósitos. Pode ser aplicada em tarefas de preenchimento de objetos, recuperação de imagens, compressão de vídeos, remoção de plano de fundos e etc.

Textura pode ser definida como um padrão visual num plano infinito 2D que, em alguma escala, tem uma distribuição estável. Então, naturalmente, pode-se obter uma amostra finita de texturas presentes neste plano, a fim de sintetizar outras amostras a partir das mesmas. Esta amostra finita pode ser extraída de inúmeras texturas distintas, o que pode se tornar um problema. Para contorná-lo, assume-se que a amostra deve ser grande o suficiente para capturar a distribuição textural da imagem (Efros e Leung (1999)). Além disso, a síntese de textura é responsável pela fusão de regiões contínuas com mínima imperfeição e percepção da operação, garantindo a qualidade visual.

A abordagem proposta por Criminisi, Perez e Toyama (2004) visa eliminar e redefinir objetos em uma imagem digital com informações de regiões vizinhas. Este método utiliza a síntese de textura para preencher as regiões que contém padrões texturais bidimensionais com estocasticidade moderada. Para isso, gera amostragem de novas texturas a partir de áreas fonte da imagem e faz uma cópia para as áreas alvo.

Na Fig. 1, é apresentado o algoritmo de Criminisi, Perez e Toyama (2004), no qual se supõe que uma imagem em que existe uma área fonte Φ e uma área alvo Ω , contornada por $\delta\Omega$, claramente distinguível (Fig. 1 (a)). O objetivo é sintetizar a área delimitada pelo fragmento Ψ_p , o qual é centralizado no ponto $p \in \Phi$, ilustrado na Fig. 1 (b). Então, são computados os candidatos mais prováveis para preencher Ψ_p apresentados em $\delta\Omega$, por exemplo, $\Psi_{q'}$ e Ψ_q'' na

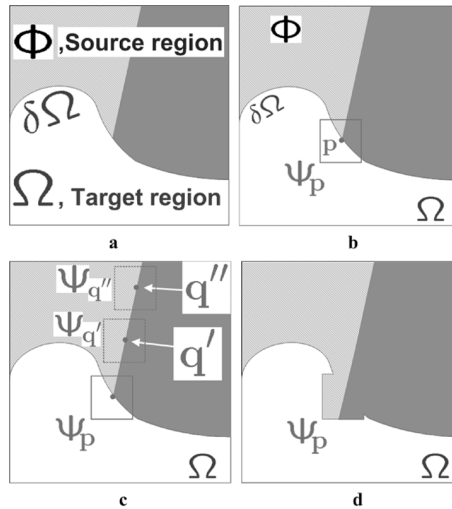


Figura 1: Síntese de textura baseada em modelos: (a) Imagem original. (b) Fragmento Ψ_p centralizado em $p \in \Phi$. (c) Candidatos mais prováveis $\Psi_{q'}$ e $\Psi_{q''}$. (d) Candidato mais provável é propagado para o fragmento alvo.

Fig. 1 (c). Entre os candidatos, existe um que mais se aproxima do fragmento alvo, tendo seus pixels correspondentes copiados no fragmento Ψ_p . Este processo é repetido até o preenchimento completo de Ω . Na Fig. 1 (d) pode-se notar que a textura, bem como a estrutura (a linha que separa as regiões cinzas, clara e escura), são propagadas sobre o fragmento Ψ_p .

No algoritmo, cada pixel mantém um valor de cor (que é nulo nos *pixels* a serem preenchidos) e um valor de confiança, que reflete a confiança no valor de cor com base no momento em que este pixel foi definido. Durante a execução do algoritmo, os fragmentos localizados no contorno $\delta\Omega$ recebem um valor de prioridade temporário, definindo a ordem de preenchimento. Assim, um processo iterativo é executado na seguinte sequência:

1. Computando prioridades dos fragmentos: O algoritmo realiza a tarefa de síntese através de uma estratégia de preenchimento *best-first*, que depende inteiramente dos valores de prioridade que são atribuídos a cada fragmento na frente de preenchimento. O cálculo da prioridade é tendenciosa para aquelas regiões que: a) estão sobre a continuação do bordas fortes e b) são cercadas por *pixels* de alta confiança.
2. Propagação das informações de estrutura e de textura: Uma vez que todas as prioridades na frente de preenchimento são computadas, o fragmento $\Psi_{\hat{p}}$ de maior prioridade é encontrado, sendo depois preenchido com dados extraídos da região fonte Φ .
3. Atualização de valores de confiança: Depois que o fragmento $\Psi_{\hat{p}}$ é redefinido com novos valores de *pixel*, o valor de confiança é atualizado na área delimitada por $\Psi_{\hat{p}}$.

