

Detecção de Componentes Conectados e Classificação de Imagens

Rodrigo Ferreira Guimarães
Departamento de Ciência de Computação e Faculdade de Tecnologia
Universidade de Brasília, Brasília
Email: rodrigofegui@aluno.unb.br

Resumo—The abstract goes here.

I. INTRODUÇÃO

Este trabalho trata da detecção de componentes conectados e da classificação de imagens, a partir desta detecção. Foi construído, com o auxílio da toda a turma, um banco de dados com mais de 200 imagens, sendo estas divididas igualmente em duas classes: **classe 1** compreende as imagens que possuem pessoas com a maior parte do corpo exposta; enquanto que para a **classe 2** compreende as imagens com as pessoas vestidas.

O trabalho objetiva realizar: **a)** a binarização das imagens a partir da segmentação por cor de pele; **b)** a rotulação para detectar a quantidade de elementos conectados na imagem binária; **c)** a diferenciação do maior elemento conectado, através da sua área; e por fim, **d)** a classificação das imagens dentre as opções de classes.

12 de Setembro, 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; média aritmética e desvio padrão; do espaço de cor YCbCr; e da F-measure.

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*



Figura 1. Sinalização dos sentido cardeais da Rosa-dos-Ventos. Fonte [6].

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:

- **Adjacência de 4:** os dois *pixels* são adjacentes de 4 se ambos possuírem o mesmo nível de cinza e se q está na $N_4(p)$;
- **Adjacência de 8:** os dois *pixels* são adjacentes de 8 se ambos possuírem o mesmo nível de cinza e se q está na $N_8(p)$;
- **Adjacência de m :** os dois *pixels* são adjacentes de m se ambos possuírem o mesmo nível de cinza e se respeitarem as condições: **a)** q está na $N_4(p)$ ou **b)** 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 R como não pertencente.

C. Média Aritmética e Desvio Padrão

Uma das formas de mensurar o comportamento de fenômenos ao longo do tempo - considerando o tempo como a generalização para a variável independente do fenômeno e não apenas como o tempo decorrido em, por exemplo, segundos - é através da *média aritmética* ($M_e(X)$). Como foi definido anteriormente, uma imagem é uma função discreta e para tal é possível definir a média aritmética como sendo o somatório das intensidades dos *pixels* dividido pela quantidade de *pixels* somados, tendo a expressão geral:

$$M_e(X) = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i,$$

onde x_1, x_2, \dots, x_n são as intensidades observadas nos *pixels* de estudo.

Além da $M_e(X)$, a estatística fornece várias ferramentas para a representação da variabilidade de um conjunto de dados, uma vez que grupos distintos podem apresentar a mesma média; neste projeto será utilizada a ferramenta denominada *desvio padrão* ($D_P(X)$), tendo a expressão geral:

$$D_P(X) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - M_e(X))^2}$$

D. 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 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*.

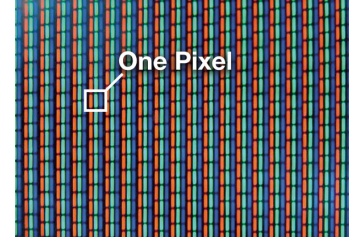


Figura 2. Aproximação num monitor, sendo possível perceber os *pixels* e os *LED* do espaço de cor *RGB*. Fonte [5].

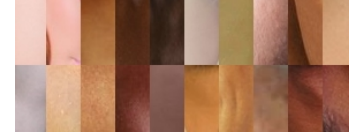


Figura 3. Paleta desenvolvida com recortes de pele de imagens, base para a comparação.

E. F-Measure (F_1)

Em análise estatística de classificação binária, o F_1 é a medida de acurácia de testes, precisão; considera tanto os verdadeiros positivos (P_V) como os falsos (P_N), possuindo como resultado um valor entre 0 e 1 respeitando a relação: (pior) $0 \leq resultado \leq 1$ (melhor), pela expressão geral:

$$F_1 = \left[\frac{2P_V}{2P_V + P_N} \right]$$

III. METODOLOGIA

Com as definições feitas é possível prosseguir para o desenvolvimento do projeto. Para tanto foram utilizados alguns recursos disponibilizados pela *OpenCV* [4] (*Open Source Computer Vision Library* ou, em português, Biblioteca de Visão Computacional em Código Aberto). O projeto foi dividido em etapas:

A. Construção da Paleta de Cores

Para que haja a caracterização de cor de pele é importante ter uma base de comparação; como o banco de dados é compartilhado por toda a turma, foi definido que as 20 imagens fossem destinadas para a construção da paleta, seguindo o critério que deva conter 10 fotos de cada classe, sendo, ainda, estas subdivididas em 5 de cada gênero, e que estas fotos sejam as mesmas para toda turma, diferenciando, somente, quanto aos pedaços de pele a serem recortados, o resultado é demonstrado na Figura 3.

Procurou-se construir uma paleta que continha transição entre os tons de pele, dentro de cada imagem, pois dessa forma a gama de tons característicos como tons de pele seria um grupo com menos restrições que as já impostas.

A partir da paleta de cores foram definidos as médias aritméticas e os desvios padrões para o espaço de cores *YCbCr*, sendo estes os valores de referência para as futuras comparações.

B. Binarização das imagens

Com a construção da paleta de cores, o passo seguinte é a binarização das imagens.

Para tal tarefa foi realizada a conversão de espaços de cores, do *RGB* para o *YCbCr*, pois uma pele não deixa de ser uma pele com a variação de luminosidade. Após isso, os *pixels* de cada plano de crominância foram comparados com os valores determinados pela paleta de cores; caso estivesse no intervalo $\text{média} \pm \text{desvio padrão}$, então era marcado como 1 (branco) ou 0 (preto), caso contrário. Dessa forma, a imagem é binarizada com relação à cor de pele.

