

Grupo 5

Nome (RA):

Lucas Alves Racoci (156331)

Luiz Fernando Rodrigues da Fonseca (156475)

Rafael Zane (157079)

Rodrigo Noronha Máximo (157209)

Tema

Utilização do Spark para distribuir a aplicação do algoritmo Sobel, de detecção de bordas em imagens, e também do filtro da mediana.

Plano de trabalho

1. Resultados e testes iniciais

Foi instalado o Apache Spark e o Pyspark para poder utilizar as funcionalidades do Spark no Python. Como o trabalho envolve imagens, foi instalado também o OpenCV, que proporciona funcionalidades de processamento de imagens. Por fim, foi instalado o Jupyter Notebook, que é uma aplicação web, open-source, que permite criar e compartilhar documentos, com o intuito de facilitar a execução dos códigos por nós utilizados.

Primeiramente realizamos a leitura de imagens transformando-as em RDD de modo que fossem lidas como arquivos texto. Em sequência, foram feitos testes de leitura de imagens no HDFS em formato binário, onde foi realizada a conversão do binário para uma matriz Numpy, com o auxílio do OpenCV.

2. Próximas etapas

O algoritmo de Sobel utiliza-se de duas máscaras (matrizes quadradas) para detectar bordas nas imagens.

Para a próxima etapa, o modelo de maps e reduces adotado será da seguinte forma:

- Serão dadas diversas imagens como entrada.
- As imagens serão quebradas em pequenos blocos, que serão passados para os mappers.
- Será utilizado como chave o identificador da imagem a qual o bloco pertence, de modo a serem identificadas corretamente durante a etapa de reduce.
- Os mappers irão executar parte do algoritmo do Sobel em cada bloco.
- Os reducers vão juntar os blocos com mesma chave, e compor a imagem completa com as bordas detectadas.