

Um algoritmo alternativo para a alocação de janelas de fundo de céu aplicado à missão CoRoT

Luciana Guidon Coelho, Hae Yong Kim, Fábio de Oliveira Fialho

Resumo — CoRoT é um satélite franco-europeu com participação brasileira dedicado à sismologia estelar e à busca por exoplanetas. Este trabalho apresenta um algoritmo alternativo para a alocação de janelas para a estimação de ruído de fundo de céu para o canal de exoplanetas do CoRoT. A correção de fundo de céu é um procedimento padrão na redução de dados fotométricos que consiste na subtração do nível de fundo de céu das medições fotométricas de uma estrela em observação.

Palavras-Chave — CoRoT, ruído de fundo de céu, processamento digital de imagem.

Abstract— CoRoT is a French-led European satellite with Brazilian participation dedicated to stellar seismology and the search for exoplanets. This paper presents an alternative algorithm for window allocation in order to estimate the sky background noise of the CoRoT exoplanet channel. The sky background correction is a standard procedure in photometric data reduction which consists of subtracting the sky background mean level from photometric measurements of an observed star.

Keywords— CoRoT, sky background noise, digital image processing.

I. INTRODUÇÃO

O satélite franco-europeu CoRoT (*Convention, Rotation and planetary Transits*) faz parte de uma missão espacial dedicada à sismologia estelar e à busca por exoplanetas. Ambos os programas científicos baseiam-se em uma fotometria de alta precisão e requerem observações ininterruptas de longa duração. Para a descoberta de exoplanetas, o objetivo é detectar a ocultação temporária da estrela ao redor da qual orbita o planeta em questão, ou seja, detectar um trânsito planetário (Figura 1), o que pode ser traduzido por uma diminuição tênue do fluxo de fótons estelar coletado pelos detectores do tipo *charged coupled device* (CCD) [1] do fotômetro. Uma descrição mais detalhada sobre a missão CoRoT pode ser encontrada em [2].

Existem alguns ruídos e diversas fontes de luz que atingem os CCDs e que não são provenientes das estrelas em estudo. Tais ruídos e fontes de luz geram uma iluminação chamada de “fundo de céu” nos CCDs que é não homogênea e que precisa ser corrigida antes do início da exploração científica dos dados obtidos nas observações do satélite. A correção de fundo de céu é um procedimento padrão na redução de dados fotométricos. Consiste na subtração do nível médio de fundo de céu das medições fotométricas de uma estrela.

Para a determinação do nível médio, o CoRoT utiliza um conjunto de janelas de fundo de céu. Essas janelas precisam ser alocadas de forma representativa, garantindo que as inomogeneidades de fundo de céu presentes nos CCDs não

permaneçam poluindo as curvas de luz após as correções instrumentais.

Este trabalho apresenta, na introdução, uma breve descrição do problema e sua contextualização. Depois, apresentamos um novo algoritmo para a alocação de janelas de fundo de céu. A Seção II descreve o método de alocação de janelas de fundo de céu em utilização pela missão CoRoT, assim como suas limitações. Na Seção III são descritos três métodos de estimação de fundo de céu, dentre eles o método em uso pelo satélite. Em seguida, a Seção IV descreve brevemente os CCDs; e a Seção V detalha o método proposto neste trabalho. Por fim, a Seção VI apresenta a continuação deste trabalho.

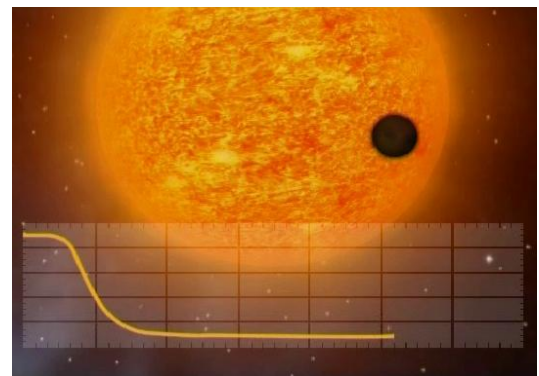


Fig.1. Representação artística do efeito de trânsito planetário na curva de luz de uma estrela (CNES).