2.4. Decomposição da Imagem

O algoritmo *cartoon+texture* (*cartoon* é uma outra denominação de estrutura bastante usada na literatura) decompõe qualquer imagem f na soma de uma parte de estrutura, u , em que aparecem apenas as formas grosseiras da imagem, e uma parte de textura v contendo padrões de oscilação. Tal decomposição $f = u + v$ tem comportamento análogo aos clássicos filtros passa-altas e passa-baixas do processamento de sinais. A parte de estrutura de uma imagem contém bordas fortes, e portanto baixas e médias frequências, enquanto que a textura tem frequências altas, podendo ainda conter médias. Desta forma, os algoritmos de decomposição linear não

conseguem fazer uma separação clara entre estrutura e textura, uma vez que, geralmente eles apenas suavizam as bordas, extraem as altas frequências e as introduzem na imagem de textura. Frequentemente, o filtro passa-baixas acaba não atenuando algumas componentes de textura, que permanecem erroneamente na parte de estrutura.

Em Buades et al. (2011) é proposta uma solução usando um problema variacional aproximado e rápido, obtido pela aplicação de um par de filtros não-lineares passa-baixas e passa-altas. Neste algoritmo, uma decisão é tomada para cada ponto da imagem, definindo se o mesmo pertence à parte de estrutura ou à parte de textura. Este processo é feito pelo cálculo de uma variação total local da imagem em torno do ponto, e comparando-a com a variação total local depois da aplicação de um filtro passa-baixas. Quando a imagem é convoluída por um filtro passa-baixas, pontos de estrutura de uma imagem tendem a ter uma variação total local oscilando lentamente. Por outro lado, pontos de textura mostram forte decaimento de variação total local após esta operação.

A parte de *cartoon* mantém os valores originais da imagem em pontos denominados como não-texturais. Em pontos identificados como de textura, esta parte assume o valor filtrado. Já nos pontos em que a decisão é confusa, uma média ponderada entre os mesmos é feita. A parte de textura v é simplesmente a diferença entre a imagem original f e sua parte de estrutura u .

Outro ponto importante é o parâmetro de escala da textura, para especificar o grau de decomposição. Não há uma decomposição única de uma imagem em textura e *cartoon*. Uma textura pode ser mantida na parte de estrutura para baixos valores do parâmetro de escala. O parâmetro de escala no algoritmo é crucial, e deve ser escolhido empiricamente.

Tal como indicado no algoritmo, a decomposição *cartoon+texture* apenas requer a aplicação de dois filtros passa-baixas sobre o gradiente da imagem, que são executados diretamente por uma convolução discreta. Para informações mais detalhadas sobre esta técnica, consultar Buades et al. (2011).

2.5. Mapeamento das regiões de estrutura e de textura

Uma estratégia foi proposta para mapear áreas de estrutura e textura de uma imagem de satélite com base na decomposição descrita na subseção 2.4. Este processo começa transformando a componente que contém a informação de textura, ou seja, a imagem v , em uma imagem binária com valores 1 para áreas heterogêneas de textura e 0 para áreas de estrutura. Em seguida, um filtro não-linear de mediana é aplicado para tornarem homogêneas (suavizadas), as áreas onde pequenas lacunas de uma determinada classe são cercadas por regiões onde predomina uma outra característica.

Essa etapa é crítica devido à presença de nuvens e sombras na imagem, e é realizada de modo a definir corretamente as técnicas a serem utilizadas para cada região. Isso acontece porque as nuvens e as sombras sempre serão componentes da estrutura, então, para definir qual técnica usar para removê-las, deve-se observar as regiões circundantes à esta, isto é, seu contexto. Portanto, como resultado da aplicação do filtro, as regiões a serem redefinidas são mapeadas na imagem binária para regiões de estrutura e textura, para que sejam aplicados o *inpainting* e a síntese textura na imagem de entrada, respectivamente.

