

# Segmentação de texto e operações morfológicas

Nathália Harumi Kuromiya  
Engenharia da Computação - Graduação  
E-mail: n175188@dac.unicamp.br  
RA 175188

22 de maio de 2019

## 1 Introdução

Esse trabalho tem o objetivo de mostrar como a aplicação de operações morfológicas podem auxiliar no reconhecimento de texto em imagens a partir do reconhecimento e classificação de componentes. Para a elaboração desse trabalho, foi necessário um estudo maior sobre operações morfológicas, incluindo dilatação, erosão e fechamento. Para esse fim, utilizou-se o livro[1] e os slides de morfologia[2], disponibilizados pelo professor *Hélio Pedrini*.

## 2 Trabalho Proposto [3]

O trabalho proposto se resume em: dada uma imagem de entrada no formato .PBM, o algoritmo reconhece as linhas e palavras contidas nessa imagem e produz duas imagens de saída: uma com as linhas destacadas, outra com as palavras destacadas. Essas duas imagens também estão no formato .PBM. O programa também retorna o número de linhas e palavras reconhecidos na imagem.

Para efeito de estudos, o algoritmo produzido também retorna imagens .PNG com o mesmo conteúdo, porém com os destaques coloridos.

### 3 Programas e bibliotecas utilizadas

A versão de Python[4] utilizada para implementação foi 3.5.4 e as bibliotecas de auxílio foram NumPy[5] 1.11.2 e OpenCV[6] 3.1.0.

### 4 Entradas e Saídas

A imagem de entrada e as de saída possuem o formato .PBM (*Portable Bit-Map*). Como citado anteriormente, para efeitos de estudo, foram produzidas imagens .PNG (*Portable Network Graphics*).

## 5 Implementação e Resultados

### 5.1 Imagem-referência

Para comparação das técnicas utilizadas, usaremos a Figura 1 como base para todos os experimentos.

310

IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION, VOL. 7, NO. 3, JUNE 1991

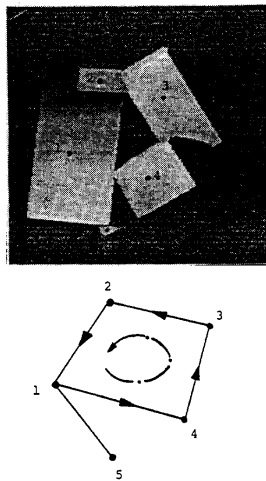


Fig. 4. Range image of an AMBIGUOUS scene and the corresponding graph.

sensory feedback is carried out in a local reflexive mode rather than in a planned mode with one exception, that is, when a pathological state is detected.

3) *States*: This is a finite set of states describing the environment of the Turing machine as perceived by the sensors. If new sensors are added, the set of states is partitioned to describe the scene as perceived by the additional sensors. For example, if a sensor capable of determining the “touch” relations of objects in the scene is added, then the set of five states, can be partitioned (a finer partition) to describe both the “touch” and “on-top-of” relations. The states of the machine are:

- |            |   |
|------------|---|
| Empty      | If there are no vertices in the diagraph, i.e., an empty diagraph.  |
| Dispersed  | If there no edges in the diagraph, i.e., a null diagraph (Fig. 2).  |
| Overlapped | If there are at least two vertices connected with an edge (Fig. 3).   |
| Ambiguous  | If there is one or more directed cycles in the diagraph (Fig. 4).   |
| Unstable   | This category is not tested by the analysis of the graph but through analysis of the contact point/line of the object with the support plane. If this contact is a point or a line, it is classified as unstable. See Fig. 5. |

Figura 1: Figura-referência

## 5.2 Segmentação de linhas

Para a segmentação de linha, começamos pelo pré-processamento da imagem original a fim de encontrar as componentes que formam as linhas.

### 5.2.1 Pré-processamento

Para o pré-processamento, utilizou-se as funções *dilate*, *erode* e *bitwise\_not* do OpenCV[6].

Primeiramente, a imagem original foi invertida, dado que a biblioteca utilizada usa como valor-referência para pretos e brancos o inverso do apresentado pela imagem de extensão .PBM.

