

## Detecção de Mudança para Identificação de Objetos Móveis em Vídeo UHD

Marielcio Gonçalves Lacerda  
Camila Souza dos Anjos Lacerda  
Leidiane do Nascimento Andrade  
Elcio Hideiti Shiguemori

Instituto de Estudos Avançados (IEAv) – Força Aérea Brasileira,  
São José dos Campos-SP, Brasil  
{marielcio, camilasa, leidiane, elcio}@ieav.cta.br

**Abstract.** Several sensors provide a very large range of data with different characteristics. In this line, this work aims to demonstrate an application for Data Fusion. An algorithm in OpenCV designed to detect moving objects in Ultra HD video purchased by Iris camera, installed on the Zvezda module of the International Space Station (ISS). The study area is located in Vancouver, Canada. The images are from Deimos-2 satellite in 1C level. The first stage of the research was to detect moving objects between successive frames of high resolution video. The detection of moving objects resulted in new images containing motion information, thus it was possible to separate the moving objects in the scene. In order to highlight the changes were applied one thresholding in the change detection image creating a binary image. So as to reduce the possibility of false positives, especially those generated by the movement of the sensor, was applied a morphological operation erosion. Finally, a new video was created showing the moving objects over time, allowing monitoring of these objects. In the second stage, there was the match between the georeferenced panchromatic image and the video frames of moving object, which allowed the identification of the object position associated with the each frame of video to monitor the displacement of objects.

**Palavras-chave:** — Detecção de mudança, Fusão de Dados, Change Detection, Data Fusion.

### 1. Introdução

Desde o surgimento do sensoriamento remoto até os dias atuais, nota-se constante evolução dos sistemas sensores. Estes sistemas tornaram-se cada vez mais eficazes e capazes de observar remotamente diferentes alvos presentes na superfície da Terra. A melhoria da resolução espacial e espectral dos sensores são alguns dos principais benefícios alcançados com o progresso desta tecnologia, que tem favorecido, em especial, a identificação de alvos urbanos, que geralmente se apresentam nas imagens pequenos e diversificados (Anjos, 2016). Outro benefício alcançado com o progresso do sensoriamento remoto é a utilização de imagens obtidas por sensores remotos a bordo de VANT's (Veículos Aéreos Não Tripulados), o que também tem favorecido a identificação de alvos urbanos, devido ao fato de poder agregar novas perspectivas aos dados de sensoriamento remoto utilizados neste tipo de estudo (Rebouças et al., 2013).

A utilização de VANT's tem adicionado novas perspectivas ao Sensoriamento Remoto, principalmente no campo da detecção de mudanças ocorridas na superfície terrestre. Como esta aplicação tem avançado rapidamente, surgem a cada dia novos desafios ao processamento e análise destas informações, sejam elas imagens estáticas (no presente trabalho tratada somente como imagem) ou sequências de imagens com intervalo de tempo menor que 1 segundo (neste trabalho tratada como vídeo). O crescente desenvolvimento e emprego de VANT's tem propiciado a captura de imagens e vídeos para uma grande gama de aplicações. Entre esses desafios pode-se mencionar o uso de um grande volume de dados, o que requer um grande esforço de processamento devido à sua complexidade e quantidade de informação. Outro desafio é a questão da integração de dados de diferentes fontes ou plataformas, que podem ser tratadas através da abordagem investigativa de Fusão de Dados (*Data Fusion*).

De acordo com Dalla Mura (2015), entre uma variedade de aplicações, a Fusão de Dados tem impactado a classificação e mapeamento de áreas urbanas, uma vez que o uso de imagens

provenientes de diferentes sensores, pode levar a um melhor desempenho da classificação, fornecendo informações adicionais e resolvendo conflitos.

Os vídeos de sensoriamento remoto sendo eles provenientes de VANT's ou ainda provenientes de plataformas espaciais, como no caso do presente trabalho, estão abrindo oportunidades para novas aplicações de sensoriamento remoto. Para explorar esse novo tipo de dados são necessários algoritmos sofisticados para interpretação da cena. De acordo com Mou e Zhu (2016), as principais dificuldades no sensoriamento remoto por vídeo derivado de plataformas espaciais são: a detecção de objetos em movimento devido à relativa baixa resolução espacial do vídeo adquirido a partir do espaço; e a avaliação das desvantagens que o movimento da câmera gera na interpretação da cena.