II. O MÉTODO DE ALOCAÇÃO DE JANELAS DE FUNDO DE CÉU EM USO

O CoRoT utiliza 200 janelas de fundo de céu a serem alocadas por CCD - sendo que são dois CCDs para o canal de exoplanetas - com tamanho de 10x10 pixels e as mesmas devem ser alocadas automaticamente em espaços livres de estrelas e de pixels quentes e frios.

O método atualmente utilizado divide o CCD em n regiões quadradas, com n podendo ser diferente a cada corrida (ciclo de observação do céu). Então, para cada região quadrada, aplica-se uma filtragem tipo *smooth*. Sobre o resultado desta filtragem determinam-se os mínimos locais e as janelas são alocadas nesses mínimos. A quantidade de janelas é distribuída igualmente entre as n regiões, como mostrado na Figura 2. Este método quebra o problema original em subproblemas e toma como a solução do problema original o conjunto de soluções dos subproblemas. Infelizmente, o agrupamento das soluções dos subproblemas nem sempre é uma boa solução do problema original, principalmente, na presença de forte gradiente de fundo de céu. A Figura 2.a é um exemplo claro deste efeito. Neste exemplo, estrelas posicionadas no centro do CCD terão

seus fundos de céu muito mal estimados, pois nenhuma janela de fundo de céu foi alocada nesta região. Outro ponto negativo desta abordagem é que ela não leva em conta o impacto da escolha do método de estimação de fundo de céu na determinação das janelas de fundo de céu. Se decidir utilizar o método da janela mais próxima para estimar o fundo do céu, a alocação das mesmas deve ser feita de modo a minimizar a maior distância entre uma estrela e a sua janela mais próxima. Por outro lado, se o método escolhido for a triangularização, a escolha das janelas deve minimizar o número de extrapolações nas estimações de fundo de céu. Este tópico é detalhado na Seção III.

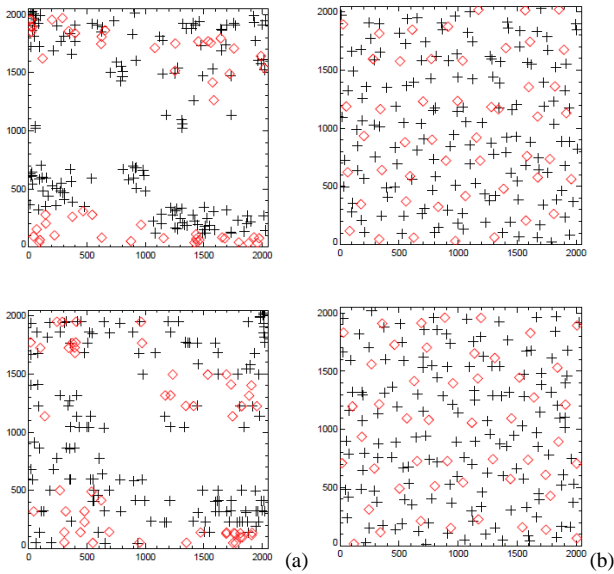


Fig. 2. (a) Distribuição das janelas de fundo de céu na corrida IRA01. As janelas amostradas a 512 segundos são representadas por cruces e as amostradas a 32 segundos por losangos (Drummond et al., 2008). O CCD foi dividido em 4 regiões apenas, o que resultou em uma alocação de janelas não homogênea. (b) Distribuição das janelas na corrida LRc01. O CCD foi dividido em 49 regiões, o que resultou em uma distribuição mais homogênea das janelas.

III. TÉCNICAS DE ESTIMAÇÃO DE FUNDO DE CÉU

Em [3] são apresentadas algumas técnicas de estimação de fundo de céu das estrelas em observação a partir dos valores observados nas janelas. Tais técnicas de estimação de fundo de céu interferem diretamente na solução que se espera obter para o problema em questão.

