

## Enriquecimento Semântico dos Dados do Tribunal de Justiça de Santa Catarina

Eduardo Machado Carboni

Visão Computacional – Programa de Pós Graduação em Ciência da Computação - Universidade Federal de Santa Catarina / Tribunal de Justiça de Santa Catarina

### Introdução

O Tribunal de Justiça de Santa Catarina possui uma extensa base de dados com *terabytes* de informações armazenadas em alguns banco de dados relacionais.

Através de metodologias de Visão computacional seria possível utilizar os dados que a instituição possui e os que virá a possuir para enriquecer semanticamente sua base de dados e auxiliar na tomada de decisões.

### Problema

Embora existam muitos dados no Tribunal de Justiça de Santa Catarina que possuem um grande potencial de exploração, eles podem ser ainda mais aproveitados.

Pode-se inferir que talvez através de *Data Mining* e *Data Science* "a iniciativa privada extrai mais conhecimento destes dados do que o próprio TJSC".

Quanto mais semântica for inserida aos dados do TJSC, mais úteis eles vão ser para a própria instituição.

### Proposta

Buscar dados brutos na base de dados do TJSC e, utilizando técnicas de Visão Computacional, aumentar semanticamente sua informação.

Obter o endereço relacionado às partes dos processos e classificá-los em Urbano ou Rural.

A classificação proposta daria maior semântica aos dados de processos possibilitando um agrupamento de casos em diferentes classes, como exemplo as propostas, sendo urbana ou rural.

### Desenvolvimento – Busca de Dados

Os dados de origem são de uma base MySQL do sistema judicial com milhões de registros de endereços. Na figura 1 é demonstrada superficialmente uma consulta que busca o endereço das partes dos processos.

```

1 SELECT pr_id_processo
2       ,em_ingradosou
3       ,em_num_ingradouso
4       ,em_bairro
5       ,em_des_localidade_endereco
6 FROM pr_id_pr
7      ,em_pp_id_parte_pp
8      ,em_pp_id_en
9 WHERE pr_id_pr = processo
10      and pp_id_pr = _essoa;
    
```

id_processo	ingradouso	num_ingradouso	bairro	des_localidade_endereco
1	Servidão Variable Maria Mecca...	0	Arruá da Fátima	Paflopa
2	Rua 17 DE CEZARIO	2	SANTA ELIZABETHA	Antônio Carlos
3	Rua 700 E	1	Várzea	Itapema
4	Rua Morada do Sol	2	Parque	Porto Belo
5	Rua 04 de Novembro	1	Centro	Antônio Carlos
6	Estrada Geral Rachadel	0	Rachadel	Antônio Carlos
7	Rua João Crescencio de Silva	0	Berra do Araruá	Paflopa
8	Rua São João	2	Ponte do Itanaraí	Paflopa
9	Rua Carlos Manoel Sampaio T...	1	Cambá	Governador Celso Ramos
10	Rua São André Vieira	0	São Sebastião	Paflopa
11	Rua dos Arcangjos	1	Canoinha Nova	São José
12	Rua Honório de Miranda Gomes	1	Fundo	Biguaçu
13	Rua Manoel Justino de Oliveira	1	Boa Vista	Biguaçu

Figura 1 – Consulta na base de dados

### Desenvolvimento – Encontrar Mapa

São buscadas as imagens de satélite do endereço através de um Jupyter Notebook em Python

Busca pelo endereço  
Encontra as coordenadas (Latitude e Longitude)  
Salva em um diretório com referência ao processo



Figura 2 – Busca de imagem de endereço

Figura 3 – Busca de imagem de endereço

### Desenvolvimento – Desfocar Mapa

Tendo os mapas encontrados, iniciou-se pela aplicação do filtro de mediana do OpenCV com *ksize* 11 para desfocar as imagens conforme abaixo:

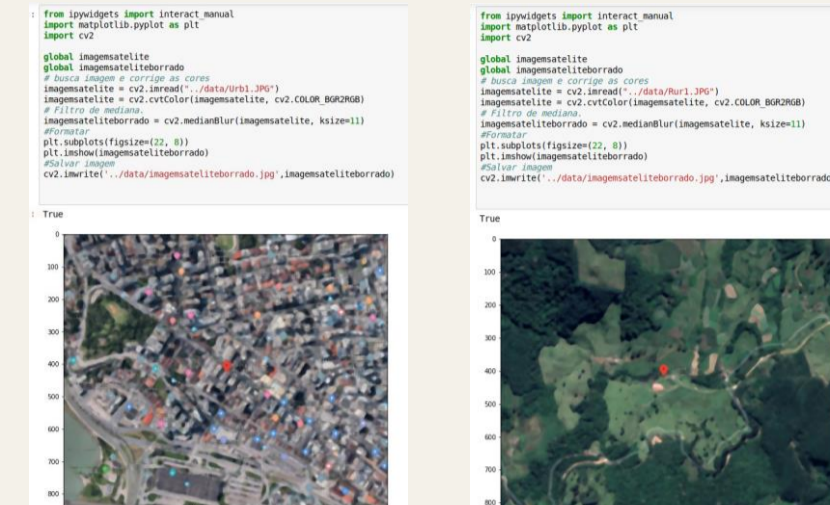


Figura 4 – Código e desfoco de imagem

Figura 5 – Código e desfoco de imagem

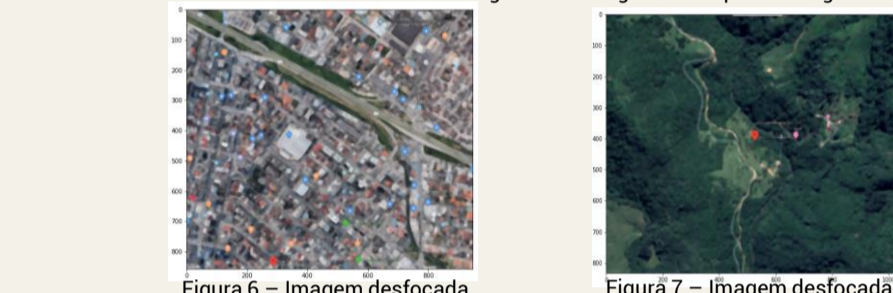


Figura 6 – Imagem desfocada

Figura 7 – Imagem desfocada

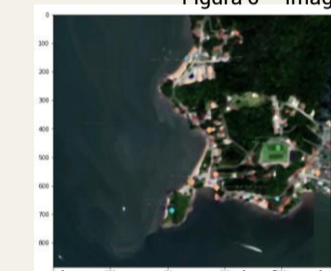


Figura 8 – Imagem desfocada



Figura 9 – Imagem desfocada

### Desenvolvimento – Encontrar Bordas

A detecção das bordas foi realizada através da implementação de Canny com o OpenCV tendo os seguintes parâmetros: (threshold1=50, threshold2=300, apertureSize=3, L2gradient=True)



