

Rastreamento de olhos em tempo real utilizando webcam no espectro de luz infravermelho

Antônio Celso Caldeira Júnior
Universidade Federal de Minas Gerais
Departamento de Ciência da Computação
caldeira@dcc.ufmg.br

Resumo

Este trabalho trata do rastreamento dos olhos e da extração de informação de distância até o usuário em tempo real utilizando uma webcam de baixo custo alterada para emitir e captar apenas luz no espectro infravermelho.

1. Introdução

Uma das áreas de pesquisa em Visão Computacional é dedicada para a implementação de sistemas que detectam e extraem informações dos movimentos e gestos do usuário[1, 2, 3, 4, 5]. Este trabalho tem por objetivo extrair informações a partir da localização dos olhos do usuário utilizando recursos mínimos.

O objetivo é atingido utilizando idéias baseadas no trabalho de Grauman[5] e adaptadas para a utilização da característica do piscar involuntário do usuário para a detecção em tempo real e sem treinamento dos olhos do usuário.

A principal contribuição deste trabalho é utilizar webcams de baixo custo e executar em tempo real utilizando apenas a informação do piscar do usuário em conjunto com o espectro infravermelho (não refletido pela retina).

2. Trabalhos Relacionados

O trabalho de Grauman[5] descreve um sistema que utiliza o piscar e o movimento das sombra-celhas para manipular a iteração de clicks de um mouse. Ele descreve o seu algoritmo que utiliza padrões pré-definidos de olhos para treinar o sistema e melhorar as taxas de acerto.

O trabalho de Chau[1] e Betke[3] descreve uma arquitetura que utiliza as idéias de Grauman de forma

robusta, garantindo que altas velocidades enquanto mantem as taxas de acertos elevadas.

Estes artigos não tratam da questão do infravermelho, a retina não reflete a luz infravermelha o que pode garantir uma maior facilidade na detecção do olho. Os trabalhos citados utilizam desta característica para fazer o *pattern matching* mas não utiliza-na em conjunto com o piscar.

3. Metodologia

O primeiro passo para a execução do trabalho foi a alteração de uma webcam de baixo custo para a detecção de apenas luz no espectro do infravermelho. Isto foi feito utilizando-se alguns tutoriais da internet[10, 12]. A webcam utilizada foi uma Multilaser[8] de baixo custo, também foi utilizado um canhão para iluminação infravermelho[9], este canhão foi fixado no topo da webcam. A metodologia de alteração da câmera foi um retalho dos tutoriais citados. A alteração da câmera consiste em três passos básicos:

1. Retirar o filtro infravermelho
2. Adicionar o filtro de luz visível
3. Adicionar iluminação infravermelho (duas camadas de filme queimado de máquina de retrato analógica)

As Figuras 1 e 2 mostram o filtro infravermelho retirado da webcam e a webcam remontada com o filtro de luz visível e iluminação infravermelho, respectivamente.

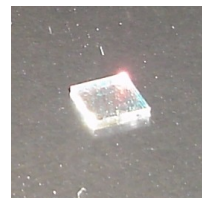


Figura 1. Filtro infravermelho



Figura 2. Webcam remontada

A idéia original era utilizar a retina para identificar o olho do usuário, entretanto, como foi utilizada uma câmera de baixo custo só foi possível fazer isso quando o olho estava muito perto da câmera, restrição considerada proibitiva visto que o objetivo não era utilizar câmeras instaladas em óculos, mas fixadas próximo do monitor. A Figura 3 mostra uma foto tirada utilizando a webcam montada com iluminação infravermelha a cinco centímetros da câmera. Pode-se perceber que a retina é claramente identificada pelo fato de não refletir o infravermelho. O ponto branco dentro do círculo é reflexão da córnea.



Figura 3. Retina vista no infravermelho

Após a montagem da webcam foi implementado o sistema de Visão Computacional utilizando o OpenCV[7]. A Figura 4 mostra um diagrama da estrutura do sistema em baixo e em alto níveis. Em alto nível, a principal característica do sistema é a utilização do piscar para detectar dos olhos, e, a partir daí rastrea-los. A utilização do piscar faz com que a detecção aconteça sem nenhum tipo de treinamento do sistema com olhos detectados. No diagrama baixo nível a cada novo quadro é utilizado um buffer de tamanho dois, isto é, dois quadros anteriores são armazenados, e, a partir deles, é calculado o histórico de movimento no intervalo de tempo de três frames. É importante notar que uma piscadela dura aproximadamente 10 frames em um filme de 30 fps (0,3 segundos)[3]. No caso da câmera utilizada o frame rate foi de 10 FPS, este é o motivo da escolha de dois quadros anteriores no buffer (dois anteriores + atual = três). Em posse destes três quadros é realizada a operação de diferença entre eles e aplicado o limiar para a binarização desta imagem. Na imagem binária são calculados os contornos da

imagem, possibilitando assim, o cálculo das elipses que melhor se ajustam à imagem. Neste ponto, dependendo da quantidade de movimento do usuário, podem surgir várias elipses, apenas quando são encontradas duas elipses o algoritmo assume que estas duas elipses são os olhos do usuário e começa o processo de rastreamento utilizando o algoritmo de fluxo óptico de Lucas-Kanade (versão iterativa). Finalmente a distância euclidiana é calculada a partir dos centros das duas elipses rastreadas.