A detecção de mudança é definida como o processo de identificação de mudanças no estado de um objeto ou fenômeno em momentos diferentes (Singh, 1989). Este estudo utiliza a Fusão de Dados (trabalha com vídeo e com imagem ortorretificada), para obter a detecção de mudanças no ambiente urbano, especialmente a identificação de objetos em movimento.

No contexto de Fusão de Dados, o presente artigo aborda a identificação de objetos em movimento em um vídeo obtido por uma câmera instalada na Estação Espacial Internacional, utilizando como referência estática uma imagem orbital de alta resolução espacial.

## 2. Metodologia de Trabalho

A metodologia empregada neste estudo visa a identificação de objetos em movimento no vídeo UHD (*Ultra High Definition*), e determinação da posição geográfica dos objetos em movimento encontrados na primeira etapa do desenvolvimento, associando-se os frames do referido vídeo com a imagem pancromática georreferenciada. Para tal, foram implementadas em OpenCV técnicas de diferença de imagens, limiarização, morfologia matemática (erosão e dilatação) além de técnicas de casamento de padrões (*matching*).

### 2.1. Dados Utilizados

O conjunto de dados utilizado neste estudo consiste de uma imagem pancromática e um vídeo UHD, cedidos pela IEEE em parceria com a Deimos Imaging e UrtheCast, por ocasião do "The 2016 IEEE GRSS Data Fusion Contest". A imagem pancromática foi adquirida em 30 de maio de 2015, sobre Vancouver, no Canadá por meio do satélite DEIMOS-2 e possui resolução espacial de 1 metro, no nível 1C (produto calibrado e radiometricamente corrigido, manualmente ortorretificado e reamostrado).

O vídeo UHD em cores também foi adquirido sobre Vancouver em 2 de julho de 2015. O vídeo foi obtido pela câmera de alta resolução IRIS, que está instalada no módulo Zvezda da Estação Espacial Internacional (ISS). A câmera Iris possui um detector CMOS utilizado para capturar vídeos RGB com uma distância de amostra de solo de até 1 metro. O formato do frame é  $3840 \times 2160$  pixels e cobre aproximadamente  $3,8 \text{ km} \times 2,1 \text{ km}$  no terreno.

### 2.2. Detecção de Objetos em Movimento

Tendo em vista a detecção de objetos em movimento no vídeo, foi realizada a detecção de mudança por meio da diferença entre imagens de frames consecutivos, denominada "imagem diferença". Foi realizada a captura de dois frames consecutivos, nos tempos  $t_1$  e  $t_2$ , para a detecção de mudança e identificação de objetos móveis.

Normalmente, um vídeo obtido por uma estação espacial exigiria considerar o deslocamento da plataforma além do movimento da Terra, no entanto, neste estudo preliminares tais movimentos foram desconsiderados devido o intervalo de tempo entre os quadros ser inferior a 1 segundo.

A Figura 1 apresenta um frame do vídeo, com a finalidade de demonstrar a escala da imagem, a área abrangida e a dificuldade na detecção de objetos móveis. Em destaque na Figura 1 registra-se o recorte da imagem que gerou a área selecionada para testes apresentada na Figura 2.



Figura 1. Frame do vídeo UHD adquirido pela câmera Iris, instalada na ISS.

Apesar das dificuldades inerentes ao imageamento disponível, objetos em movimento, como carros, barcos e trens foram identificados no vídeo. A Figura 2 apresenta, em destaque, alguns desses objetos.

