

# Pedra, Papel, Tesoura, Lagarto e *Spock*

Diego Brian, Pedro Aurélio e Rodrigo Guimarães

Departamento de Ciência de Computação e Faculdade de Tecnologia

Universidade de Brasília, Brasília

E-mails: diegobleite@hotmail.com, pedro.almeidabsb@gmail.com e rodrigofegui@aluno.unb.br

Matrículas: 14/0136371, 14/0158103 e 14/0170740

## I. INTRODUÇÃO

Este trabalho visa fixar os conceitos aprendidos ao longo de um semestre de aulas. Para atingir tal objetivo foi desenvolvido um programa que realiza a detecção de gestos manuais, de forma a implementar o jogo popular: **Pedra, Papel, Tesoura, Lagarto e *Spock*** [10].

04 de dezembro, 2015

## II. EMBASAMENTO TEÓRICO

Para que haja um correto entendimento sobre o desenvolvimento deste projeto é importante abordar alguns aspectos relevantes, como as definições: de uma imagem; das relações básicas entre *pixels*: adjacência, conectividade, região e contorno; do espaço de cor YCbCr; operações morfológicas; segmentação, dentre outras coisas.

### A. Imagem e Pixels

Uma imagem pode ser definida como uma função bidimensional do tipo  $f(x, y)$ , onde  $x$  e  $y$  são as coordenadas espaciais e a amplitude de  $f$  em qualquer ponto de coordenadas  $(x, y)$  é denominado de intensidade da imagem naquele ponto. Quando  $x$ ,  $y$  e  $f$  são valores finitos e discretos, a imagem é denominada imagem digital, tendo esta significância aos computadores digitais. Um dado elemento com coordenadas  $(x, y)$  e intensidade  $f$  é denominado de *pixel* (picture element ou, em português, elemento de imagem), dessa forma, entende-se que uma imagem é constituída por um ou mais *pixels*.

Para a manipulação dos *pixels* é necessário saber as relações básicas entre eles, como, por exemplo, a vizinhança. Os conceitos a serem apresentados consideram uma imagem em nível de cinza.

Cada *pixel*  $p$  pode possuir três tipos de vizinhança, semelhantes a Rosa-dos-Ventos da Figura 1: *vizinhança de 4*, *vizinhança diagonal* e *vizinhança de 8*; onde para o primeiro, denotado por  $N_4(p)$ , são considerados os quatro vizinhos horizontais e verticais, seguindo as orientações N-S-L-O da Rosa-dos-Ventos; para o segundo, denotado por  $N_D(p)$ , são considerados os vizinhos das diagonais, seguindo as orientações NE-SE-SO-NO; enquanto que o terceiro tipo, denotado por  $N_8(p)$ , é a junção dos dois anteriores.

### B. Relações Básicas entre Pixels

Com o conceito de vizinhança construído é possível entender o conceito de adjacência, entre dois *pixels*  $p$  e  $q$ , que, também, possui três tipos, sendo eles: **a) Adjacência de 4**: os

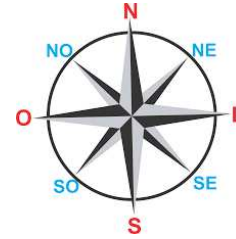


Figura 1: Sinalização dos sentidos cardeais da Rosa-dos-Ventos. Fonte [7].

dois *pixels* são adjacentes de 4 se ambos possuírem o mesmo nível de cinza e se  $q$  está na  $N_4(p)$ ; **b) Adjacência de 8**: os dois *pixels* são adjacentes de 8 se ambos possuírem o mesmo nível de cinza e se  $q$  está na  $N_8(p)$ ; **c) Adjacência de  $m$** : os dois *pixels* são adjacentes de  $m$  se ambos possuírem o mesmo nível de cinza e se respeitarem as condições: **c.1)**  $q$  está na  $N_4(p)$  ou **c.2)**  $q$  está na  $N_D(p)$  e não há interseção entre a  $N_4(p)$  e  $N_4(q)$ .