2.6. Adequação das técnicas para imagens de satélite

Os métodos de restauração de regiões descritos até aqui, foram projetados e testados especificamente para imagens sintéticas de cenas comuns. Em se tratando de processamento de imagens obtidas por sensores orbitais, devem ser levadas em consideração algumas características peculiares, como o tamanho e alto grau de heterogeneidade entre os *pixels*, resultantes do grande campo de visada do sensor e da alta variação de reflectância dos objetos

presentes na superfície terrestre, respectivamente.

O tamanho da imagem está intrinsecamente ligado ao número de operações e ao consumo de memória exigidos pelas técnicas. A máquina usada para testes contém um processador de dois núcleos operando à 1.8Ghz, e memória RAM de 3GB, sendo considerada um equipamento *desktop* usual. Para este computador, ocorria estouro de memória para a operação de *inpainting* e processamento “sem fim” para a síntese de textura.

Com o objetivo de contornar este problema, optou-se por fazer um processamento por blocos na imagem, ou seja, em cada operação dividiu-se a imagem em blocos de tamanhos iguais e equivalentes ao tamanho de imagens sintéticas, permitindo mais economia de memória. No entanto, a busca por fragmentos candidatos, relativa à síntese de textura, ainda era um processo muito lento. Para acelerá-lo, adotou-se uma estratégia local para a busca por blocos.

Supondo que a continuação das áreas cobertas por nuvens e sombras estão nos seus arredores, então não faz sentido procurar por substitutos em regiões muito distantes. Logo, é buscado na k -ésima vizinhança, a partir do pixel central $f(i, j)$, o fragmento fonte que mais se aproxima da região alvo. O conceito de vizinhança é dado em Gonzalez e Woods (2008), sendo k a camada que determina a distância de $f(i, j)$ ao vizinho. Por padrão, k foi convencionado ser 10 vezes o tamanho do fragmento alvo. Portanto, aplicada esta modificação, a síntese de textura reduziu consideravelmente seu tempo de processamento, mantendo a qualidade.

3. Resultados

Não há na literatura métricas bem definidas de avaliação quantitativa para algoritmos de *inpainting*. Ao longo deste trabalho foram testadas as medidas de avaliação PSNR (local e global), Kappa e a soma das diferenças absolutas. Concluiu-se que nenhuma delas avalia apropriadamente as diferentes abordagens para redefinir regiões. Por exemplo, certos resultados da aplicação da abordagem híbrida, ou até mesmo usando apenas a síntese de textura, resultam visualmente em um preenchimento de regiões que parece ser mais consistente do que os obtidos com o *inpainting* por suavização (que às vezes implicam em grandes borrões). No entanto, olhando para as avaliações quantitativas citadas, é comum que a última alcance melhores resultados.

Com relação à avaliação qualitativa, a Fig. 2 (a) mostra uma imagem do sensor CBERS 2, com a combinação das bandas 341 (*true color*), afetada pela presença de nuvens densas e sombras sobre áreas de textura (urbana) e estrutura (vegetação). Na Fig. 2 (b) é ilustrada uma máscara (em preto) contendo essas regiões a serem redefinidas. Esta máscara foi produzida pelo algoritmo citado na Seção 2.1. As Figs. 2 (d) e (e) mostram os resultados isolados do *inpainting* e da síntese de textura, respectivamente. A partir destes resultados, então é decidido aplicar o método híbrido. A Fig. 2 (c) mostra uma imagem binária, depois de passar a imagem v por um filtro de mediana, contendo regiões de textura (branco) e estrutura (preto). A partir deste mapeamento, a Fig. 2 (f) mostra o resultado da abordagem híbrida, onde se tem a união das vantagens das técnicas aplicadas em regiões adequadas do mesmo.

Para uma análise minuciosa do *trade-off* de cada técnica, a Fig. 3 mostra um zoom da área urbana da Fig. 2, em que é visível a predominância de textura nesta região. Como esperado, o *inpainting* gerou borrões destacados pelos círculos vermelhos, ao contrário da síntese de textura, que conseguiu redefinir a área de maneira mais coerente. O contrário é constatado quando se analisa a região de vegetação da Fig. 4, mais homogênea, e favorável portanto ao *inpainting*. Neste caso, a síntese de textura deixa alguns fragmentos incoerentes, mostrados nos círculos, se contrapondo às regiões circundantes àquela redefinida.