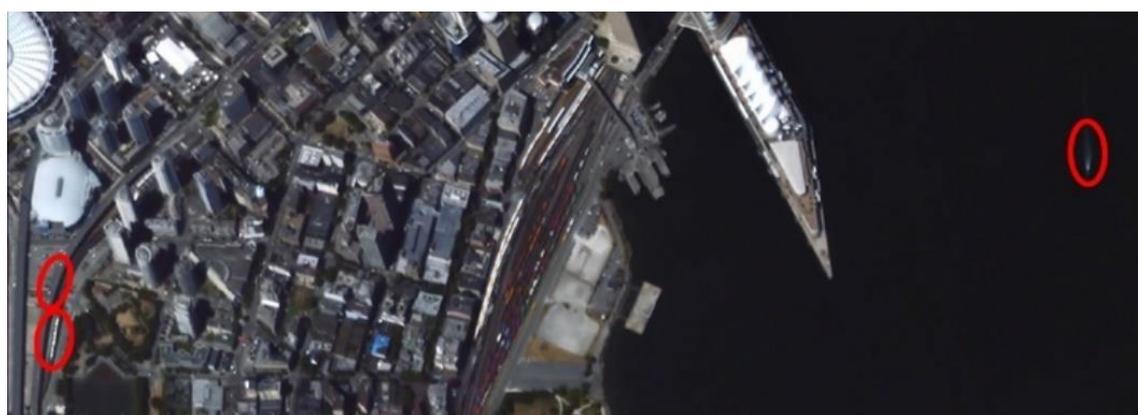


Figura 2. Ampliação do frame vídeo UHD, destacando alguns objetos em movimento.

Os frames, inicialmente obtidos em RGB, foram submetidos a uma transformação para tons de cinza, adotando-se o Método da Intensidade, amplamente utilizado e aceito em diversos trabalhos, sendo:

$$\text{Intensidade} = 0,21 \times \text{Vermelho} + 0,72 \times \text{Verde} + 0,07 \times \text{Azul}$$

As proporções empregadas nessa fórmula decorrem das proporções das células fotossensíveis da retina humana, denominadas como cones. Na Figura 3, tem-se um exemplo do frame transformado para tons de cinza pelo Método da Intensidade.



Figura 3. Frame do vídeo UHD, convertido para tons de cinza pelo Método da Intensidade.

Devido à movimentação da câmera, as imagens diferença podem apresentar borrões ou distorções que erroneamente podem ser considerados como objetos móveis, tais objetos são chamados de falsos positivos. Visando minimizar o problema de indicação de falsos positivos na imagem diferença, foi aplicada a técnica de limiarização na imagem, ou seja, criou-se uma imagem binária com apenas dois valores, determinados por um limiar (*threshold*) pré-estabelecido, baseado na análise da imagem e nas respostas para diferentes *thresholds* testados. Diversos testes foram realizados para definir o limiar mais adequado e para uma imagem diferença de 8 bits, um limiar igual a 20 apresentou um bom resultado. Desta forma, para gerar uma imagem binária limiarizada considerou-se que todos os pixels com valores inferiores a 20 assumiriam o valor de 0 (preto), enquanto todos os valores superiores ao limiar 20 assumiriam o valor de 255 (branco), sendo considerados como objetos móveis.

Ainda com o propósito de ignorar o ruído e minimizar os erros, técnicas morfológicas de erosão e dilatação foram aplicadas em todas as imagens diferença. A técnica de erosão é uma operação de filtragem morfológica em que os detalhes da imagem inferiores ao elemento estruturante são filtrados (removidos) da imagem. Ao contrário da erosão, que é uma técnica de diminuição ou afinamento, a dilatação “aumenta” ou “engrossa” os objetos presentes em uma imagem binária. A forma específica e a extensão desse espessamento são controladas pelo formato do elemento estruturante utilizado. (Gonzalez e Woods, 2010).

Por fim, visando a determinação da localização dos objetos em movimento já identificados, criou-se uma tabela onde os pares ordenados de coordenadas (x, y) desses pontos identificados em cada frame fossem armazenados, para que fosse possível uma associação com a imagem georreferenciada. A Figura 4 apresenta um exemplo de impressão na tela de um trecho do registro desses pares ordenados.



