



Segmentação de imagens tridimensionais utilizando o sensor Microsoft Kinect

Lucas Viana Barbosa¹; Wanderson Rigo²; Manassés Ribeiro³

INTRODUÇÃO

Os sistemas de visão artificial vêm auxiliando o ser humano nas suas mais variadas tarefas. Sua aplicabilidade se estende por diversas áreas como medicina, engenharia, agricultura, meio ambiente, entre outras [2]. Uma área de grande interesse atualmente está relacionado à segmentação de imagens tridimensionais que pode ser utilizada em diversas aplicações desde jogos até ferramentas de apoio à pessoas com necessidades especiais [2] [3]. Neste trabalho propõe-se a segmentação de imagens em profundidade capturadas por meio do sensor Microsoft Kinect. Esse sensor, que tem sido utilizado em diversas aplicações pela sua capacidade em fornecer interação humano-computador, foi desenvolvido pela empresa PrimeSense, para o console de jogos Xbox da Microsoft, e é capaz de realizar uma varredura no ambiente, capturando imagens tridimensionais. O sensor é composto por sensores RGB (*Red, Green, Blue*), de profundidade e microfones, sendo que o sensor de profundidade consiste em um projetor de laser infravermelho combinado com um sensor monocromático (CMOS).

Este trabalho objetiva propor um conjunto de técnicas que possibilitem a segmentação em imagens de profundidade utilizando o sensor Microsoft Kinect.



O presente trabalho foi realizado com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – Brasil.

¹ Aluno da graduação, Instituto Federal Catarinense, Campus Videira. Curso de bacharelado em Ciência da Computação. Email: lucas.viana.b.93@gmail.com

² Professor Orientador do Instituto Federal Catarinense, Campus Videira. Curso de bacharelado em Ciência da Computação. E-mail: wanderson.rigo@ifc-videira.edu.br

³ Professor Co-orientador do Instituto Federal Catarinense, Campus Videira. Curso de bacharelado em Ciência da Computação. E-mail: manasses@ifc-videira.edu.br

Inicialmente discorre-se sobre os sistemas de visão artificial, depois são abordados os procedimentos metodológicos utilizados no trabalho, seguido dos testes realizados e por fim as discussões e conclusões fomentadas pelos resultados obtidos.

METODOLOGIA

Este trabalho desenvolveu-se entre julho de 2013 e junho de 2014 totalizando 13 meses via:

1. Revisão bibliográfica sobre as técnicas de segmentação de imagens em profundidade mais utilizadas atualmente.
2. Obtenção de imagens capturadas pelo sensor Microsoft Kinect, para aplicação das técnicas de segmentação revisadas anteriormente.
3. Tentativa de segmentação das imagens obtidas na Internet (banco de imagens).
4. Análise dos resultados obtidos.

Inicialmente foi realizado um levantamento bibliográfico sobre as técnicas de segmentação de imagens existentes atualmente. O estudo incluiu: algoritmos e técnicas de reconhecimento facial, extração de características faciais, classificação de folhas de árvores (através de formato, cor, textura e características), extração automática de características de imagens, histogramas de gradientes orientados (para detecção de humanos em imagens), reconhecimento de padrões [3], reconhecimento real de objetos (utilizando invariantes do momento), métodos de *threshold* para histogramas gradientes (em escala de cinza) [4], detecção de bordas (*edge detection*)[1], entre outros assuntos implícitos nos citados.

Em seguida, fez-se necessária a utilização do Microsoft Kinect para efetuar a captura das imagens que seriam utilizadas no projeto. A meta era obter imagens de faces de pessoas e utilizá-las para efetuar a segmentação de objetos de interesse (nariz e orelhas, por exemplo). Como o campus não dispunha do sensor, tentou-se encontrar na web algum banco de imagens capturadas com o Kinect, para então sugerir as técnicas de segmentação.

Encontrou-se um banco de dados contendo fotos bidimensionais (cerca de que 15.000 fotos de pessoas) disponibilizado na internet por um grupo de pesquisadores de imagens em profundidade (Biwi Head Kinect Head Pose Database), juntamente com um pequeno trecho de código fonte (escrito em C++), com a função de abrir uma imagem do banco e transportá-la para um vetor bidimensional. Como o código não provia nenhuma função além da citada, foi necessário efetuar a inclusão da biblioteca OpenCV⁴ (*Open Source Computer Vision Library*), para que fosse possível efetuar o processamento das imagens (visualização, segmentação, etc.).

Após isso, um grave problema foi encontrado. Como o banco de dados não provia nenhum tipo de documentação (as tentativas de contato com o pesquisador que o disponibilizou não foram respondidas), ficou extremamente difícil descobrir como os arquivos binários disponibilizados deveriam ser processados, dificultando o entendimento de como deveria ser feito o processamento das imagens disponibilizadas. Apesar da falta de documentação, via grande esforço, tentou-se decodificar os arquivos binários (os resultados são exibidos na seção seguinte).

Infelizmente, como o campus ainda não dispõe de um sensor Microsoft Kinect não foi possível o processamento de imagens próprias. Além disso, as imagens presentes no banco utilizado, apesar de terem sido capturadas com um Kinect, não eram especificamente as imagens mais favoráveis a este trabalho, pois eram imagens capturadas com a câmera do Kinect, e não com seu sensor de profundidade. O máximo que poderia ser feito era a segmentação bidimensional (caso o banco apresentasse sua documentação), pois as imagens não continham dados de profundidade.

RESULTADOS E DISCUSSÕES

Apesar das dificuldades encontradas, conseguiu-se efetuar a segmentação e, em seguida, dar destaque das silhuetas das imagens das pessoas fotografadas e gravadas no banco. Alguns exemplos imagens oriundas do banco de imagens são mostrados a seguir:

⁴ <http://opencv.org>

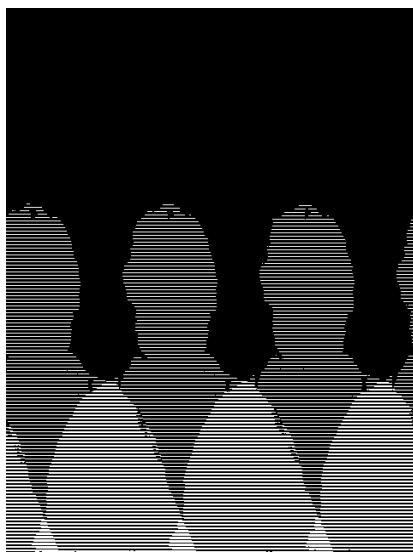


Imagen 1.



Imagen 2.

Abaixo mostra-se a imagem 1 e a imagem 2 segmentadas. Note o destaque das silhuetas (em branco).



Como pode-se observar, não foi possível efetuar o destaque correto das silhuetas presentes nas imagens, pois, sem a documentação do banco de imagens, foi impossível descobrir como lidar com os arquivos binários disponibilizados (o que não seria o caso se pudéssemos gerar nossas próprias imagens com um Kinect).

CONSIDERAÇÕES FINAIS

A proposta inicial do projeto foi uma ótima idealização de aplicação da visão computacional em outras áreas do conhecimento (jogos digitais, medicina, segurança, etc.), o que acabaria facilitando e poupano trabalho humano (trazendo consequentemente mais segurança e confiabilidade para sistemas computacionais que lidam com imagens e ainda dependem da avaliação humana). Porém,

dificuldades insuperáveis foram encontradas no decorrer do projeto (a indisponibilidade do sensor Microsoft Kinect e insucesso na busca por um banco contendo as imagens necessárias para o andamento do mesmo). Tais intercorrências acabaram limitando/podando os resultados que almejava-se alcançar. No entanto o *know-how* adquirido pelo pesquisador fica como principal legado deste projeto, além do caminho trilhado, que pode ser explorado por outros pesquisadores em trabalhos futuros.

REFERÊNCIAS

- [1] CANNY, J. *A computational approach to edge detection and Machine Intelligence, IEEE Transactions on, Pattern Analysis* 1986. PAMI-8, n. 6, p. 679-698, 1986. ISSN 0162-8828.
- [2] FORSYTH, David A.; PONCE, Jean. *Computer vision: a modern approach*. Upper Saddle River: Prentice Hall, 2011.
- [3] GONZALEZ, Rafael C.; WOODS, Richard E. *Processamento de Imagens Digitais*. São Paulo: Edgar Blücher, 2005.
- [4] OTSU, N. *A threshold selection method from gray-level histograms*. Transactions on Systems, Man and Cybernetics, IEEE 1979. v. 9, n. 1, p. 62-66, jan. 1979.