

## ESTRUTURAÇÃO DE UM ALGORITMO BASEADO EM BIG DATA E TÉCNICA DE TOMOGRAFIA PARA ANÁLISE DA AMOSTRA DE SOLO AGRÍCOLA

GABRIEL MARCELINO ALVES<sup>1</sup>, PAULO ESTEVÃO CRUVINEL<sup>2</sup>

<sup>1</sup> Doutorando em Ciência da Computação, UFSCar/São Carlos-SP, (19) 3634-1100, gabriel.marcelino@ifsp.edu.br

<sup>2</sup> Doutor em Engenharia, Pesquisador, Embrapa Instrumentação/São Carlos-SP, (16) 2107-2800, paulo.cruvinel@embrapa.br

Apresentado no

XLIV Congresso Brasileiro de Engenharia Agrícola - CONBEA 2015

13 a 17 de setembro de 2015- São Pedro – SP, Brasil

**RESUMO:** Na agricultura do século XXI tem sido verificado, para o processo de auxílio à tomada de decisão, significativo aumento de dados e informações provenientes de solos, plantas, clima, sistemas de manejo, bem como dos insumos utilizados, entre outros. O termo *Big Data* pode ser associado a este cenário como a capacidade de extrair informações de diferentes fontes inferindo novas linhas de raciocínio e integração entre sistemas. É possível observar iniciativas que empregam *Big Data* a fim de aprimorar processos de tomada de decisão em diversas áreas, a exemplo da agrícola. Este trabalho tem por objetivo apresentar uma estrutura algorítmica com o emprego de *Big Data* voltado para reconstrução 3D de imagens tomográficas de Raios X de alta resolução para a análise física de amostras de solos agrícolas. Resultados mostram que a estrutura baseada no modelo de programação *MapReduce* utilizado em ambientes *Big Data* para seleção das projeções tomográficas e o algoritmo de reconstrução *Filtered Back-Projection* (FBP) proporciona escalabilidade bem como a possibilidade de redução de tempo no processo de reconstrução de imagens e de análise de um maior número de amostras de solo em um mesmo tempo considerado. Esta abordagem pode fornecer a capacidade de escala. Finalmente, pode ser observada como um resultado de que a utilização de uma solução heurística pode habilitar a modelagem computacional de trazer oportunidades para sua utilização, principalmente em problemas que envolvem a utilização de engenharia na agricultura.

**PALAVRAS-CHAVE:** Tomografia agrícola, Big Data, Precisão na agricultura.

### AN ALGORITHM BASED ON BIG DATE AND TOMOGRAPHIC RECONSTRUCTION FOR THE ANALYSIS OF AGRICULTURAL SOIL SAMPLES

**ABSTRACT:** In the twenty-first century the agriculture is facing the needs for improvements in the processes for decision making, since has being observed a significantly increasing on the amount of data and information, which come from a broad number of different sources, i.e., from soils, plants, climate, management systems, as well as the inputs, among others. The term *Big Data* can be associated with this scenario and it is associated with the ability to extract information from different sources, as well as inferring new lines of reasoning and integration between information from different systems. This paper aims to present an algorithmic structure based on the use of *Big Data* and X-ray tomographic 3D image reconstruction of high-resolution for physical analysis of agricultural soil samples. The structure consists of adopting the *MapReduce* programming model for the selection of the tomographic projections and the use of an image reconstruction algorithm based on the *Filtered Back-Projection* (FBP). This approach can provides the ability to scale, as well as the possibility of reducing time in the image reconstruction process, i.e., increasing the speed of the soil samples analysis in time. Finally, was observe as a result that the use of a heuristic solution enabled the computational modeling discussing and its opportunities of use, mainly in terms of its usability in problems involving the use of engineering in agriculture.