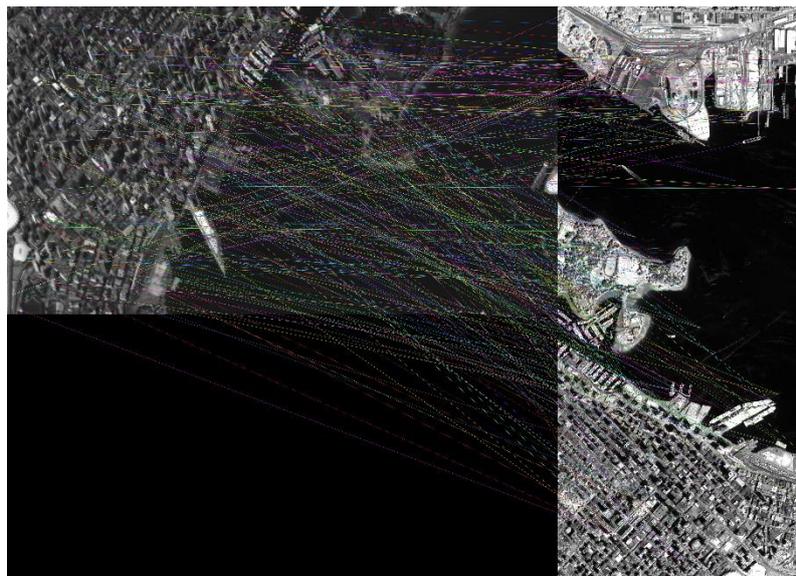


Figura 5. Exemplo da execução do matching implementado no OpenCV.

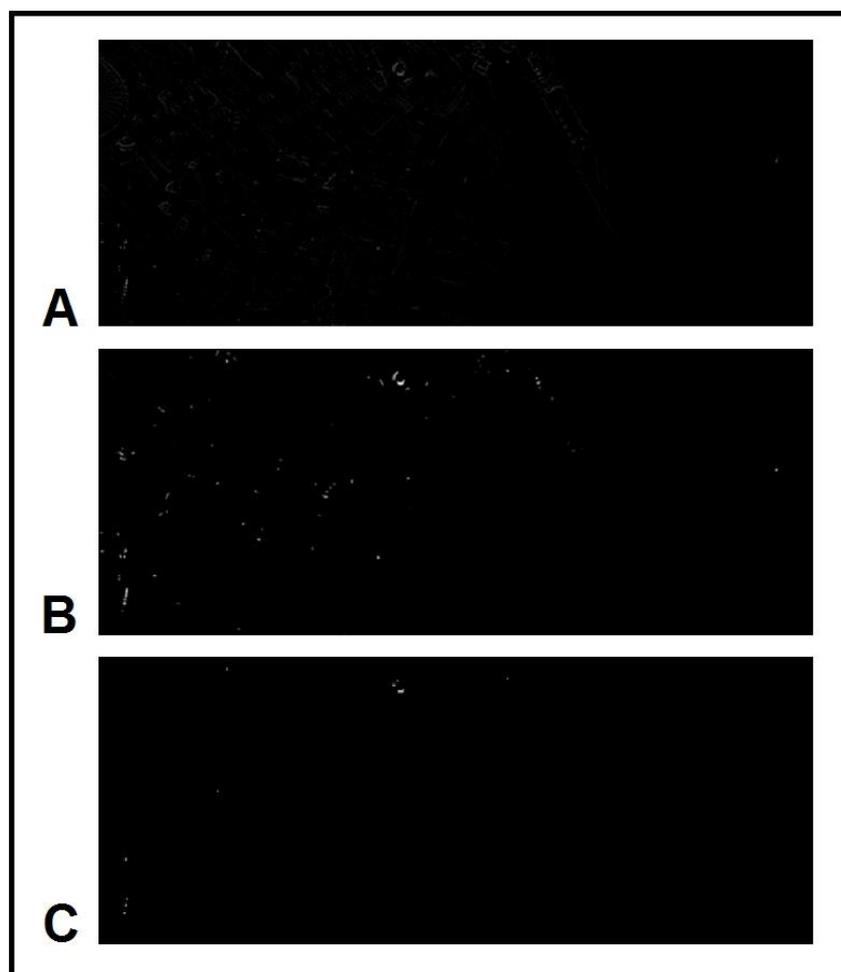


Figura 1. (A) Imagem Diferença obtidas entre dois frames consecutivos; (B) Imagem limiarizada; (C) Imagem após a aplicação das operações morfológicas Erosão e Dilatação.

Os procedimentos apresentados na Figura 6, demonstram os resultados da detecção de objetos móveis para um par frames consecutivos, o qual apresenta os resultados intermediários.