Considerando um subconjunto  $S$  de imagem, dois *pixels* pertencentes a  $S$  são conectados, considerando um dos critérios de adjacência, caso exista um caminho entre eles que também pertença a  $S$ . Para qualquer *pixel*  $p$  em  $S$ , o conjunto de *pixels* que proporciona a união a  $q$  é denominado de *conjunto conectado*. Caso  $q$  seja adjacente a  $p$ , este é chamado de *componente conectado*.

Uma região  $R$  é definida como um conjunto conectado e o contorno  $C$  de  $R$  é definido como sendo os *pixels* que possuem vizinhos que tanto pertencem a  $R$  como não pertencente.

### C. Espaço de Cores

As imagens coloridas são construídas através de espaços de cores, por exemplo o RGB (*Red, Green e Blu* ou, em português, Vermelho, Verde e Azul). Os monitores de computadores e *smartphones*, além dos televisores, são constituídos por vários *pixels*, como demonstrado na Figura 2, sendo que cada *pixel* é constituído por 3 LEDs, basicamente, para cada cor do RGB.

Considerando que cada *pixel* em cada plano de um espaço de cores é descrito por 1 byte ou 8 bits; ao implementar este conhecimento no espaço de cores RGB é possível ter  $256^3$  cores diferentes. Mesmo tendo esta grande gama de cores é sabido que o olho humano é mais sensível à luminância (brilho) do que à cor; considerando que o RGB, em que a

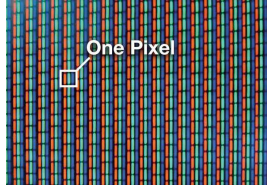


Figura 2: Aproximação num monitor, sendo possível perceber os *pixels* e os *LED* do espaço de cor *RGB*. Fonte [8].

luminância está espalhada pelas componentes, não é adaptado para se aproximar da sensibilidade humana é preciso utilizar outro espaço de cores que faça tal adaptação.

Como opção tem o espaço de cores *YCbCr* que possui as componentes: luminância (*Y*) e as crominâncias (*Cb* e *Cr*), onde a luminância é mais significativa do que as crominâncias. É importante ressaltar que, devido à constituição dos monitores, e similares, é necessário que haja formas de conversão entre os espaços de cores, independente de qual seja o escolhido, para o padrão *RGB*; é possível tal conversão entre o *RGB* e o *YCbCr*.

#### D. Operações Morfológicas

A morfologia, geralmente, é considerada apenas como uma especialização da biologia que trabalha com a forma e a estrutura dos seres vivos. Além desse conceito, há a morfologia matemática que estuda o processamento e análise de estruturas geométricas. Dessa forma, tem que para o processamento de imagens é aplicado a morfologia matemática, pois dessa forma é realizada a extração de regiões de uma imagem, representação e descrição das formas de uma determinada região.

Dessa forma, considerando imagens binárias, tem-se algumas operações, sendo elas: erosão; dilatação; fechamento; abertura; *top-hat*; *bottom-hat*; dentre outras, sendo que algumas destas serão exploradas neste projeto.

1) *Erosão*: A erosão é definida como sendo:

$$A \ominus B = \{z \mid (B)_z \subseteq A\},$$

ou seja, a erosão de *A* por *B* é o conjunto de todos os pontos *z*, tal que *B* transladado de *z* está contido em *A*, como demonstrado na Figura 3 (*B* é chamado *elemento estruturante*). Dessa forma ao considerar que *B* não tem elementos comuns com o fundo é possível reescrever como sendo:

$$A \ominus B = \{z \mid (B)_z \cap A^c = \emptyset\}.$$

2) *Dilatação*: A dilatação é definida como sendo:

$$A \oplus B = \{z \mid (\hat{B})_z \cap A \neq \emptyset\},$$

ou seja, a dilatação de *A* por *B* é o conjunto de todos os deslocamentos *z* de forma que *B* e *A* se sobrepõem, em pelo menos, por um elemento, como demonstrado na Figura 4; um modo alternativo:

$$A \oplus B = \{z \mid [(\hat{B})_z \cap A] \subseteq A\}.$$

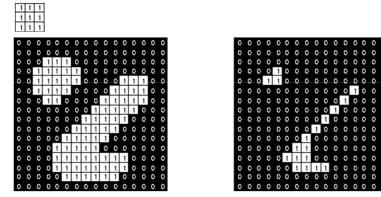


Figura 3: Exemplo da operação morfológica: Erosão. À esquerda é a imagem original, à direita é a imagem erodida. Fonte: fornecido pelo professor.