**KEYWORDS:** Agricultural tomography, Big data, Precision in agriculture.

**INTRODUÇÃO:** A Tomografia Computadorizada (TC) encontra aplicações em diversas áreas. Na agricultura ela encontra destaque como método de análise e investigação na física dos solos, ou ainda, na possibilidade de avaliar plantas e árvores vivas (CRUVINEL *et. al.*, 1990; PIRES *et. al.* 2010; PEREIRA & CRUVINEL, 2015). Inserida nesse contexto, a reconstrução 3D configura-se como relevante instrumento na TC, uma vez que possibilita a análise volumétrica do interior de um corpo ou objeto de forma não invasiva embora se verifique dificuldades em obter modelos 3D de objetos de forma automática, rápida e precisa (CRUVINEL *et. al.*, 2009; KHILAR *et. al.*, 2013). Por outro lado, o emprego de *Big Data* no processo de reconstrução foi recentemente iniciado, principalmente na tratativa das informações a serem reconstruídas tridimensionalmente e no desenvolvimento de novos algoritmos (ZHAO *et. al.*, 2013; DITTER *et. al.*, 2014). O termo *Big Data* está associado com a capacidade de extrair informações de diferentes fontes inferindo novas linhas de raciocínio e integração entre sistemas, além de buscar novas maneiras de analisar dados superando os métodos tradicionais de análise da informação nas diversas áreas do conhecimento, a exemplo da agricultura e, para isto, uma das técnicas empregadas é o modelo de programação para processamento paralelo e distribuído de dados conhecido como *MapReduce* (DEAN & GHEMAWAT, 2008; SAGIROCLU & SINANC, 2013; HIRAFUJI, 2014). Este trabalho tem por objetivo apresentar uma estrutura algorítmica para o emprego de *Big Data* em tomografia computadorizada de Raios-X de alta resolução para a análise física de amostras de solos agrícolas.

**MATERIAL E MÉTODOS:** Para a estruturação do algoritmo foi considerada a arquitetura apresentada na Figura 1 que visa à análise de amostras de solos agrícolas com emprego de *Big Data*.

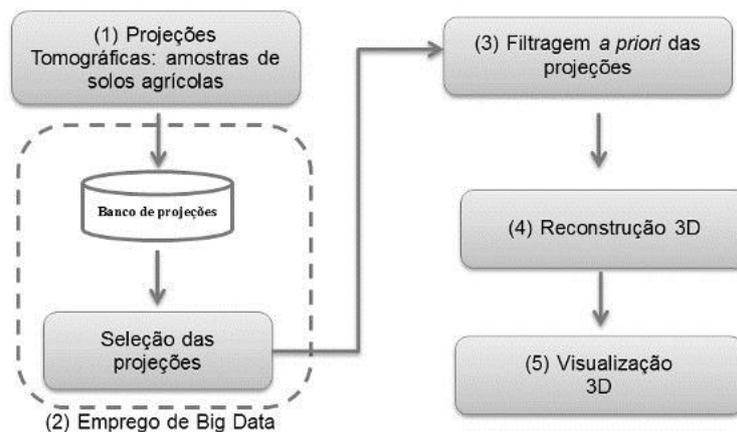


FIGURA 1. Diagrama de blocos da arquitetura para estruturação do algoritmo de análise de amostras de solos agrícolas com o emprego de *Big Data*.

A primeira etapa refere-se à aquisição de projeções oriundas de amostras de solos agrícolas a partir de tomógrafos para uso na área de agricultura instalados na Embrapa Instrumentação (Embrapa/CNPDI), localizada na cidade de São Carlos/SP. Salienta-se que foi considerada a coleta de informações em tomógrafos de alta resolução cuja ordem de grandeza das matrizes de dados para cada amostra é de 1,5 GigaByte. A segunda etapa consiste no armazenamento das projeções obtidas em ambiente computacional voltada para *Big Data*. Essa etapa deverá ser realizada de forma paralela adotando o modelo *MapReduce*, que mapeia os arquivos de entrada para um conjunto de nós que, após processados, geram um conjunto intermediário de dados que é dividido e replicado para as tarefas de redução em que cada nó consome os dados obtidos na fase intermediária. Portanto, a função “*Map*” será responsável por analisar as projeções e gerar, no passo intermediário, a lista de candidatas a serem submetidas ao processo de reconstrução. A função “*Reduce*” define o conjunto de projeções por meio da seleção a partir da lista produzida. Uma vez selecionado o conjunto, a terceira etapa consiste em filtrar as projeções selecionadas a fim de prepará-las para a reconstrução. Nesse passo deverão ser avaliados critérios que identifiquem a necessidade da filtragem, bem como medidas que caracterizem o resultado da aplicação do filtro na imagem. A etapa seguinte consiste na reconstrução 3D da imagem proveniente de amostras agrícolas por meio do algoritmo *Filtered Back-Projection* (FBP), o qual