A abordagem híbrida, portanto, combina as vantagens de ambas as técnicas, aplicando cada uma em seu estado mais favorável, resultando em uma remoção e substituição de nuvens e

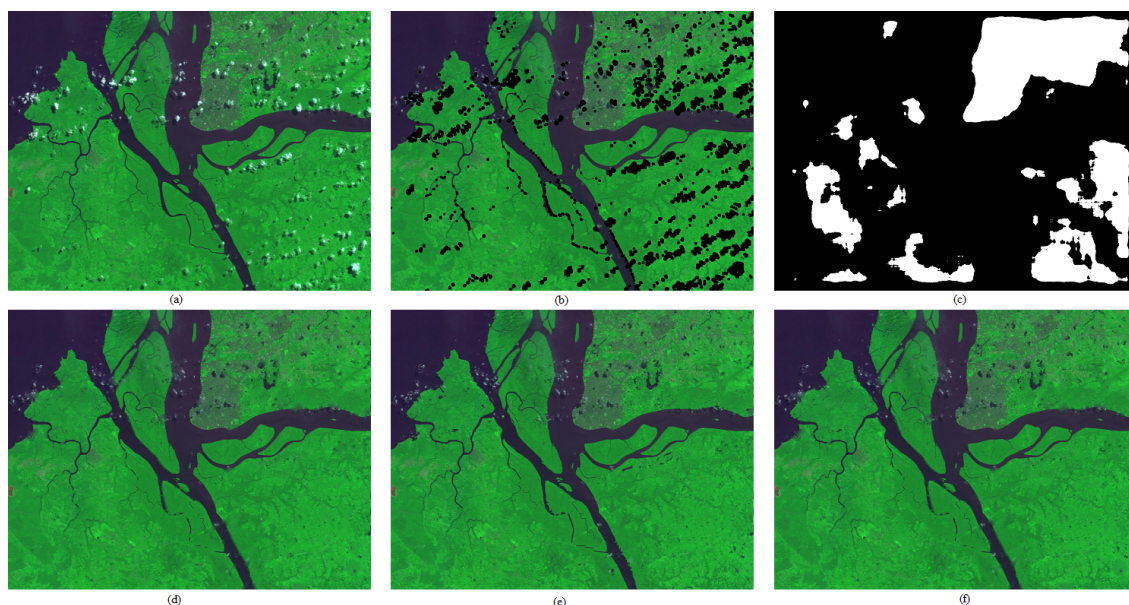


Figura 2: Processo de remoção de nuvens e sombras para imagens de satélite: (a) Imagem Original. (b) Máscara com as regiões detectadas. (c) Resultado do *inpainting*. (d) Resultado da Síntese de Textura. (e) Imagem binária representando a decomposição da imagem em textura e estrutura. (f) Resultado da abordagem híbrida.

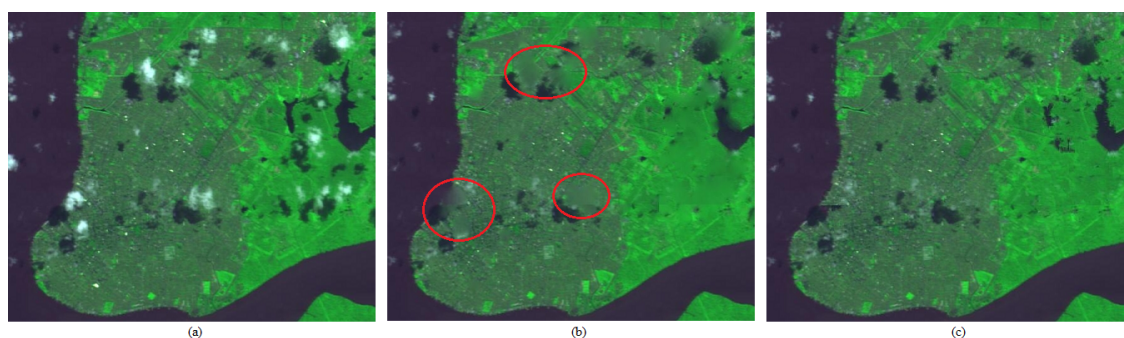


Figura 3: Zoom da área urbana da Fig. 2: (a) Imagem Original, (b) Resultado do *inpainting*, e (c) Resultado da Síntese de Textura.

sombras mais eficiente.