Após, dado os elementos estruturantes  $e_1$ ,  $e_2$  e  $e_3$ , em que:

$e_1$  = elemento estruturante de 1 pixel de altura e 100 de largura

$e_2$  = elemento estruturante de 200 pixels de altura e 1 de largura

$e_3$  = elemento estruturante de 1 pixel de altura e 30 de largura

Foram aplicados:

- Um fechamento (dilatação seguido de erosão) entre a imagem original  $I$  e  $e_1$  ( $I_1 = I \bullet e_1$ );
- Um fechamento entre  $I$  e  $e_2$  ( $I_2 = I \bullet e_2$ );
- A intersecção entre os resultados dos dois fechamentos ( $I_3 = I_1 \cap I_2$ );
- A intersecção foi submetida a outro fechamento com  $e_3$  ( $I_f = I_3 \bullet e_3$ ).

O resultado de todas essas operações resultou nas componentes ilustradas na Figura 2.

### 5.2.2 Pós-processamento

Após o processamento, utilizou-se a função *getConnectedComponentsWithStats* do OpenCV[6] para obter as coordenadas das componentes mostradas na figura 2. Com essas componentes selecionadas, era necessário escolher quais componentes contém texto e quais não. Para isso, foi necessária a criação de regras de filtragem.

A primeira regra é baseada na razão de pixels pretos pela quantidade de pixels total em cada componente.

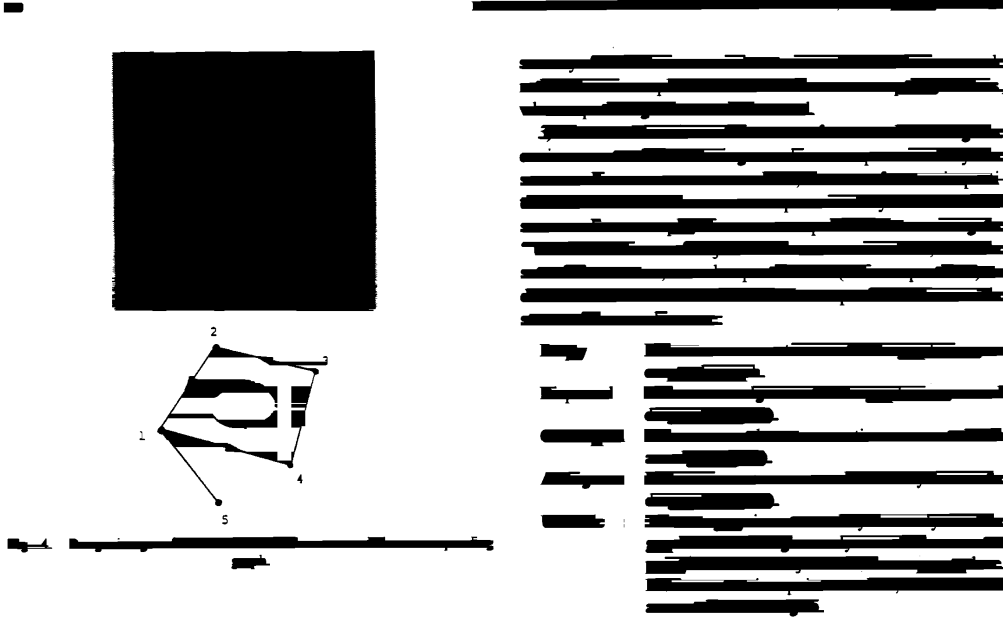


Figura 2: Componentes encontradas depois do pré-processamento

As taxas aceitas foram fixadas em:

$$0.5 < BP_{ratio} < 0.8$$

A obtenção dessas taxas foi experimental. Com uma taxa mínima maior, alguns números da imagem não foram reconhecidos como texto e com uma taxa máxima menor, muitas linhas foram descartadas.

A segunda regra foi baseada na razão entre transições de pixel (branco-preto) pela quantidade de pixels pretos na componente.