agrupa as projeções filtradas no passo anterior, realizando a retroprojeção e agrupando os pixels reconstruídos. A última etapa refere-se à visualização 3D que consiste em preparar a imagem para observação por meio de interface computacional, em modo gráfico, que possibilite a extração de medidas e visualizações de faixas de interesses das amostras e cuja operação pelo usuário final é de fácil compreensão e interação. Foram selecionados para as primeiras análises e primeiros estudos os Latossolos Vermelhos Distróficos (LVd), os quais são solos minerais com teores de  $\text{Fe}_2\text{O}_3$  entre 8% e 18%, de textura argilosa e porosidade alta. Foram coletadas amostras indeformadas na camada superficial do solo (primeiros 10,0 cm), utilizando cilindros de PVC, com borda biselada, de 6,4 cm de altura, 3,7 cm de diâmetro interno e volume aproximado de  $68,8 \text{ cm}^3$ . As imagens tomográficas foram obtidas utilizando um sistema de microtomografia de raios-X, modelo 1172 da SkyScan, composto por um tubo de raios-X de microfoco com fonte de alta tensão (100 kV), um porta-amostra com manipulador de precisão e um detector baseado em cintilador de óxido de gadolínio e câmera CCD com resolução de 10Mpixels (4000 x 2300 pixel) conectados a um computador com algoritmos para seleção de projeções, controle e aquisição das imagens.

**RESULTADOS E DISCUSSÃO:** A estruturação do algoritmo por meio da arquitetura apresentada oferece uma abordagem para o emprego de *Big Data* em aplicações de tomografia computadorizada, particularmente na área agrícola. O modelo *MapReduce* é voltado para computação paralela e possibilita abstrair, dos desenvolvedores, a complexidade da paralelização como distribuição dos dados, tolerância a falhas e balanceamento de carga permitindo que eles se dediquem à elaboração das funções “*Map*” e “*Reduce*” o que simplifica o processo de construção do sistema de reconstrução 3D de imagens tomográficas. Quanto à reconstrução, o algoritmo FBP permite iniciar o cálculo imediatamente após a primeira imagem de projeção estar disponível no ambiente o que o torna adequado para a paralelização, pois propicia distribuir a computação tornando a solução escalável para conjuntos maiores de dados. A Figura 2 apresenta a visão geral do algoritmo do FBP adequado ao modelo *MapReduce*.

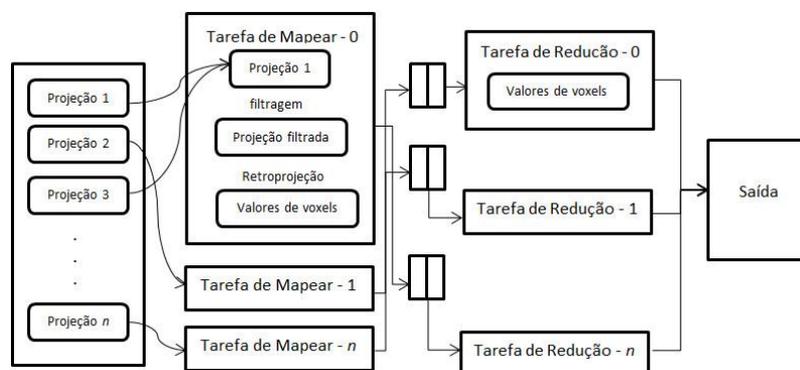


FIGURA 2. Visão geral do algoritmo FBP adequado ao modelo *MapReduce*.

As projeções são lidas do local de armazenamento e são alocadas pelas tarefas de mapeamento que realizam a filtragem e a retroprojeção e transferem voxels individuais para a fase intermediária. As tarefas de redução acumulam todos os valores intermediários produzindo o volume final reconstruído. A Figura 3 apresenta exemplo de imagens tomográficas de amostras de solo agrícola, para os dois sistemas de manejo considerados e reconstruídas com base nesse processo. A técnica utilizada no processamento de imagens tomográficas de alta resolução é útil para detectar variações tanto em macro porosidade como em micro porosidade de amostras de solos agrícolas, bem como auxiliar na caracterização de sistemas de manejo. Além disso, pode possibilitar uma visualização tridimensional, podendo contribuir para o conhecimento de como são formados os sistemas de redes de poros do solo, assim como auxiliar no entendimento de como a estrutura do solo pode afetar a retenção de água e sua dinâmica no solo.