Figura 10 – Código e bordas de imagem



Figura 11 – Código e bordas de imagem

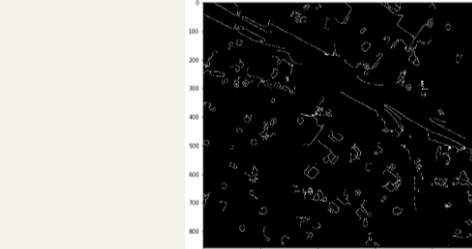


Figura 12 – Bordas de imagem

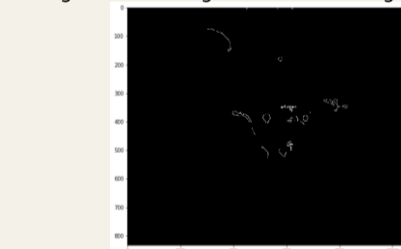


Figura 13 – Bordas de imagem

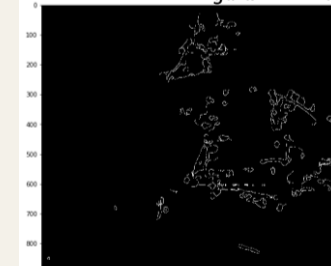


Figura 14 – Bordas de imagem

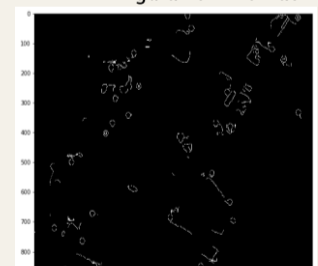


Figura 15 – Bordas de imagem

### Desenvolvimento – Mapas

Neste trabalhos foram utilizadas as imagens de satélites híbridas e desfocadas. Porque? Abaixo a comparação com as outras opções

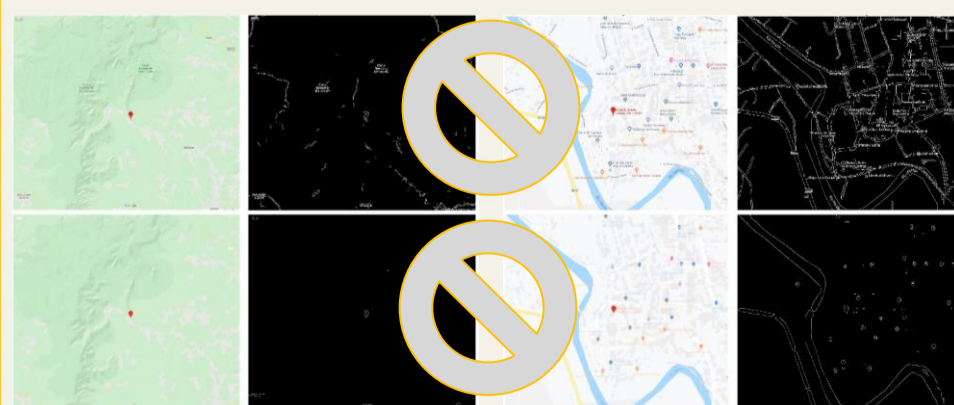


Figura 16 – Bordas de imagem original e borrada

Figura 17 – Bordas de imagem original e borrada

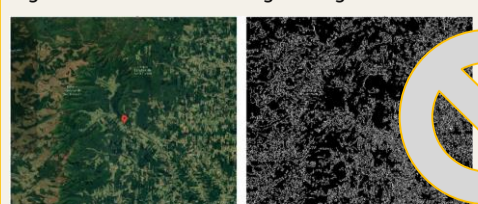


Figura 18 – Bordas de imagem original e borrada

Figura 19 – Bordas de imagem original e borrada

### Desenvolvimento – Classificar Mapas

Nos testes foi escolhido o valor 3000 para ser o limiar entre Área Rural e Área Urbana. São contados os pixels brancos da imagem e, se for maior que 3000 é Área Urbana. Se for menor é Área Rural:

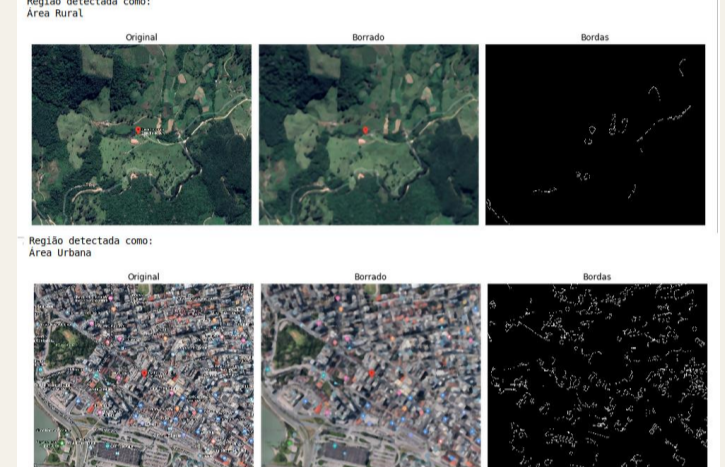
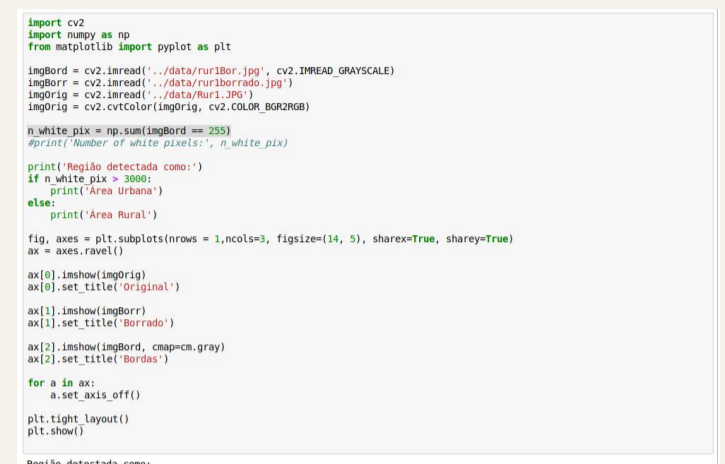