Vale ressaltar que esse processo foi realizado para todos os pares de frames do vídeo de referência. Em seguida um vídeo foi construído unindo as imagens resultantes do processo descrito acima (Imagem Diferença Limiarizada com Operações de Erosão e Dilatação) com o intuito de permitir a visualização binária de objetos em movimento ao longo do tempo, acompanhando o deslocamento dos mesmos.

A Figura 7 apresenta frames do vídeo original, juntamente com os respectivos frames no vídeo resultado, tornando possível a comparação entre as imagens originais e o acompanhamento dos objetos em movimento. São apresentados 4 (quatro) frames (20, 23, 26 e 33 segundos) para tornar possível a verificação do desempenho das técnicas utilizadas para a detecção de objetos em movimento. Tais frames são apresentados em pequenos recortes para facilitar a visualização dos objetos móveis.

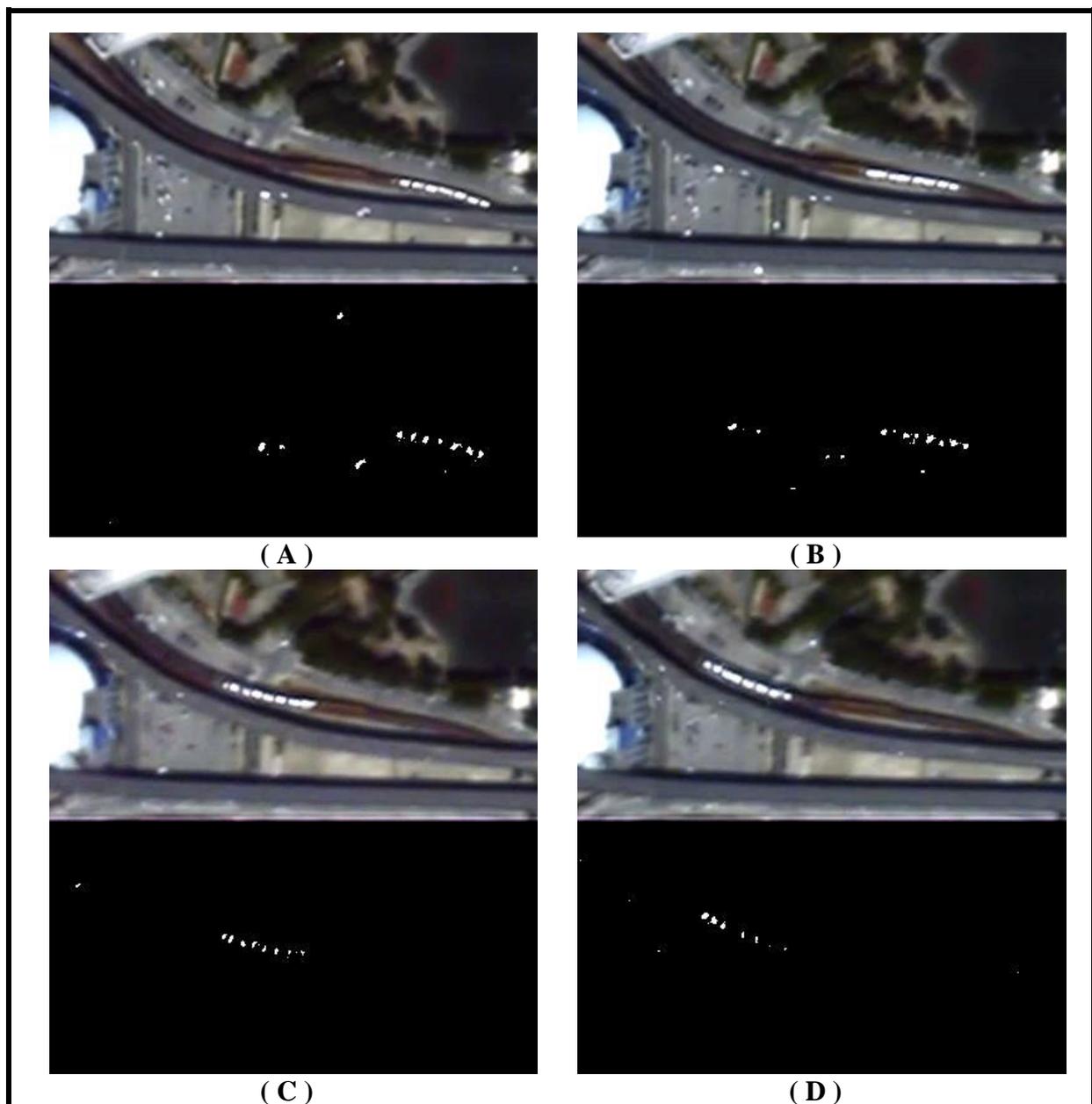


Figura 7. Detecção de objetos em movimento. Frames do vídeo original (RGB) e seu respectivo frame do vídeo resultado: (A) Frame 20 segundos; (B) Frame 23 segundos; (C) Frame 26 segundos; e (D) Frame de 33 segundos.

Por meio do *matching* entre o frame em análise e a imagem pancromática georreferenciada, consegue-se realizar o registro do frame, tornando-o georreferenciado. Como o frame em análise é o mesmo utilizado durante a execução do programa para gerar a imagem diferença atual, todos os pontos são comuns ao frame e a imagem diferença final. Como durante o processamento, os dados (x, y) de todos os pontos identificados como objetos em movimento na imagem diferença são armazenados em uma tabela, para se determinar as coordenadas geográficas desses pontos, basta fazer a associação da tabela da imagem diferença com a tabela das coordenadas do frame para determinar as coordenadas geográficas dos pontos identificados como objetos em movimento.

#### 4. Conclusões

O uso de vídeo obtido a partir de uma plataforma espacial é uma inovação no campo do sensoriamento remoto. Os resultados obtidos mostram que as abordagens utilizadas para identificar objetos em movimento são promissoras para o problema proposto. Apesar das dificuldades relacionadas ao deslocamento da plataforma, os objetos móveis foram detectados e a respectiva localização geográfica foi obtida para cada frame do vídeo.