4. Conclusão

Neste trabalho foi apresentada uma nova forma de realizar a remoção e redefinição de nuvens e de sombras em imagens de satélite, através de uma abordagem híbrida. No processo, foram propostas algumas modificações nas técnicas de *inpainting* e de síntese de textura, assim como um mapeamento das regiões de estrutura e de textura na imagem, em que serão aplicadas tais técnicas, respectivamente.

Na avaliação qualitativa foi evidente que a abordagem híbrida supera a utilização das técnicas de forma separada. É visível a permanência de nuvens e sombras não detectadas nestas figuras, o que mostra que a melhora da detecção é uma tarefa fundamental a ser realizada em trabalhos futuros.

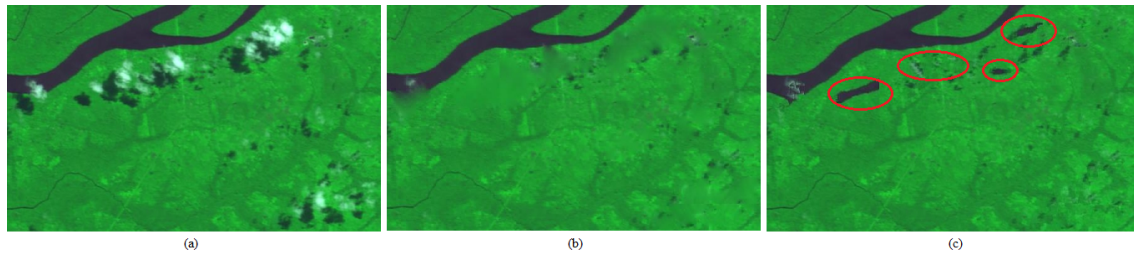


Figura 4: Zoom da área de vegetação da Fig. 2: (a) Imagem Original, (b) Resultado do *inpainting*, e (c) Resultado da Síntese de Textura.

Agradecimentos

Este trabalho é parcialmente financiado pela Fundação Amazônia Paraense de Amparo à Pesquisa e Vale S/A por meio do edital 001/2010. Também é suportado pela Universidade Federal do Pará e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico.

Referências

- BERTALMIO, M. et al. Image inpainting. In: *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. [S.l.]: ACM Press/Addison-Wesley Publishing Co, 2000. p. 417–424.
- BUADES, A. et al. Cartoon+texture image decomposition. In: . [s.n.], 2011. Disponível em: <http://dx.doi.org/10.5201/ipol.2011.blmv_ct>.
- CRIMINISI, A.; PEREZ, P.; TOYAMA, K. Region filling and object removal by exemplar-based image Inpainting. In: *IEEE Transactions On Image Processing*. [S.l.: s.n.], 2004. v. 13(9), p. 1200–1212. IEEE Computer Society.
- EFROS, A.; LEUNG, T. Texture synthesis by non-parametric sampling. In: *International Conference on Computer Vision*. [S.l.]: IEEE Computer Society, 1999. p. 1033–1038.
- GARCIA, D. Robust smoothing of gridded data in one and higher dimensions with missing values. In: . [S.l.]: Elsevier, 2010. v. 54, n. 4, p. 1167 – 1178.
- GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. [S.l.]: Addison-Wesley Publishing Company, 2008.
- HOAN, N. T.; TATEISHI, R. Cloud removal of optical image using SAR data for ALOS applications. Experimenting on simulated ALOS data. In: *Beijing: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. [S.l.: s.n.], 2008.
- LIU, Y.; WONG, A.; FIEGUTH, P. Remote sensing image synthesis. In: *Geoscience and Remote Sensing Symposium (IGARSS)*. [S.l.]: IEEE International, 2010. p. 2467 –2470.
- MAALOUF, A. et al. A bandelet-based Inpainting technique for clouds removal from remotely sensed images. In: *IEEE Transactions On Geoscience and Remote Sensing*. [S.l.: s.n.], 2009. v. 47, n. 7, p. 2363–2371.
- SOUSA, D.; SIRAVENHA, A. C.; PELAES, E. A hybrid approach to clouds and shadows removal in satellite images. In: *Computational Modelling of Objects Represented in Images III*. [S.l.]: CRC Press, 2012. p. 153–158. ISBN 978-0-203-07537-1.
- ZHANG, X.; QIN, F.; QIN, Y. Study on the thick cloud removal method based on multi-temporal remote sensing images. In: *International Conference on Multimedia Technology (ICMT)*. [S.l.]: IEEE, 2010. p. 1–3.