A. Método da janela mais próxima

Neste método, utiliza-se o ruído observado na janela mais próxima de uma estrela para estimar o ruído de fundo de céu desta estrela. Ou seja, uma janela de fundo de céu se torna *default* para cada estrela e é utilizada constantemente para a correção.

B. Método da interpolação triangular

Este método foi inicialmente proposto para casos onde não havia uma janela de fundo de céu próxima da estrela em observação. É realizada a interpolação ou extrapolação das

três janelas de fundo de céu mais próximas à estrela em observação.

C. Método do ajuste polinomial

Neste método são utilizados todos os dados disponíveis, o que oferece vários pontos extras para a correção. Pode-se usar todas as 400 janelas, sendo 200 de cada CCD.

D. Método da mediana

Este é o método utilizado atualmente pela missão CoRoT. Esta técnica calcula a mediana das 200 janelas de fundo de céu de cada CCD e este valor é considerado como o fundo de céu para todas as estrelas. Ou seja, um único valor de mediana será subtraído das curvas fotométricas de todas as estrelas em observação. Tal método é mais eficiente que os demais apresentados pelos autores, sobretudo por corrigir o problema das variações orbitais de fundo de céu, uma vez que o CoRoT mede variações fotométricas e não fotometrias absolutas.

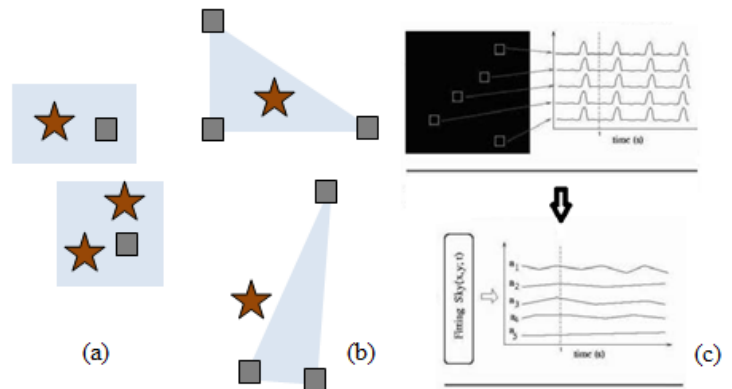


Fig. 3. (a) Método da janela mais próxima. (b) Método da interpolação triangular. (c) Método do ajuste polinomial.

IV. CCDs

O plano focal do satélite CoRoT é composto por quatro CCDs retro-iluminados, de transferência de quadro e estreitos. Cada CCD tem 2048x2048 pixels na zona de imagem e 2048x2054 pixels na zona de memória, sendo que cada pixel possui 13,5x13,5µm de tamanho, o que corresponde a 2,32 segundos de arco. O campo de visão total do satélite abrange cerca de 2,66x3,00 graus no céu. O canal de exoplanetas, no qual este trabalho é focado, utiliza dois CCDs.

A transferência de dados da zona de imagem para a zona de memória dura 0,2 segundos. O tempo de exposição real (tempo de integração de fótons) é de 31,8 segundos para o canal de exoplanetas. Sendo assim, a duração de um ciclo (integração e transferência) é de 32 segundos.

Em função da assimetria do sistema óptico do CoRoT, a difusão da luz de fundo de céu através do plano focal não é homogênea [4].

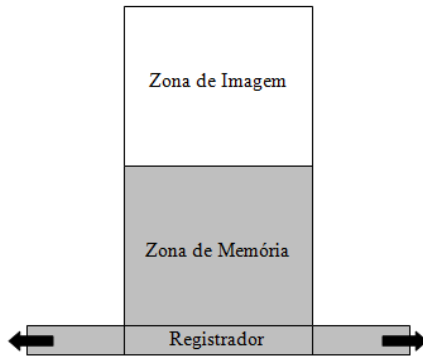


Fig. 4. Representação do CCD e as denominações de suas principais áreas.