Os planos de crominância do *YCbCr* podem sofrer operações de redimensionamento sem alteração ao olho nu, por isso as imagens foram redimensionadas para 400×400 , para agilizar o processamento do banco de dados.

C. Detecção das Regiões

O algoritmo adotado para a detecção das regiões considera a conectividade entre os *pixels*, uma vez que para tal tarefa seguiu-se os passos:

- 1) Encontre um *pixel* branco, marque-o como visitado e armazene-o como sendo o pivô de uma região;
- 2) Busque na vizinhança de 8 se há outros *pixels* brancos;
 - a) Caso encontre, marque-o como visitado e armazene-o na região do seu *pixel* pivô correspondente;
 - i) Repita a partir do passo 2).
- b) Repita enquanto houver *pixel* a ser visitado ou não houver mais *pixels* conectados a esta região;
- 3) Repita enquanto houver imagem a ser navegada.

Esta maneira de procurar as regiões deriva das *buscas em largura e profundidade* em grafos, visto que busca de maneira recursiva os *pixels* conectados à região a partir de um dado *pixel* pivô.

D. Classificação das imagens

Além das imagens a serem classificadas houve imagens de treinamento, com características semelhantes às imagens a serem classificadas - também compartilhadas com a turma. Isto foi feito para que se tenha um critério quanto à discriminação entre as classes.

Para as imagens de treinamento foram feitos os mesmos procedimentos que seriam feitos com as imagens a serem classificadas; a diferença está no fato de que foi calculado a média da área dita como pele para ambas as classes, dessa forma foi calculada a média das médias para a determinação da média limiar para a distinção entre as classes.

E. Cálculo da *F-Measure*

Com a construção da média limiar entre as classes, foi possível realizar a distinção entre as classes para as imagens lidas, ao passo que são contados os verdadeiros positivos.



Figura 4. Imagem classificada como verdadeiro positivo.

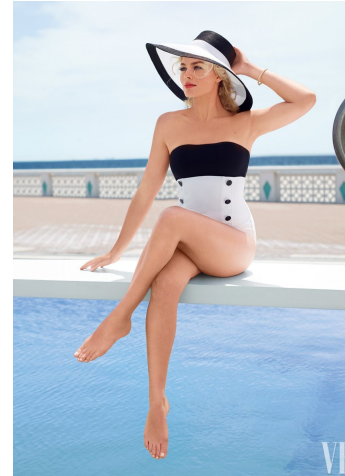


Figura 5. Original da Figura 4.

IV. RESULTADOS

Houve imagens classificadas corretamente entre as classes, considerados como os verdadeiros positivos, como demonstrado na Figura 4, tendo como origem a Figura 5; ao passo que, também, houve erros quanto à classificação entre as classes, considerados como os falsos positivos, como demonstrado na Figura 6, tendo como origem a Figura 7, oriundo da falha de binarização, comentada a seguir. Apesar disso, o algoritmo possui um *F-Measure* de 0.745645.

Houve problemas em algumas imagens quanto à binarização, como demonstrados nas Figuras ??, tendo como origem a Figura ??, pois, visualmente, é de conhecimento geral que se tratava de uma representação de uma pessoa na imagem, mas quanto aos valores dos *pixels*, a maioria foge da média estabelecida pela paleta; houve o acréscimo aos valores das médias e dos desvios padrões, a fim de minimizar este erro de reconhecimento de tom de pele.

Um aspecto positivo para o algoritmo desenvolvido é a rapidez que processa as imagens, visto que foram lidas pouco mais de 200 imagens em menos de 15 segundos de processamento.

V. CONCLUSÃO

Este trabalho proporcionou uma maior fixação do conteúdo ministrado em sala de aula, uma vez que foi necessário o



Figura 6. Imagem classificada como falso positivo.



Figura 7. Original da Figura 6.

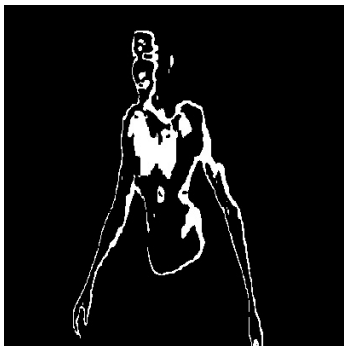


Figura 8. Problema na binarização.



Figura 9. Original da Figura ??.

REFERÊNCIAS

- [1] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.
- [2] Gonzalez, Rafael C. e Woods, Richard E., *Digital Image Processing*, 3^o ed, Pearson Ed. - ISBN: 9780131687288.
- [3] Bussab, Wilton de Oliveira; Morettin, Pedro Alberto, *Estatística Básica*, 8^o ed, Editora Saraiva - ISBN: 9788502207998
- [4] Documentation, OpenCV. *Welcome to opencv documentation*. Disponível em: <http://docs.opencv.org/index.html>, acessado em 2015.
- [5] 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.
- [6] Triade da Aprovação.*Níveis de Conhecimento: Por onde começar e até onde você deve estudar cada assunto?*. Disponível em: <http://triadedaaprovacao.com/niveis-de-conhecimento-por-onde-comencar-e-ate-onde-voce-deve-estudar-cada-assunto/>, acessado em 2015;

desenvolvimento de todo o algoritmo.

Além disso, com o valor de 0.745645 para o *F-Measure*, conclui-se que o projeto mesmo com limitações quanto à correta classificação das imagens, decorrente de falhas na binarização das imagens, possui espaço para que seja melhorado em atividades futuras.