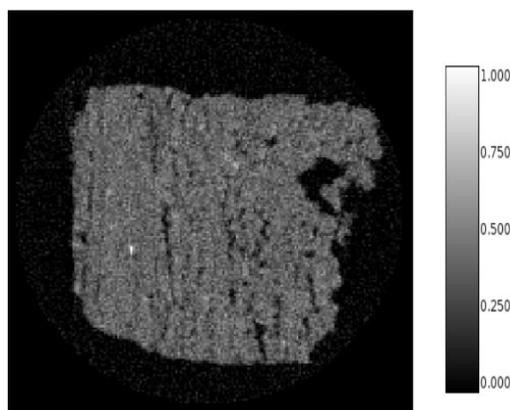


FIGURA 3. Imagem tomográfica utilizada no processo de reconstrução 3D considerando uma amostra de Latossolo Vermelho Distrófico (LVd). A escala esta calibrada em densidade, ou seja, em  $\text{g/cm}^3$ .

Os algoritmos paralelos agilizam o processo da análise uma vez que reduzem o tempo geral necessário para a construção das imagens e sua interpretação. Tais algoritmos possibilitam explorar os potenciais de processamento das arquiteturas *multicores* e das placas gráficas.

**CONCLUSÕES:** Resultados deste trabalho apresentam a estrutura de um modelo de reconstrução 3D de imagens tomográficas provenientes de amostras de solos agrícolas empregando a abordagem de *Big Data* em sua concepção. Essa abordagem oferece a possibilidade de escalabilidade bem como a possibilidade de redução de tempo no processo de reconstrução de imagens, viabilizando um maior número de análises em uma mesma janela de tempo disponível.

## REFERÊNCIAS

- CRUVINEL, P. E.; CESAREO, R.; CRESTANA, S.; MASCARENHAS, S. **X- and Y-Rays Computerized Minitomograph Scanner for Soil Science**. IEEE Transactions on Instrumentation and Measurement, vol. 39, nº 5. 1990.
- CRUVINEL, P. E.; PEREIRA, M. F. L.; SAITO, J. H.; COSTA, L. F. **Performance Improvement of Tomographic Image Reconstruction based on DSP processors**. IEEE Transactions on Instrumentation and Measurement, vol. 58, nº 9, p. 3295-3304. 2009.
- DEAN, J.; GHEMAWAT, S. **MapReduce: simplified data processing on large clusters**. Communications ACM, vol. 51, p. 107-113. 2008.
- DITTER, A.; FEY, D.; SCHON, T.; OECKL, S. **On the way to Big Data applications in industrial computed tomography**. IEEE International Congress on Big Data, p. 792-793. 2014.
- HIRAFUJI, M. **A strategy to create agricultural Big Data**. IEEE SRII Global Reference, p. 249-250. 2014.
- KHILAR, R.; CHITRAKALAR, S.; SELVAMPARVATHY, S. N. **3D Image Reconstruction: techniques, applications and challenges**. IEEE Proceedings of International Conference on Optical Imaging Sensor and Security. India. 2013.
- PEREIRA, M. F. L.; CRUVINEL, P. E. **A model for soil computed tomography based on volumetric reconstruction, Wiener filtering and parallel processing**. Elsevier. Computers and Electronics in Agriculture, p. 151-163. 2015.
- PIRES, L.; BORGES, J.; BACCHI, O.; REICHARDT, K. **Twenty-five years of computed tomography in soil physics: a literature review of the Brazilian contribution**. Soil and Tillage Research, 110 (2), p. 197-210. 2010.
- SAGIROGLU, S.; SINANC, D. **Big Data: a review**. IEEE International Conference on Collaboration Technologies and Systems. 2013.
- ZHAO, J. FU, Y.; TAN, Y.; CAO, F. **A reduction algorithm for the Big Data in 3D surface reconstruction**. IEEE International Conference on Systems, Man and Cybernetics, 2013.