

MC920 Trabalho 3 - Operadores Morfológicos

182851 - Lucy Miyuki Miyagusiku Narita

22 de maio de 2019

1 Introdução

O objetivo deste trabalho é aplicar operadores morfológicos para segmentar regiões compreendendo texto e não texto em uma imagem de entrada.

Juntamente desse relatório está sendo enviado um arquivo zip que contém o código e as imagens referentes ao projeto.

2 Execução

O projeto foi desenvolvido em Python 3.7.0, utilizando os seguintes pacotes como dependências:

- Manipulação de dados
 - `numpy`
 - `opencv`
- OCR
 - `tesseract 4.0.0-beta.1 with pytesseract`

As imagens utilizadas como entrada se encontram na pasta `imgs/`. As imagens geradas se encontram na pasta `imgs/out`.

Conforme especificado pelo projeto, apenas imagens no formato PBM (*Portable Bit Map*) são aceitas como entrada. Igualmente, apenas imagens no formato PBM são geradas como saída.

O script pode receber alguns argumentos, que são descritos abaixo:

```
usage: python lab03.py [-h] [--p P] [--op OP] [--out OUT]
                        [--ocr] [--lang LANG] [--seg {char,word,line}]
                        IMAGE_NAME
```

positional arguments:

IMAGE_NAME image file name without extension

optional arguments:

—h, —help show this help message and exit
—p P input image file path
—op OP output image path
—out OUT output image file prefix
—ocr use tesseract to parse identified text segments
—lang LANG text language. Only makes sense when using —ocr
—seg {char,word,line} search for `line`s, `char`s or `word`s

3 Processo e Decisões Tomadas

O algoritmo para identificação de texto e não-texto:

1. dilatação da imagem original com o elemento estruturante:
 - (a) Para linhas: 1 pixel de altura, 100 pixels de largura,
 - (b) Para palavras: 7 pixels de altura, 12 de largura;
2. erosão da imagem resultante com o mesmo elemento estruturante do passo (1);
3. dilatação da imagem original com o elemento estruturante:
 - (a) Para linhas: 200 pixels de altura, 1 pixel de largura,
 - (b) Para palavras: 50 pixels de altura, 1 pixel de largura;
4. erosão da imagem resultante com o mesmo elemento estruturante do passo (3);
5. aplicação da intersecção (AND) dos resultados dos passos (2) e (4);
6. fechamento do resultado obtido no passo (5) com o elemento estruturante:
 - (a) Para linhas: 1 pixel de altura, 30 pixels de largura,
 - (b) Para palavras: 5 pixels de altura, 10 pixels de largura;
7. aplicação de algoritmo para identificação de componentes conexos sobre o resultado do passo (6);

8. para cada retângulo envolvendo um objeto, calcule:
 - (a) razão entre o número de pixels pretos e o número total de pixels (altura×largura);
 - (b) razão entre o número de transições verticais e horizontais branco para preto e o número total de pixels pretos;
9. categorização em *texto* e *não texto* baseando-se nas métricas do passo (8):
 - (a) Para linhas: $0.1 < razão (8a) < 0.5$ e $razão (8b) > 0.2$
 - (b) Para palavras: $0.1 < razão (8a) < 0.6$ e $razão (8b) > 0.2$

Para encontrar as componentes conexas e desenhar as *bounding boxes*, utilizamos as funções do *openCV* ao invés do programa disponibilizado pelo professor.

4 Resultados e Discussões

As operações morfológicas foram aplicados à Imagem 1.

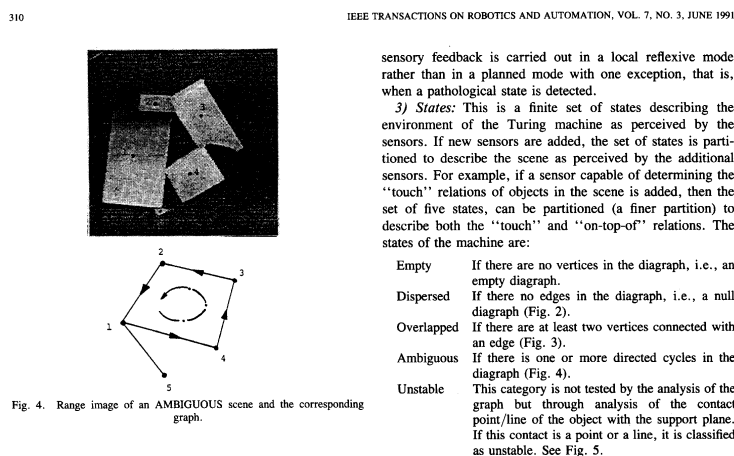


Figura 1: Imagem original

Texto identificado pelo Tesseract a partir da imagem original 1:

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

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 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.”