V. O MÉTODO PROPOSTO

A. O problema

O método de estimação de fundo de céu utilizado neste trabalho é o da janela mais próxima, onde o valor do ruído de fundo de céu de uma estrela s é considerado aproximadamente igual ao valor do ruído observado na janela w espacialmente próxima de s . Quanto mais próxima for a janela w da estrela s , melhor a estimativa do fundo de céu. Para que esta estratégia funcione, é necessário que cada estrela tenha uma janela espacialmente próxima. Em outras palavras, é necessário minimizar a maior distância entre uma estrela s e a janela w mais próxima de s .

Formalizando, o problema consiste em: dadas m posições de estrelas-alvos s_j ($1 \leq j \leq m$) e um número k de janelas, é necessário encontrar a distribuição das k janelas $W = \{w_1, w_2, \dots, w_k\}$ que minimize a função custo:

$$f(W) = \text{MAX}_{i=1}^k \left[\text{MAX}_{s_j \in \eta_i} [d(w_i, s_j)] \right], \quad (1)$$

onde $d(w_i, s_j)$ é a distância euclidiana entre a janela w_i e a estrela s_j ; e η_i é o conjunto de estrelas que pertencem à “zona de influência” da janela w_i . A zona de influência η_i de uma janela w_i é a região espacial onde a janela mais próxima é w_i . Se uma estrela $s_j \in \eta_i$ então $d(w_i, s_j) \leq d(w_l, s_j)$ para $\forall l \neq i$.

No caso do satélite CoRoT, tipicamente utilizam-se $k = 200$ janelas de fundo de céu para estimar a fotometria de $m \leq 6000$ estrelas, e a imagem tem tamanho 2048x2048 pixels.

B. Algoritmo aleatório

A solução mais imediata para o problema acima é o “algoritmo aleatório”. Ele consiste em distribuir aleatoriamente as k janelas w_i dentro da imagem certo número N de vezes, gerando as distribuições aleatórias das janelas W_1, W_2, \dots, W_N . Para cada distribuição W_j , calcula-se a função custo $f(W_j)$ da equação (1) e escolhe a distribuição W^* com o menor custo. Evidentemente, este algoritmo é ineficaz.

C. Algoritmo k -médias

O problema de minimização da média das distâncias quadráticas é bem conhecido e pode ser resolvido com o algoritmo de agrupamento por k -médias (*k-means clustering*). Infelizmente, esse algoritmo não pode ser utilizado para o nosso problema, pois o nosso objetivo é minimizar a maior distância (e não a média das distâncias quadráticas). Utilizaremos k -médias como inspiração para criar o nosso algoritmo. Também compararemos os resultados de k -médias com os do nosso.

Dada uma distribuição inicial W^0 de k janelas, o algoritmo de agrupamento por k -médias otimiza iterativamente a distribuição, gerando uma solução melhor a cada iteração W^1, W^2, \dots até a convergência em W^* . Dadas m posições de estrelas-alvos s_j e a distribuição das janelas $W^t = \{w_1, w_2, \dots, w_k\}$ na iteração t , o passo do algoritmo k -médias é:

1. Para cada janela w_i na iteração t , ache o conjunto de estrelas de η_i , isto é, aquelas estrelas que estão mais próximas de w_i do que de qualquer outra janela.
2. Mova cada janela w_i para o centro geométrico das estrelas em η_i .

O algoritmo k -médias consiste em repetir os dois passos acima até a convergência. A solução final obtida W^* depende fortemente da distribuição inicial W^0 . Assim, a distribuição inicial deve ser escolhida cuidadosamente. Por exemplo, usando o algoritmo corte mediana – *median cut*. Também pode-se repetir o algoritmo k -médias para N distribuições iniciais aleatórias e escolher a solução ótima de menor custo.