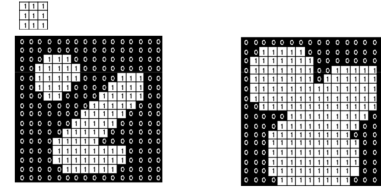


Figura 4: Exemplo da operação morfológica: Dilatação. À esquerda é a imagem original, à direita é a imagem dilatada. Fonte: fornecido pelo professor.

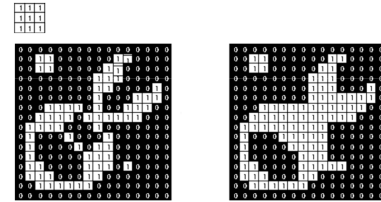


Figura 5: Exemplo da operação morfológica: Fechamento. À esquerda é a imagem original, à direita é a imagem fechada. Fonte: fornecido pelo professor.

3) *Fechamento*: O fechamento visa suavizar contornos como, também, funde as descontinuidades estreitas e alonga os golfos finos, elimina pequenos buracos e preenche lacunas num contorno, sendo definida como sendo:

$$A \bullet B = [(A \oplus B) \ominus B],$$

ou seja, consiste na dilatação de *A* por *B*, seguida da erosão do resultado por *B*, como demonstrado na Figura 5.

#### E. Segmentação

A segmentação consiste em dividir uma imagem em regiões de interesse, de forma a permitir a extração de suas características. Para atingir este objetivo há duas maneiras: **a)** transições, que considera as mudanças bruscas de intensidade; ou **b)** similaridade, que considera regiões semelhantes, tendo um critério prédefinido.

### III. METODOLOGIA

Com as definições feitas é possível prosseguir para o desenvolvimento do projeto. Dessa forma, esta seção visa apresentar os passos seguidos para a realização das atividades, desenvolvidas em *MatLab* [2]. O trabalho foi subdividido em duas partes: **a)** detectar a mão na imagem e **b)** determinar a

forma que a mão assume, por isso as partes serão tratadas separadamente.

#### A. Realce da Mão

Para o realce da mão é considerado um par de imagem: a primeira com apenas o plano de fundo (não podendo aparecer parte do corpo); e a segunda com a mão inserida no contexto. Com isso, é determinar a diferença entre as duas imagens, o que resultaria numa imagem que possui apenas a mão, idealmente; mas realizando apenas a diferença entre as imagens foram detectados muitos ruídos, por isso foi realizado um incremento à técnica.

Ambas as imagens foram convertidas para o espaço de cores  $YCbCr$ , pois as cores de pele tem maior destaque no plano  $Cr$ , auxiliando na segmentação da mão. Foram determinadas as diferenças para os dois planos de crominância, seguida pela soma das diferenças, sendo considerada como a diferença absoluta das imagens no espaço de cores  $YCbCr$ . A diferença sobre o plano da luminância também foi calculada, para a determinação seu limiar de binarização (será explicado mais adiante).

A diferença absoluta foi binarizada, seguido pelo processo morfológico de fechamento, pelo preenchimento de pequenos buracos e pela remoção de pequenos ruídos na imagem, através da erosão. Com isso, objetiva-se conseguir uma imagem com a mão segmentada e alguns poucos ruídos. Após essas operações, é realizada a detecção de regiões, que servirá de base para o refinamento das regiões, onde regiões que possuem área total inferior a 6% da área total da imagem são descartadas, pois foi percebido que se tratavam de ruído. Vale ressaltar, que também houveram mãos que foram desconsideradas nesta etapa, fruto de uma má segmentação, proporcionada pelas condições ambientais e da câmera de aquisição.