4.1 Detecção de Linhas

As imagens com os resultados parciais de cada passo são apresentadas nas Figuras 2 à 8 e o resultado com a detecção das linhas (somente texto) na Figura 9.

No processo foram encontrados 54 componentes conexos, sendo 40 considerados textos pelo algoritmo.

Texto gerado pelo Tesseract:

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

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

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 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.”

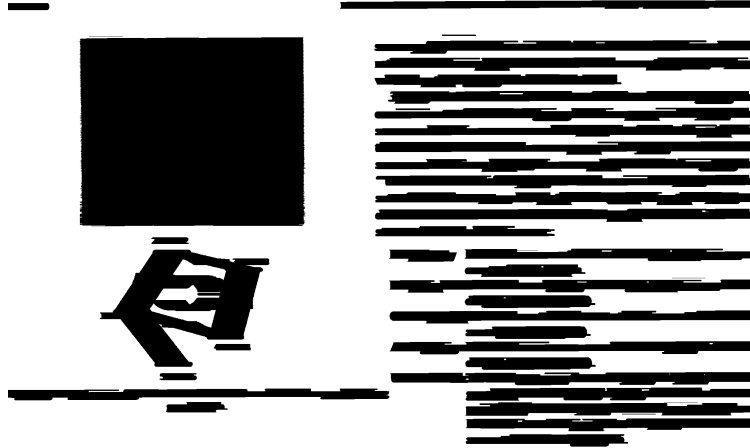


Figura 2: Aplicação do passo 1



Figura 3: Aplicação do passo 2

4.2 Detecção de Palavras

As imagens com os resultados parciais de cada passo são apresentadas nas Figuras 10 à 16 e o resultado com a detecção das palavras (somente o que foi considerado como texto) na Figura 17.



Figura 4: Aplicação do passo 3

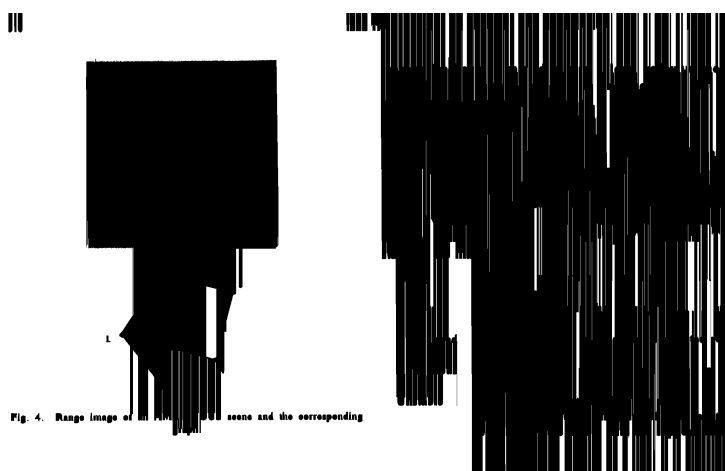


Fig. 4. Range image of the scene and the corresponding

Figura 5: Aplicação do passo 4

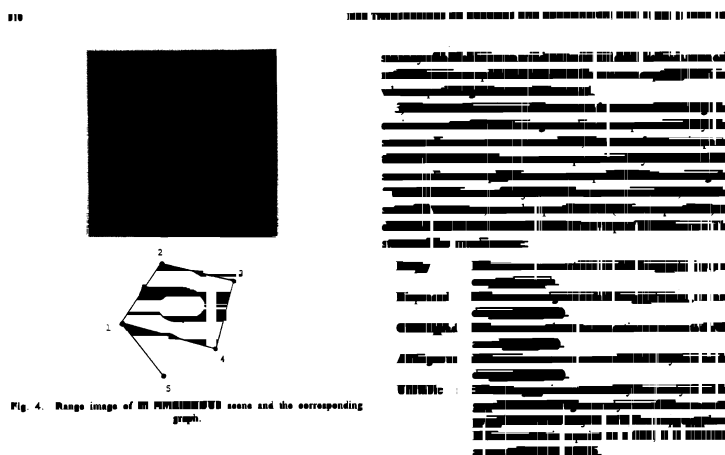


Fig. 4. Range image of the scene and the corresponding graph.