D. Algoritmo k -box

O algoritmo proposto neste trabalho, chamado de k -box, é uma versão modificada do algoritmo k -médias para minimizar a maior distância em vez de minimizar a média das distâncias quadráticas. Dadas m posições de estrelas-alvos s_j , um número k de janelas e uma distribuição das janelas na iteração t $W^t = \{w_1, w_2, \dots, w_k\}$, um passo do algoritmo k -box é:

1. Para cada janela w_i da iteração t , ache o conjunto de estrelas em η_i , isto é, aquelas estrelas que estão mais próximas de w_i do que de qualquer outra janela.
2. Ache o seu *bounding box* (daqui em diante, chamaremos simplesmente *box*) das estrelas de η_i , isto é, o menor retângulo alinhado aos eixos e que contém todas as estrelas de η_i . A Figura 5 ilustra um *box*. Mova cada janela w_i para o centro geométrico do seu *box*. Isto fará (normalmente) diminuir $\text{MAX}_{j \in \eta_i} [d(w_i, s_j)]$.

O algoritmo k -box consiste em repetir os dois passos acima até a convergência. Da mesma forma que k -médias, a solução final de k -box depende fortemente da distribuição inicial de janelas W^0 escolhida. Assim, executamos repetidamente o algoritmo k -box para N distribuições iniciais aleatórias de janelas e escolhemos a melhor solução.

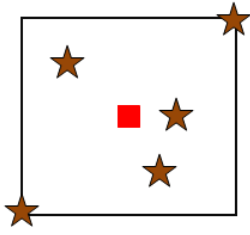


Fig. 5. *Bounding box* (ou simplesmente *box*) de uma janela (em vermelho) com as estrelas vizinhas.

E. Implementação e resultados experimentais

Comparamos a eficácia dos três algoritmos em dados simulados. Os três algoritmos foram implementados em C++ sem otimizações de velocidade.

As implementações de *k-box* e *k-médias* apresentam um gargalo: o cálculo das estrelas que pertencem à zona de influência η_i de uma janela w_i . Para isso, calculam-se as distâncias de w_i com todas as estrelas, o que é um processo computacionalmente lento. Este cálculo pode ser acelerado usando *kd-tree* ou algum outro algoritmo que acelere o cálculo do vizinho mais próximo [5].

Nos testes, geramos aleatoriamente $m = 200$ estrelas dentro de uma imagem com 1024×640 pixels. Estamos fazendo testes em imagens menores que o tamanho real do CCD, pois é mais fácil visualizar o comportamento dos algoritmos nas imagens menores. Depois, executamos o algoritmo aleatório $N=100000$ vezes e escolhemos a melhor solução. Também executamos os algoritmos *k-médias* e *k-box* $N=5000$ vezes e escolhemos a melhor solução. Estes testes foram repetidos 3 vezes. Os resultados estão na Tabela 1.

TABELA I. O custo da solução (a maior distância entre uma estrela e a janela mais próxima, em pixels) encontrada pelos três algoritmos.

	aleatório	<i>k-médias</i>	<i>k-box</i>
teste 1	131,8	95,7	85,7
teste 2	132,4	96,0	86,3
teste 3	131,9	97,7	89,2
média	132,0	96,5	87,1

Os três algoritmos levaram em média 25s, 13s e 15s respectivamente para encontrar as soluções da Tabela 1. Fica claro que o algoritmo *k-box* encontrou a melhor solução nos três testes. A Figura 6 mostra os resultados obtidos no teste 1.

Executamos o algoritmo *k-box* usando os dados reais do CoRoT e obtivemos a solução mostrada na Figura 7.