Após este processo, é avaliado o limiar de binarização da diferença das luminosidades, citado anteriormente. Caso seja superior a 158, significa que existia muita luminosidade, o que pode prejudicar a segmentação da mão, por isso é realizado o processamento, já descrito, sobre a escala de cinza de ambas as imagens-base. Este procedimento é feito para reformar as formas gerais, pois no final é feita uma junção dos resultados no espaço de cores  $YCbCr$  com os da escala de cinza.

Da imagem resultante deste processamento é realizada a detecção de bordas, pelo método de *Canny* [9], que será necessária para a determinação da forma, a seguir.

#### B. Determinação da Forma

Antes de qualquer operação, foi analisada a orientação da mão segmentada, onde ao ser considerada como uma mão na horizontal a mesma é invertida para a forma vertical, de forma a ter o mesmo processamento sobre todas as imagens.

Para a determinação da forma foram consideradas: o bloco que engloba a mão segmentada; e a imagem das bordas detectadas. Pois é selecionada uma linha para a realização de um corte horizontal na imagem, tal linha se localiza a 70% da distância do centro do bloco até o topo do mesmo. Para esta linha é, então, determinada a quantidade de dedos. Sobre esta linha,



Figura 6: Imagem do plano de fundo considerado para exibição dos resultados.

também, é determinado a primeira e a última coluna que constitui a mão.

É determinada, também, uma linha móvel que vai variar do centro do bloco +15% até centro do bloco -30%, em relação ao topo do bloco, onde a porcentagem positiva é linha, visualmente, acima do centro e analogamente para a porcentagem negativa. Isto é feito para verificar a existência ou não de dedão.

Enfim, com essas informações a forma, que corresponde à mão segmentada, é determinada. Onde são estas as escolhas:

- Se houver dedão, então:
  - Se dedos é superior a 3, então: Papel
  - Senão se tiver, exatamente, 2, então: *Spock*
  - Senão: Lagarto
  - Fim\_se
- Senão:
  - Se dedos é igual a 2, então: Tesoura
  - Senão:
    - \* Se proporção entre colunas é superior a limiar, então: Pedra
    - \* Senão: Lagarto
    - \* Fim\_se
  - fim\_se
- Fim\_se

Onde o limiar citado é um valor percentual da distância entre colunas com o comprimento do bloco, equivalente a 50%.

## IV. RESULTADOS

Como dito na Seção III há dois problemas, dessa forma esta seção também será subdividida. Todos os códigos estão disponíveis *online* [6], assim como todas as imagens-resultado obtidas com a execução dos *scripts*. Além disso, só serão expostos os resultados sobre as Figuras 6, plano de fundo, 7, plano de fundo com o acréscimo da mão em forma de papel, e 8, plano de fundo com o acréscimo da mão em forma de pedra.

#### A. Realce da Mão

Considerando as imagens-base já citadas, foi realizada a operação de diferença entre os planos, o que resultou na Figura 9 como sendo a diferença absoluta no espaço de

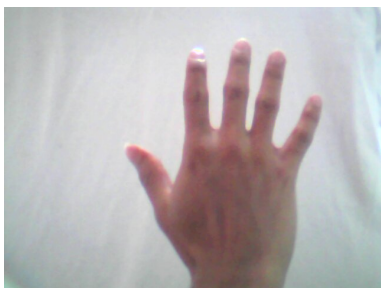


Figura 7: Imagem do plano de fundo com o acréscimo da mão, em forma de papel, considerado para exibição dos resultados positivos.



Figura 8: Imagem do plano de fundo com o acréscimo da mão, em forma de pedra, considerado para exibição dos resultados negativos.

cores  $YCbCr$ , após a binarização. Nesta imagem é possível ver que há muitas regiões brancas que não correspondem a mão. Por isso, tem a Figura 10 como o resultado da erosão, seguida pelo preenchimento de buracos e refinamento por área de região, não foi preciso analisar em escala de cinza. Por fim, tem-se a Figura 11 que demonstra a detecção de borda aplicada sobre a Figura 10. Enquanto que sobre a Figura 8 foram obtidas as Figuras 12, 13 e 14, análogo ao já apresentado.