Figura 6: Aplicação do passo 5

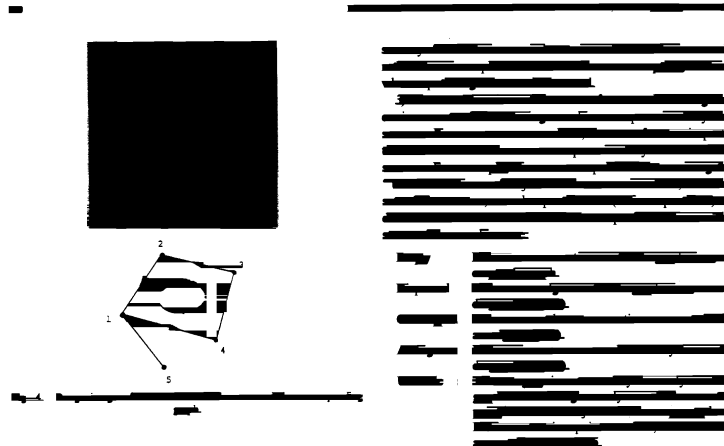


Figura 7: Aplicação do passo 6

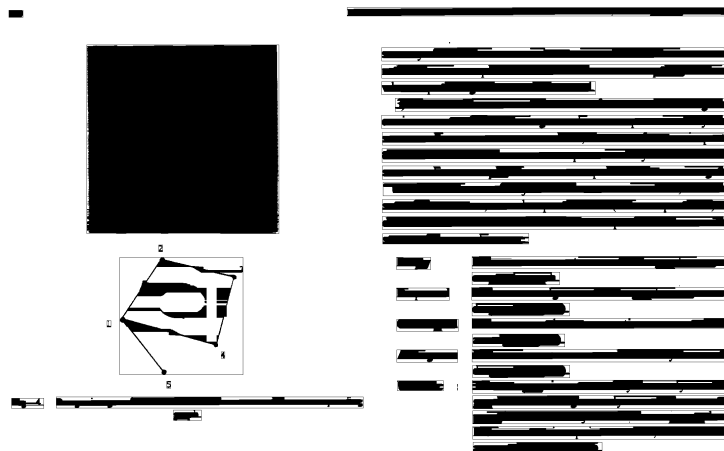


Figura 8: Aplicação do passo 7

310

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

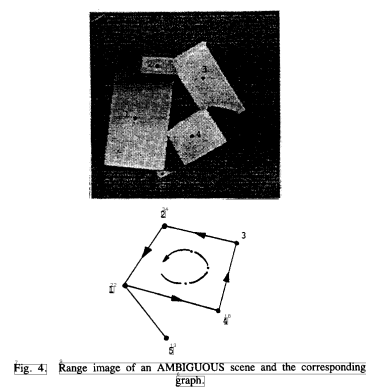


Fig. 4. Range image of an AMBIGUOUS scene and the corresponding graph.

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 9: Detecção de *texto* na Figura 1

No processo foram encontrados 249 componentes conexos, sendo 242 considerados textos pelo algoritmo.

Texto gerado pelo Tesseract:

“310 TEE TRANSACTIONS ren ROBOTICS AUTOMATION, VOL. 7, NO. cm
JUNE 199]

4 i 1 4 5

Fig. re Ree Tarte ata inl FOV elOKs scene and the corresponding graph.

sensory feedback it carried revel hl | local reflexive mode Potala than Th F planned mode with one exception, aati i when a pathological state Ts detected.

3) States: This T a finite set of States describing the environment of the Turing machine as ORK hee by the Sensors. ita new sensors added, the set rey i states tS ovase tioned ke) describe iets scene as Nous hiee by the additional sensors. For example, ti a sensor capable ray i determining the “*toych” relations of telnet in the scene is added, siatets] the...”