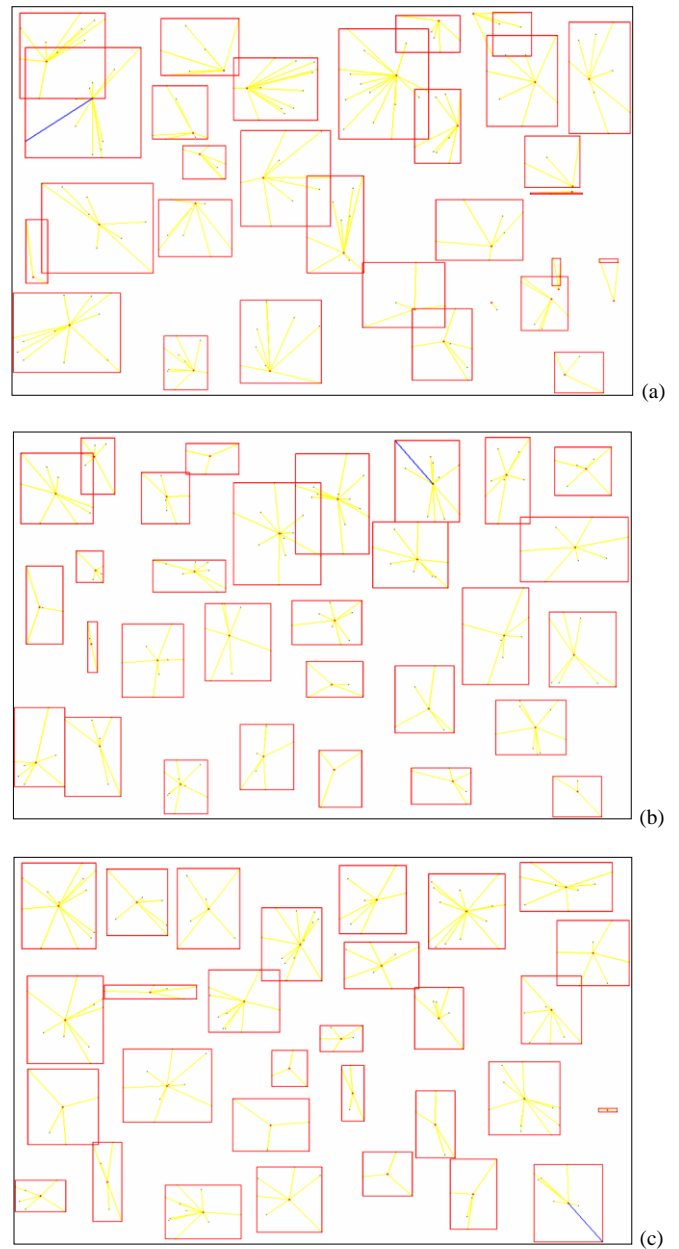


Fig. 6. Soluções obtidas pelos algoritmos aleatório (a), *k-médias* (b) e *k-box* (c). Os retângulos vermelhos mostram os *boxes* das janelas. O ponto vermelho no centro de um *box* é a janela. Os pontos azuis são as estrelas. As linhas amarelas ligam uma janela a todas as estrelas vizinhas. A linha azul é a maior distância entre uma estrela e a sua janela mais próxima.

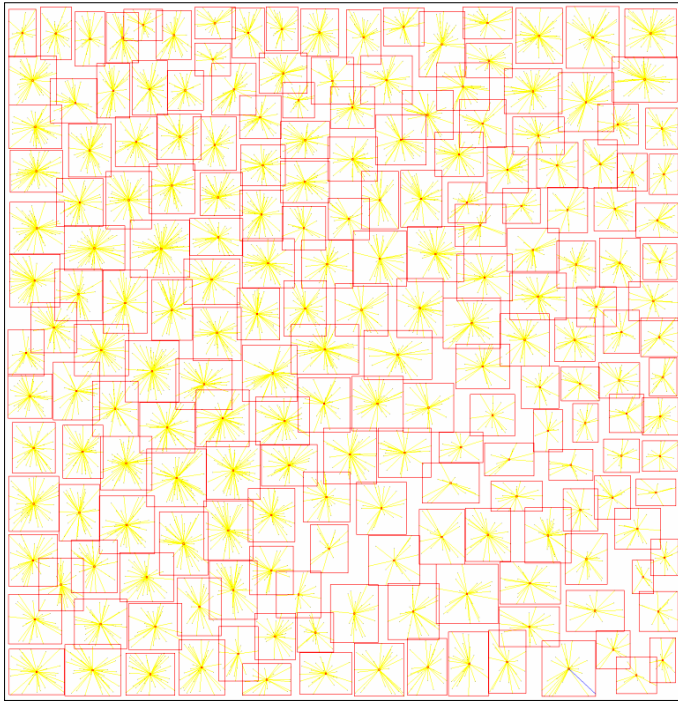


Fig. 7. Solução do k -box utilizando os dados do CoRoT, com 200 janelas de fundo de céu, 5724 estrelas (posições reais no CCD) e imagem de tamanho 2048x2048 pixels. Para esta simulação foram testadas 100 distribuições iniciais de janelas.