As taxas aceitas foram:

$$0.1 < TR_{ratio} < 0.6$$

Da mesma forma que a taxa de pixels pretos  $BP_{ratio}$  foi experimental, a taxa de transições também. Com uma taxa mínima maior, várias linhas foram consideradas não-texto. Já uma taxa máxima menor desconsiderou alguns números da imagem.

A partir dessa filtragem, as componentes restantes estão apresentadas na Figura 3 e mostram as **41 linhas** obtidas na imagem.

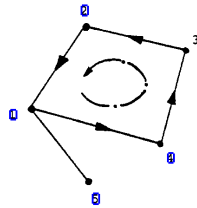
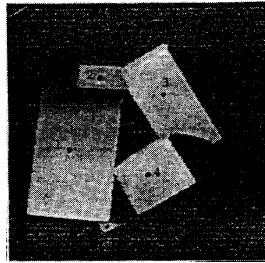


Fig. 4. Range image of an AMBIGUOUS scene and the corresponding graph.

sensory feedback is carried out in a local reflexive mode rather than in a planned mode with one exception, that is, when a pathological state is detected.

3) *States*: This is a finite set of states describing the environment of the Turing machine as perceived by the sensors. If new sensors are added, the set of states is partitioned to describe the scene as perceived by the additional sensors. For example, if a sensor capable of determining the "touch" relations of objects in the scene is added, then the set of five states, can be partitioned (a finer partition) to describe both the "touch" and "on-top-of" relations. The states of the machine are:

Empty	If there are no vertices in the diagram, i.e., an empty diagram.
Dispersed	If there no edges in the diagram, i.e., a null diagram (Fig. 2).
Overlapped	If there are at least two vertices connected with an edge (Fig. 3).
Ambiguous	If there is one or more directed cycles in the diagram (Fig. 4).
Unstable	This category is not tested by the analysis of the graph but through analysis of the contact point/line of the object with the support plane. If this contact is a point or a line, it is classified as unstable. See Fig. 5.

Figura 3: Linhas obtidas pelo algoritmo

### 5.3 Segmentação de palavras

A segmentação de palavras foi obtida um novo pré-processamento da imagem original e com o resultado obtido da segmentação de linhas.

#### 5.3.1 Pré-processamento

Primeiramente, a imagem original foi invertida, dado que a biblioteca utilizada usa como valor-referência para pretos e brancos o inverso do apresentado pela imagem de extensão .PBM.

Após, dado os elementos estruturantes  $e_1$  e  $e_2$ , em que:

$e_1$  = elemento estruturante de 1 pixel de altura e 12 de largura

$e_2$  = elemento estruturante de 10 pixels de altura e 1 de largura

Foram aplicados:

- Um fechamento (dilatação seguido de erosão) entre a imagem original  $I$  e  $e_1$  ( $I_1 = I \bullet e_1$ );

- Um fechamento entre  $I$  e  $e_2(I_2 = I \bullet e_2)$ ;
- A união entre os resultados dos dois fechamentos( $I_f = I_1 \cup I_2$ ).

O resultado de todas essas operações resultou nas componentes ilustradas na Figura 4.

200

IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION, VOL. 7, NO. 3, JUNE 1991

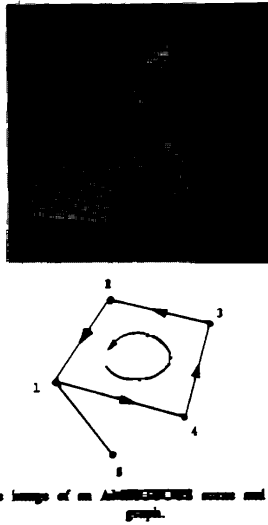


Fig. 4. Range image of an Abolition scene and the corresponding graph.

every feedback is carried out in a local reflexive mode rather than in a planned mode with one exception, that is, when a pathological state is detected.