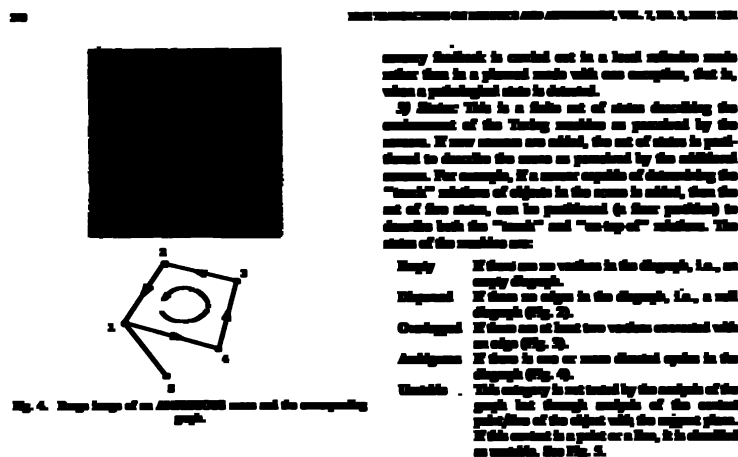


Figura 10: Aplicação do passo 1

5 Conclusão

Neste exercício pudemos verificar a utilização de alguns operadores morfológicos para fazer a segmentação entre *texto* e *não texto* de uma imagem.

A segmentação da imagem facilita o reconhecimento de caracteres por engines de OCR, uma vez que diminuimos o "ruído"(ou seja, tudo que não é texto) e ajudamos a engine a focar nos pontos que realmente importam para nós.

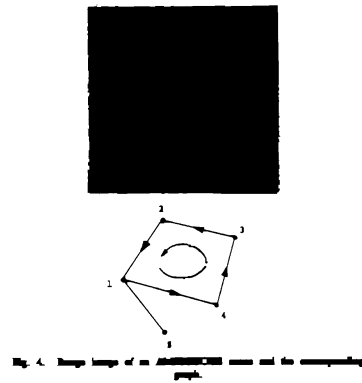


Fig. 4. Range image of an AMBIGUOUS scene and the corresponding graph.

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.

3) States: This is a finite set of states describing the environment of the Turing machine as perceived by the sensor. 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 to four positions to describe both the "touch" and "no-touch" relations. The states of the machine are:

- Empty** If there are no vertices in the diagram, i.e., an empty diagram.
- Dispersed** If there are 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 content pixels/lines of the object with the support plane. If this content is a point or a line, it is classified as unstable. See Fig. 5.

Figura 11: Aplicação do passo 2



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

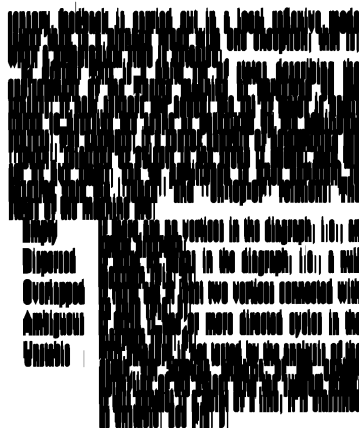


Figura 12: Aplicação do passo 3

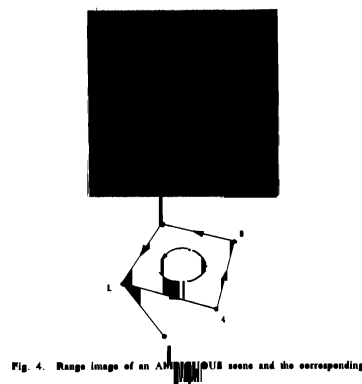


Fig. 4. Range image of an AMBIGUOUS scene and the corresponding graph.