VI. TRABALHOS FUTUROS

A. Problema com pesos

Das até 6000 estrelas observadas a cada corrida pelo satélite, nem todas têm a mesma relevância. Um dos fatores a ser considerado para diferenciar a relevância das estrelas é, por exemplo, a intensidade luminosa das estrelas. O intuito de utilizar as intensidades luminosas das estrelas como pesos é garantir que a janela de fundo de céu seja alocada mais próxima das estrelas que têm mais importância para a missão CoRoT. Considerando que a cada estrela s_j ($1 \leq j \leq m$) está associada um peso p_j , o nosso problema com peso é encontrar a distribuição das k janelas $W = \{w_1, w_2, \dots, w_k\}$ que minimize a função custo:

$$f(W) = \text{MAX}_{i=1}^k \left[p_j \times \text{MAX}_{j \in \mathcal{I}_i} [d(w_i, s_j)] \right], \quad (2)$$

onde todas as variáveis significam o mesmo que na equação (1). Ainda não está claro como minimizar esta função.

B. Alocação real das janelas

Na prática, pode não ser possível utilizar as posições das janelas determinadas pelo k -box, pois nessas posições podem existir pixels queimados ou estrelas de pequeno brilho. Assim, o próximo passo deste trabalho é determinar a real posição das janelas de fundo de céu. A posição real da janela deverá ser a mais próxima possível da posição determinada pelo método, em espaços livres de estrelas parasitas e de pixels mortos. Ou seja, o ponto de mínimo local mais próximo possível da posição da janela determinada por este método.

C. Objetivo e continuação deste trabalho

Por fim, queremos comparar os resultados que serão obtidos com os resultados apresentados em [6], onde se decidiu o método da mediana, atualmente em utilização pela missão CoRoT.

Espera-se que este trabalho possa apresentar uma alternativa quanto ao método em uso. Espera-se também que este trabalho possa contribuir para a missão PLATO que também utilizará janelas de fundo de céu para a estimação de ruído do fundo de céu. O satélite PLATO deverá ser uma espécie de continuação da missão CoRoT, com capacidade científica muito maior que o anterior.

AGRADECIMENTOS

Este trabalho tem financiamento da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Agradecemos também ao Observatório de Paris-LESIA pela disponibilização de dados e imagens da missão CoRoT.

REFERÊNCIAS

- [1] L., Pinheiro da Silva. "Caractérisation Instrumentale Appliquée à La Photométrie Très Haute Précision dans le Cadre de la Mission Spatiale CoRoT", *Tese de Doutorado – Universidade de Toulouse*, França, 2006.
- [2] M. Auvergne, "The CoRoT satellite in flight: description and performance", *Astronomy & Astrophysics*, v.506, n. 1, pp. 411-424, 2009.
- [3] R., Drummond et al., "Correcting for Background Changes in CoRoT Data", *Astronomy & Astrophysics*, v. 487, n. 3, pp. 1209-1220, 2008.
- [4] R., Samadi et al., "Extraction of the Photometric Information: Corrections", *CoRoT Book*, pp.317-325, 2007.
- [5] R. Sedgewick, "Algorithms", *Addison Wesley*, 1986.
- [6] R., Drummond, "Calibration of the CoRoT Space Mission – Correction of the Photometry for Pointing Jitter and Background Changes", *Tese de Ph.D. - Katholieke Universiteit Leuven*, Bélgica, 2007.