#### B. Determinação da Forma

A partir da Figura 11 apenas o bloco que envolve a forma foi considerado. A Figura 15 demonstra a área de interesse da imagem. Dentro deste bloco foi determinado o número de dedos levantados, que para tal imagem corresponde a 3. Após isso, foi determinado, também, se havia ou não dedão na



Figura 9: Resultado da binarização sobre a imagem da diferença absoluta no espaço de cores  $YCbCr$  da Figura 7.



Figura 10: Resultado do refinamento finalizado sobre a Figura 9.

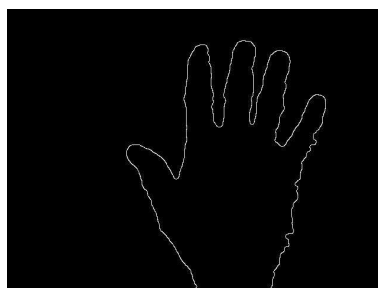


Figura 11: Resultado da detecção de borda sobre a Figura 10.



Figura 12: Resultado da binarização sobre a imagem da diferença absoluta no espaço de cores  $YCbCr$  da Figura 8.



Figura 13: Resultado do refinamento finalizado sobre a Figura 12.



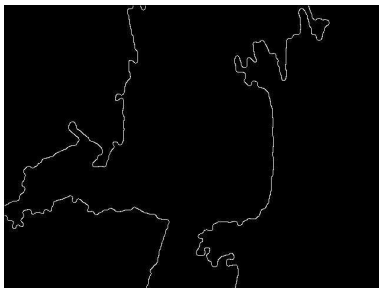


Figura 14: Resultado da detecção de borda sobre a Figura 13.

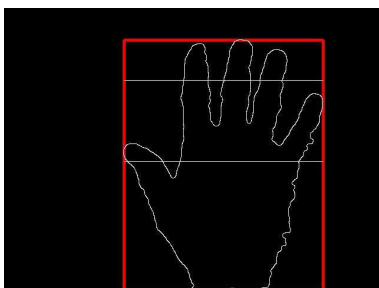


Figura 15: Bloco de interesse sobre a análise da forma da mão, em vermelho. A linha branca superior é o resultado do corte horizontal, para a detecção da quantidade de dedos levantados; já a inferior é resultado do corte horizontal, para a detecção de dedão na imagem.

imagem, sendo que para esta foi detectado que havia dedão na imagem. Enquanto que para a Figura 14 foi obtida a Figura 16, foram detectados 2 dedos e que tem dedão.

Foram consideradas 75 imagens para cada forma de mão possível, ou seja 227 imagens de teste, com dois tipos de planos de fundo e mudanças na orientação da mão. A partir dos resultados, foi construída uma estatística sobre o desempenho do projeto, originando a Figura 17. É possível perceber nesta imagem que há células com cores diferenciadas, onde: vermelho significa a menor estatística; e o resto em azul.

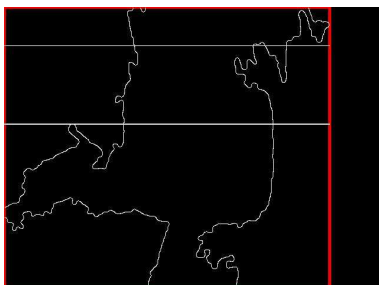


Figura 16: Bloco de interesse sobre a análise da forma da mão, em vermelho. A linha branca superior é o resultado do corte horizontal, para a detecção da quantidade de dedos levantados; já a inferior é resultado do corte horizontal, para a detecção de dedão na imagem.