No entanto, este estudo ainda está em desenvolvimento e outras abordagens devem ser estudadas, a fim de melhorar a precisão dos resultados.

#### Agradecimentos

Os autores gostariam de agradecer a Deimos Imaging e UrtheCast pela aquisição e fornecimento dos dados utilizados neste estudo

#### Referências Bibliográficas

- Anjos, C. S. **Classificação de Áreas Urbanas com Imagens Multiespectrais e Hiperespectrais Utilizando Métodos Não-Paramétricos**. 2016. 336p. Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2016.
- Bradski, G.; Kaehler, A. **Learning OpenCV - Computer Vision with the OpenCV Library**. O'Reilly Media, Sebastopol, USA, 2008. 575 p.
- Dalla Mura, D. M.; Prasad, S.; Pacifici, F.; Gamba, P.; Chanussot, J.; Benediktsson J. A. Challenges and Opportunities of Multimodality and Data Fusion in Remote Sensing. **Proceedings of the IEEE**, v. 103, n. 9, p.1585-1601, 2015.
- Fawcett, T. An introduction to ROC analysis. **Patter Recognition Letters**, v. 27, n.8, p. 861-874, 2006.
- Gonzales, R. C. Woods, R. E. **Processamento Digital de Imagens** Pearson Prentice Hall, São Paulo, Brasil, 2010. 624 p.
- Mou, L.; Zhu, X. X. Spatiotemporal Scene Interpretation of Space Videos via Deep Neural Network and Tracklet Analysis. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016, Beijing - China. **Proceedings...** Online. Disponível em: <<http://ieeexplore.ieee.org/document/7729468/>>. Acesso em: 15 nov. 2016.
- Rebouças, R. A.; Habermann, M; Shiguemori, E. H. Identificação de objetos móveis com uso de imagens aéreas obtidas por VANT. In: Simposio Brasileiro de Sensoriamento Remoto (SBSR), 2013, Foz do Iguaçu. **Anais...** São José dos Campos: INPE, 2013. Artigos, p. 8293-8300. Online. Disponível em: <<http://www.dsr.inpe.br/sbsr2013/files/p1599.pdf>>. Acesso em: 15 nov. 2016.
- Singh, A. Review Article Digital change detection techniques using remotely-sensed data. **International Journal of Remote Sensing**, v. 10, n. 6, p.989-1003, 1989.