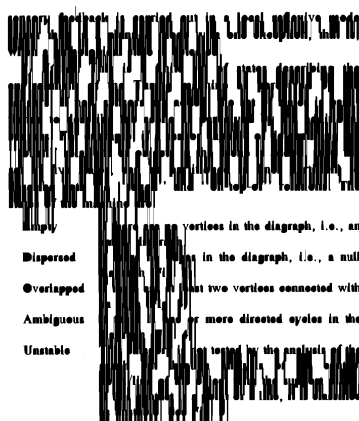


Figura 13: Aplicação do passo 4

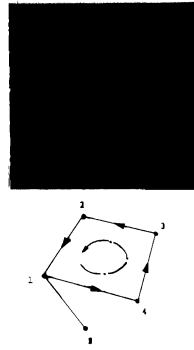


Fig. 4. Range image of an AMBERSCOIL scene and the corresponding graph.

sonary 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 Tacing machine as perceived by the sensors. If new sensors are added, the set of states is partitioned to describe the scene as partitioned 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:

- Empty** If there are no vertices in the diagram, i.e., an empty diagram.
- Dispersed** If there are 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 14: Aplicação do passo 5

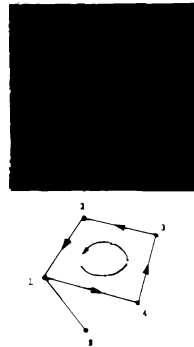


Fig. 4. Range image of an AMBERSCOIL scene and the corresponding graph.

sonary 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 Tacing machine as perceived by the sensors. If new sensors are added, the set of states is partitioned to describe the scene as partitioned 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:

- Empty** If there are no vertices in the diagram, i.e., an empty diagram.
- Dispersed** If there are 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 15: Aplicação do passo 6

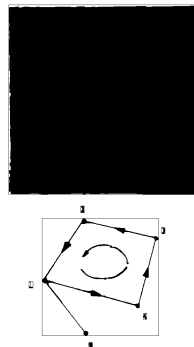


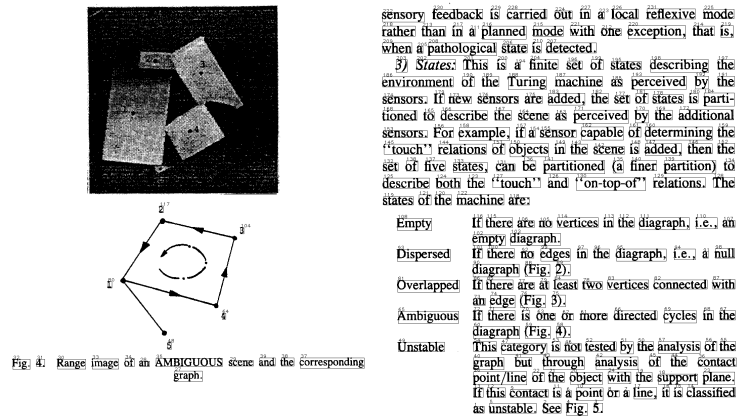
Fig. 4. Range image of an AMBERSCOIL scene and the corresponding graph.

sonary 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 Tacing machine as perceived by the sensors. If new sensors are added, the set of states is partitioned to describe the scene as partitioned 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:

- Empty** If there are no vertices in the diagram, i.e., an empty diagram.
- Dispersed** If there are 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 16: Aplicação do passo 7



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 are 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 17: Detecção de *texto* na Figura 1

Pelos resultados aqui apresentados, podemos ver que as partes segmentadas performaram pior em relação ao texto gerado pela engine de OCR. Isso pode ser explicado pelo fato do Tesseract ser otimizado para trabalhar com sentenças completas [2], ao invés de linhas ou palavras sozinhas. Mesmo utilizando as diferentes opções de segmentação, o resultado não teve um aumento significativo na qualidade. É possível argumentar também que não foram feitas outras configurações adicionais na engine / no dicionário por falta de conhecimento da ferramenta e falta de tempo para se habituar a ela. Além disso, o texto utilizado como input foi cortado rente às bordas do retângulo envolvente, o que aparentemente degrada a performance do Tesseract ¹.

Referências

- [1] Pedrini, H., *MC920 Introdução ao Processamento Digital de Imagens - Morfologia Matemática*, (Universidade Estadual de Campinas, UNICAMP. 2019). [Online] available at http://www.ic.unicamp.br/~helio/disciplinas/MC920/aula_morfologia.pdf
- [2] *Tesseract OCR*. [Online] available at <https://github.com/tesseract-ocr/tesseract/wiki/ImproveQuality#dictionaries-word-lists-and-patterns>

¹<https://web.archive.org/web/20151209085049/https://code.google.com/p/tesseract-ocr/issues/detail?id=398>