Figura 20 – Código, imagens e identificação de áreas

### Resultados e Conclusões

Os endereços testados foram classificados em área urbana ou área rural normalmente. Abaixo outros resultados:

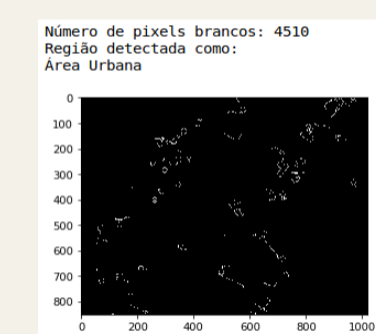


Figura 21 – Identificação de área Urbana

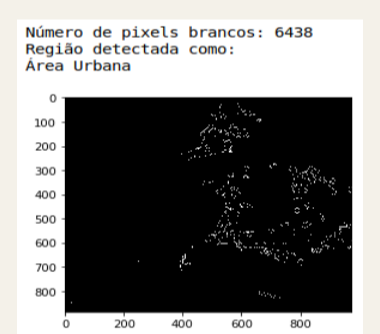


Figura 22 – Identificação de área Urbana

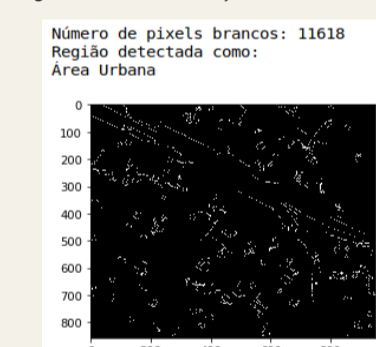


Figura 23 – Identificação de área Urbana

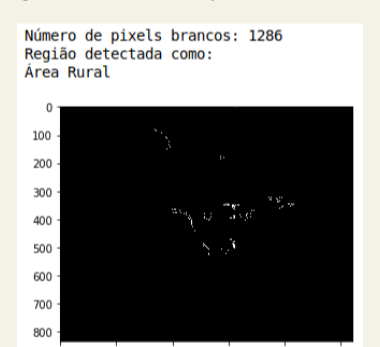


Figura 24 – Identificação de área Rural

Tendo o resultado demonstrado, é possível constatar que o objetivo de classificar os endereços em áreas foi atingido com sucesso.

### Trabalhos Futuros

Os trabalhos futuros deste projeto vão além do que pode ser explorado com os conceitos de visão computacional, mas nesta linha é possível destacar o seguinte:

Através da obtenção de imagens do Google Street View, cruzar com informações de especialistas e classificar os locais de acordo com as aparências sociais de suas construções.

Reconhecer imagens de envolvidos nos processos em postagens em redes sociais no período dos fatos narrados nos registros, encontrando locais, emoções e companheiros das pessoas.

### Contato

Aluno:

Eduardo Machado Carboni

Email: [eduardocarboni@gmail.com](mailto:eduardocarboni@gmail.com)

Poster:

