

Rastreamento e Reconhecimento Estrutural de Mão

Autores

Daniel Ferreira Iwahashi

Orientador

Luis Augusto Consularo

Apoio Financeiro

Fapic

1. Introdução

A interação com objetos virtuais em uma aplicação de Realidade Aumentada é um problema comumente resolvido com dispositivos rastreadores ópticos que tomam marcadores como referência. Estes marcadores têm uma forma previamente conhecida e sua forma na imagem projetada é utilizada para inferir posição e orientação. É desejável que seja possível interagir sem quaisquer marcadores, porém, para isso é necessário rastrear gestos ou partes do corpo como a face, as mãos ou o próprio corpo. Este artigo propõe o estudo do problema de rastreamento de mãos, contribuindo, além disso, com a implementação de um casamento entre uma representação estrutural adquirida do vídeo em tempo-real e modelos estruturalmente simplificados armazenados.

Um objeto para rastreamento bastante usado na Realidade Aumentada é um marcador, que consiste em uma placa de papelão com um simples desenho (padrão) dentro de uma moldura quadrada. Este padrão é previamente conhecido pelo software em uma biblioteca de padrões, dando a liberdade de redimensionamento, rotação, corte ou deformação de algum objeto virtual. Um dos problemas é a limitação do usuário por ter que portar esses marcadores em seu corpo, para tornar possível a interação.

Um grande anseio para um dispositivo de Realidade Aumentada é não precisar portar mais nenhum objeto de extensão de nosso corpo para interagir com a virtualidade. Uma parte de nosso corpo, condizente com tal papel, é a mão, porém, surgem dificuldades como a mutação da forma da mão, a cada momento em uma posição e tamanho, mas pode ser tomada também como uma solução para identificar gestos e liga-los a uma ação de interação no ambiente.

Uma estratégia adotada é reconhecer a cor da pele nesse vídeo capturado em tempo real. Obtida a segmentação, a próxima etapa é encontrar pontos relevantes da forma, que permitirão identificar pontas de dedo e reentrâncias das mãos e, conseqüentemente, separar os dedos, a palma da mão e a região do pulso (CONSULARO, 1999).

Uma vez que as partes relevantes estejam separadas, é necessário definir uma estrutura para o gesto da mão. Essa estrutura, representada por grafos, é comparada com uma estrutura em uma biblioteca de gestos, também representada por grafos, em um grau de similaridade, estabelecendo-se uma relação entre o modelo estrutural e a estrutura adquirida em tempo real (STENGER, 2001).

2. Objetivos

O mapeamento da realidade sobre a virtualidade, independente de algum objeto extensor é um grande desejo. Uma parte do corpo interessante para tal ação é a mão, portanto, é necessário reconhecê-la dentro de cada quadro de um vídeo.

Segmentando a mão, que é nossa forma de interesse, encontrando regiões de alta curvatura, que na mão são identificadas como pontas de dedo e áreas entre os dedos, conhecidas como reentrâncias. Essas áreas de alta curvatura são áreas de interesse da mão e, a partir delas, podemos definir pontos e estruturá-la através de grafos e depois compará-las a um vocabulário de modelos de gestos, também estruturados em forma de grafos, podendo assim, interagir numa cena virtual simples.

Capturada por uma câmera web, através de seu CCD e transferida por uma interface serial USB ou FireWire, a imagem digital tem uma estrutura matricial onde cada elemento dessa matriz possui informações de intensidade de luz captada pelos fotossensores da câmera. O acesso dessas informações se faz por meio de um ponteiro na memória fornecido pela chamada do sistema operacional ao driver do dispositivo de aquisição. Tendo, então, acesso a essas informações, o próximo objetivo é colocar em prática o processo de identificação e localização do objeto na cena real, no caso a mão e os gestos executados pela mão do usuário.

O Anexo 1 mostra os objetivos parciais para se chegar ao resultado final.

3. Desenvolvimento

Utilizando um algoritmo para segmentação da cor da pele proposto por CHENG-CHIN (CHENG-CHIN, et al, 2003) o resultado não foi satisfatório

Em testes, utilizando o espaço de cores HSV, observou-se que a cor da pele sempre se localizava em uma determinada região do espaço, mesmo em tonalidades mais escuras da pele. d *alpha*. d entre cada aresta do modelo e cada aresta de entrada, e as diferenças de inclinação α entre um par de arestas do grafo de modelo e um par do grafo de entrada. Com soma da diferença de distância e as diferenças de inclinação, obtém-se um custo. Cada aresta de entrada é comparada com todas do modelo, retornando apenas a comparação de menor custo, varrendo-se todas as arestas do grafo de entrada e comparando com um do modelo. e, em cada par de aresta temos um ângulo de inclinação

Tendo em conta mais essa informação, estabeleceu-se mais uma condição para que o elemento seja tomado como cor da pele, trabalhando com valores menores que 90 para o Tom (H representado no eixo de cor verde no gráfico) e observou-se uma grande melhora na segmentação da mão.

Para corrigir o problema do reconhecimento de região de não-interesse (falso-positivo), a saída encontrada foi segmentar apenas regiões de pele conectadas, fazendo assim, com que partes que não fossem de interesse, como ruídos dispersos, fossem eliminadas, resultando na segmentação adequada.

Uma vez que a região de interesse tenha sido identificada, a próxima etapa é encontrar as extremidades dessa região, ou seja, identificar o que é uma ponta de dedo ou uma reentrância e separá-las do que é tido como palma da mão. Porém, para tal tarefa, é necessário definir a borda dessa segmentação.

Encontrar as bordas é uma etapa relativamente simples, basta correr a vizinhança deslocando o ponteiro como se fosse uma bengalhinha sempre para um mesmo sentido, dando a volta em todo o objeto. Os demais

pixels, ou seja, os pixels em que a bengala não passou, que são pixels do miolo, são ignorados, resultando apenas o contorno da mão.

Este contorno, contudo, é muito ruidoso (em forma de escada) e, para analisarmos e extrairmos algo que seja extremidade com tal resultado, necessitamos de relativamente mais processamento e podendo, provavelmente, gerar falsos positivos de extremidades. Então, é necessário suavizar esse circuito que forma a mão, extraindo a média das posições de cada pixel componente do contorno e seus adjacentes. Essa adjacência é chamada de janela, que é composta por um pixel central e seus vizinhos. Uma janela de tamanho três, por exemplo, significa um pixel central (que recebe a média das posições de seus vizinhos), um pixel à direita e outro à esquerda. Se aumentarmos o tamanho dessa janela, obtemos como resultado uma maior suavização de contorno.

Depois de suavizado, podemos detectar as regiões de alta curvatura. Essa detecção se dá por meio de janelas (exatamente como visto anteriormente, porém, sem média, e sim cálculo de curvatura).

A curvatura se dá pela divisão da distância entre os pixels extremos da janela dividido pelo comprimento de arco dessa janela.

Calculada a curvatura, obtemos qual pixel, de todo o contorno, é um pixel de alta curvatura.

Numa detecção perfeita, num gesto de mão aberta, com todos os dedos em frente à câmera, temos como resultado nove extremidades, como mostra o Anexo 2, em que os pontos vermelhos são as extremidades.

Para o reconhecimento estrutural de uma mão, é necessário fazer uma comparação (um casamento) de uma mão em frente à câmera web e algum modelo estrutural armazenado. Essa estrutura é muito bem representada por um grafo de gesto, em que cada extremidade da mão é um vértice desse grafo. Entre dois vértices temos uma aresta, que possui uma distância

Encontrada as extremidades da mão no vídeo, é possível construir um grafo estrutural de entrada, podendo ser casado com o grafo de gesto de modelo. O Anexo 2 representa a construção de um grafo de entrada em tempo real.

O método utilizado para fazer a busca do casamento entre os grafos de modelo e de entrada é o beam-search, no qual são examinadas as ramificações em uma árvore de comparações entre o modelo e a entrada, guardando, em uma fila de prioridades, as ramificações mais bem sucedidas em cada nível (CESAR et al, 2005). O casamento é baseado em comparações, em que são conferidas as diferenças de distância.

Sedláek (SEDLÁEK, 2004) sugere que o aumento da saturação torna a segmentação mais fácil, contudo, para aumentar esse valor, é necessário converter o espaço de cores RGB para HSV (H = Tom, S = Saturação e V = Valor de Brilho). É possível transformar os pixels da imagem da mão para este espaço, mas mesmo assim a segmentação resultou em muitos ruídos. Tais ruídos podem ser eliminados utilizando técnicas de processamento de imagens digitais, como a erosão e a dilatação, ou seja, a imagem binária do objeto segmentado é corroída e depois dilatada na mesma proporção.

4. Resultados

A detecção da mão no vídeo foi de um bom resultado, cuja robustez se refletiu nos resultados das próximas etapas do projeto, como mostra o Anexo 3a.

O próximo resultado é encontrar o limite do objeto segmentado, que significa extrair as bordas da

segmentação, para poder depois, detectar pontas de dedo e reentrâncias, como mostra o Anexo 3b.

As extremidades de uma mão são representadas pelas pontas de dedo e pelas regiões entre dedos, caracterizadas por serem regiões de alta curvatura sobre as demais regiões. Encontradas nesse contorno e plotadas na tela, como mostra o Anexo 3c.

A partir daí, efetuar um casamento de grafo-modelo com grafo-entrada, a partir de comparações entre cada aresta desses grafos e, após isso, reconhecer estruturalmente uma mão no vídeo, como mostra o Anexo 3d.

O uso de grafos para a representação estrutural de gestos de mão possibilita uma grande liberdade de gerarmos uma biblioteca de gestos, contendo estruturas com um dedo, dois dedos, três dedos e assim por diante.

Com o reconhecimento estrutural, é possível, por exemplo, substituir um padrão de marcador utilizado em uma aplicação de Realidade Aumentada por um gesto de mão. Um segundo marcador por um segundo gesto, etc., podendo rastrear posições da mão na tela.

5. Considerações Finais

Os resultados deste trabalho se mostraram úteis para a segmentação de partes do corpo (segmentação de objetos de cor da pele), na detecção de contornos, na identificação de partes de interesse em um contorno, na identificação de estruturas em imagens e, por fim, no reconhecimento estrutural de mão.

Muito do que foi realizado neste trabalho são métodos já conhecidos, porém a imposição das restrições da aquisição de vídeo sempre sugere novos desafios para que os resultados sejam alcançados.

A própria biblioteca de Realidade Aumentada utilizada neste trabalho (ARToolkit) não contempla a identificação estrutural e nem o reconhecimento de gestos, de modo que este trabalho contribui também para que, depois de organizado na forma de uma biblioteca, os resultados possam ser reutilizados por alunos e pesquisadores da área de Realidade Aumentada.

Deste trabalho também resulta uma contribuição para, pelo menos, duas áreas de pesquisa de grande interesse atual: a interação 3D com objetos virtuais e o reconhecimento de gestos. No primeiro caso, duas ou três câmeras poderiam ser utilizadas simultaneamente para que a estrutura 3D da mão pudesse ser reconhecida. No segundo, a variação da estrutura poderia em função do tempo poderia ser explorada para reconhecer gestos em movimento.

Referências Bibliográficas

CHENG-CHIN, Chiang; WEN-KAI, Tai; MAU-TSUEN, Yang; YI-TING, Huang; CHI-JAUNG, Huang. A novel method for detecting lips, eyes and faces in real time. *Real-Time Imaging*. v. 9, nro 4, p277-87, 2003.

CESAR, Roberto M.; BENGOTXEA, Endika; BLOCH, Isabelle; LARRAÑAGA, Pedro. Inexact graphmatching for model-based recognition: Evaluation and comparison of optimization algorithms. *Pattern Recognition*. v. 38, p2099-2113, 2005.

CONSULARO, Luís A.; COELHO, Regina C.; FERNANDES, M.M. Tutorial sobre ARToolkit. Disponível em . Acesso em: 10 de setembro de 2005.

CONSULARO, L.A.; CESAR JR., RM. Quadtree-based Inexact Graph Matching for Image Analysis, In: Proceedings of the XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAP'05). P1530-1834, Manaus-AM, Brasil,2005.

CONSULARO, L.A.; CESAR JR.; R.M.; COSTA, L.F. Synergos and its Applications to Contour Segmentation, In: PROCEEDINGS OF THE I INTERNATIONAL SEMINAR ON BIOELECTRONIC INTERFACES AND III WORKSHOP ON CYBERNETIC VISION. p77-83, Campinas-SP, Brasil,1999.

DEITEL, H. M. Como Programar em C. LTC: Rio de Janeiro, 1999. 486pp.

SEDLÁ EK, M.

SIGAL, Leonid; SCLAROFF, Stan; ATHITSOS, Vassilis. Skin Color-Based Video Segmentation Under Time-Varying Illumination. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 26, nro 7, p862-77. 2004.

SORIANO M., HUOVINEN S., MARTINKAUPPI B., LAAKSONEN M. Using the skin locus to cope with changing illumination conditions in color-based face tracking. In: IEEE NORDIC SIGNAL PROCESSING SYMPOSIUM, Kolmarden, Sweden, 2000. p. 383–6.

STENGER, B.; MENDONÇA, P.R..S.; CIPOLLA, R. Model-based 3D tracking of an articulated hand. In: PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, Vol. II, p310-315, Kauai, Estados Unidos, 2001.

TENENBAUM, Aaron M. Estruturas de Dados Usando C. Makron: São Paulo, 1995. 884pp.

WikiPedia Encyclopedia. Transformation from RGB to HSV. Disponível em: Acesso em 27 de outubro de 2005.

Evaluation of RGB and HSV models in Human Faces Detection. In: 10TH CENTRAL EUROPEAN SEMINAR ON COMPUTER GRAPHICS (CESCG). Disponível em: . Acesso em: 14 de fevereiro de 2006.

Anexos