		Reconhecido Completo					
Esperado	%	Pedra	Papel	Tesoura	Lagarto	Spock	Desconhecido
	Pedra	38,67	5,33	0,00	0,00	9,33	46,67
	Papel	0,00	97,33	0,00	0,00	0,00	2,67
	Tesoura	0,00	4,00	76,00	0,00	5,33	14,67
	Lagarto	0,00	0,00	0,00	16,00	0,00	84,00
	Spock	0,00	9,33	0,00	0,00	80,00	10,67

Figura 17: Estatísticas geradas sobre o aproveitamento do projeto.

## V. CONCLUSÃO

Com o desenvolvimento do projeto foi possível reforçar os conteúdos teóricos visto em sala, demonstrando que o objetivo principal foi alcançado. Foram utilizadas as ferramentas disponibilizadas pelo *MatLab*, que possuem grande poder de atuação no processamento de imagens, o que facilitou a fixação dos conceitos, mas é esperado que futuramente seja feita uma aplicação em *OpenCV* [3] (*Open Source Computer Vision Library* ou, em português, Biblioteca de Visão Computacional em Código Aberto), assim como uma implementação de interface gráfica, já inicializada, mas não concluída.

O resultado do projeto foi bastante satisfatório, pois foram reforçados os tópicos apresentados ao longo do semestre. Além disso, é possível perceber, ao considerar as estatísticas, que o método apresentado possui um aproveitamento mediano, pois 2/5 das porcentagens foram acima de 80%, a terceira maior foi próximo deste valor e, infelizmente, a pedra e o lagarto obtiveram os piores aproveitamentos. Foram encontrados problemas na segmentação, principalmente para o lagarto, o que influenciou na baixa de alguns índices.

Venho por meio deste me ratificar, Rodrigo, sobre os erros dos dados estatísticos apresentados em sala, pois havia um parâmetro errado durante a execução do *script*. Há grande variação quanto aos resultados obtidos, o que indica que o modo de segmentar a mão deve ser incrementado.

## REFERÊNCIAS

- [1] Gonzalez, Rafael C. e Woods, Richard E., *Digital Image Processing*, 3<sup>o</sup> ed, Pearson Ed. - ISBN: 9780131687288.
- [2] MathWorks. *MATLAB and Simulink for Technical Computing*. Disponível em: <https://www.mathworks.com/index.html>, acessado em 2015.
- [3] Documentation, OpenCV. *Welcome to opencv documentation*. Disponível em: <http://docs.opencv.org/index.html>, acessado em 2015.
- [4] Espinoza, B., Dorea, C. e Zaghetto, A. *Material didático utilizado em aula*.
- [5] Bussab, Wilton de Oliveira; Morettin, Pedro Alberto, *Estatística Básica*, 8<sup>o</sup> ed, Editora Saraiva - ISBN: 9788502207998
- [6] 4Shared. *Trabalho Final\_IPI*. Disponível em: [http://www.4shared.com/zip/IqMwwD.J8ce/TrabalhoFinal\\_IPI.html](http://www.4shared.com/zip/IqMwwD.J8ce/TrabalhoFinal_IPI.html).
- [7] Triade da Aprovação. *Níveis de Conhecimento: Por onde começar e até onde você deve estudar cada assunto?*. Disponível em: <http://triadedaaprovação.com/niveis-de-conhecimento-por-onde-comenar-e-até-voce-deve-estudar-cada-assunto/>, acessado em 2015.
- [8] One Climbs. *Three: The Exploration of Archetypal Symbols Series*. Disponível em: <http://oneclimbs.com/2011/03/14/three-the-exploration-of-archetypal-symbols-series/>, acessado em 2015.

- [9] Do Vale, G. M. e Dal Poz, A. P. *Processo de Detecção de Bordas de Canny*. Disponível em: [http :  
//ojs.c3sl.ufpr.br/ojs/index.php/bcg/article/viewFile/1421/1175](http://ojs.c3sl.ufpr.br/ojs/index.php/bcg/article/viewFile/1421/1175),  
acessado em 2015.
- [10] Youtube. *Pedra-papel-tesoura-lagarto-Spock*. Disponível em: [https :  
//www.youtube.com/watch?v = abQj0pQkSOY](https://www.youtube.com/watch?v=abQj0pQkSOY), acessadp em  
2015.