5) *State*: This is a finite set of states describing the environment of the Turing machine as perceived by the sensors. If new sensors are added, the set of states is partitioned to describe the scene as perceived by the additional sensors. For example, if a sensor capable of determining the "touch" relations of objects in the scene is added, then the set of five states, can be partitioned (in four partitions) to describe both the "touch" and "on-top-of" relations. The states of the machine are:

<i>Empty</i>	If there are no vertices in the digraph, i.e., an empty digraph.
<i>Dispersed</i>	If there are no edges in the digraph, i.e., a null digraph (Fig. 2).
<i>Overlapped</i>	If there are at least two vertices connected with an edge (Fig. 3).
<i>Amibiguous</i>	If there is one or more directed cycles in the digraph (Fig. 4).
<i>Unstable</i>	This category is not tested by the analysis of the graph but through analysis of the contact point/line of the object with the support plane. If this contact is a point or a line, it is classified as unstable. See Fig. 5.

Figura 4: Componentes de linhas encontradas

### 5.3.2 Pós-processamento

Após obter a figura 4, utilizou-se as componentes de linhas obtidas na segmentação de linha. Nesse sentido, o método aplicado foi: para cada componente de linha, encontrar as componentes de palavras baseada na imagem 4.

Esse método foi aplicado para evitar uma segunda filtragem, ou seja, para evitar componentes que já tinham sido classificados como não-texto. Dessa forma, foi possível não só contar as **242 palavras** contidas na imagem, bem como contar quantas palavras há em cada linha.

Os resultados obtidos foram apresentados na Figura 5.

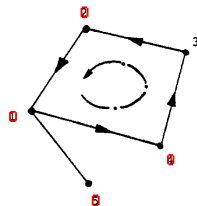
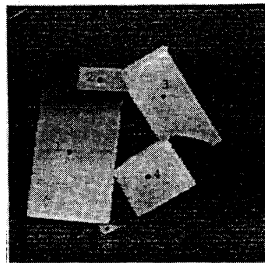


Fig. 4. Range image of an AMBIGUOUS scene and the corresponding graph.

sensory feedback is carried out in a local reflexive mode rather than in a planned mode with one exception, that is, when a pathological state is detected.

3) *States*: This is a finite set of states describing the environment of the Turing machine as perceived by the sensors. If new sensors are added, the set of states is partitioned to describe the scene as perceived by the additional sensors. For example, if a sensor capable of determining the "touch" relations of objects in the scene is added, then the set of five states, can be partitioned (a finer partition) to describe both the "touch" and "on-top-of" relations. The states of the machine are:

Empty	If there are no vertices in the diagraph, i.e., an empty diagraph.
Dispersed	If there no edges in the diagraph, i.e., a null diagraph (Fig. 2).
Overlapped	If there are at least two vertices connected with an edge (Fig. 3).
Ambiguous	If there is one or more directed cycles in the diagraph (Fig. 4).
Unstable	This category is not tested by the analysis of the graph but through analysis of the contact point/line of the object with the support plane. If this contact is a point or a line, it is classified as unstable. See Fig. 5.

Figura 5: Figura-referência

## 6 Conclusão

A implementação do segmentador de linhas/palavras se mostrou eficiente ao que tange a imagem testada. Ainda que uma palavra não tenha sido reconhecida e outras duas tenham sido reconhecidas como uma única palavra, a taxa de sucesso do algoritmo foi alta. Porém, ao mesmo tempo, o tamanho dos elementos estruturantes se mostrou sensível ao tamanho da fonte do texto na imagem. Isso faz com que o segmentador não seja eficiente para todo e qualquer tipo de imagem. Ainda assim, o trabalho teve sucesso ao aliar operações morfológicas para obter os resultados esperados.

+—————+—

## Referências

- [1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. Pearson, 2017. 1

- [2] H. Pedrini, “Aula: Morfologia.” [Online]. Available: [https://www.ic.unicamp.br/~helio/disciplinas/MC920/aula\\_morfologia.pdf](https://www.ic.unicamp.br/~helio/disciplinas/MC920/aula_morfologia.pdf) 1
- [3] —, *trabalho 3*. [Online]. Available: <http://www.ic.unicamp.br/~helio/disciplinas/MC920/trabalho3.pdf> 1
- [4] “Python 3.” [Online]. Available: <https://www.python.org/> 2
- [5] “Numpy.” [Online]. Available: <https://www.numpy.org/> 2
- [6] “Opencv.” [Online]. Available: <https://opencv.org/> 2, 3