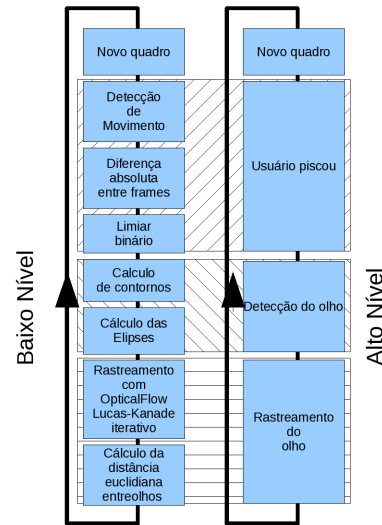


Figura 4. Metodologia

4. Resultados

Na etapa de aquisição do frame é utilizada uma imagem de 8-bits e apenas um canal. Os canais de cores são adicionados para fim exclusivo dos marcadores identificando os olhos. A Figura 5 mostra um exemplo de imagem adquirida pela câmera onde foram detectados os olhos do usuário.



Figura 5. Imagem adquirida adicionada de marcadores

A detecção de movimento gera duas imagens de saída no sistema. A Figura 6 é resultado da diferença

entre os quadros do buffer e o quadro atual. A Figura 7 é o resultado da aplicação de um limiar para binarizar a imagem.



Figura 6. Diferença entre quadros no momento do piscar



Figura 7. Limiar aplicado na imagem de diferenças

Na imagem binária representada pela Figura 7 é aplicado o algoritmo de cálculo dos contornos onde, por sua vez, é aplicado o algoritmo de cálculo e ajuste contornos à elipses, este último é apresentado na Figura 8.



Figura 8. Elipses calculadas

O algoritmo de detecção de elipses retorna o número de elipses encontradas. Com esta informação foram realizadas diversas otimizações para minimizar a quantidade de falsos-positivos e maximizar a taxa de detecção.

A primeira otimização implementada foi assumir que o algoritmo encontrou corretamente os olhos quando apenas uma ou duas elipses forem retornadas (parâmetro do sistema) pois é não é possível que o usuário possua três olhos.

A segunda otimização implementada foi exigir que os olhos são detectadas quando uma ou duas elipses (parâmetro do sistema) forem detectadas em dois quadros consecutivos; isto garante que falsos-positivos levantados por movimentação aleatória ou ruídos sejam filtrados, entretanto isto aumenta o tempo do piscar do usuário para 0,4 segundo, que é um pouco acima da média. Este problema pode ser resolvido utilizando uma taxa de quadros por segundo mais levado, pois neste caso, pode-se variar o escopo de quadros analisados em até 10 sem necessidade de esperar que o piscar do usuário seja maior que o normal.

O algoritmo de detecção de elipses retorna o centro de cada uma das elipses encontradas. Assim, em posse deste dado, é possível calcular a distância euclidiana entre estes dois centros. O resultado é dado em pixels. A Figura 9 mostra uma representação da máxima possível (os dois centros em cantos opostos da elipse) subtraída da distância entre os centros das elipses.



Figura 9. Representação da distância: a barra vermelha corresponde à distância máxima possível subtraída da distância entre os centros das elipses encontradas (escrito na imagem). O desenho verde representa a webcam, o desenho azul representa a cabeça do usuário

Em posse da informação de distância do usuário até a câmera pode-se criar uma aplicação que, por exemplo, diminui o brilho do monitor quando o usuário se aproxima dele, para evitar danos à retina.

6. Conclusões

O algoritmo implementado apresenta algumas restrições pelo fato da utilização do piscar como principal característica para a detecção dos olhos. A principal restrição é a necessidade do usuário permanecer imóvel para a detecção de seus olhos. Entretanto, todo o sistema foi desenvolvido requisitando apenas a definição de apenas um limiar (de binarização da imagem). Foram aplicadas várias técnicas de otimização para garantir uma eficiência mínima do sistema. Vale lembrar que o algoritmo implementado não faz uso de técnicas de treinamento prévio com imagens de olhos, garantindo a sua eficácia independente de treinamento.

7. Trabalhos Futuros

A necessidade de avaliar a performance e a taxa de acerto e de falsos-positivos se torna evidente neste trabalho, então, o primeiro trabalho futuro é esta avaliação.

A implementação de uma aplicação de redução de brilho a partir da distância do usuário à câmera, descrita no final da seção 5 também seria importante para validar a informação extraída através do sistema

Ouro trabalho futuro é utilizar uma webcam normal no espectro da luz visível e avaliar o desempenho do sistema.

Por fim, utilizar uma técnica de treinamento para delimitar a região de interesse e reduzir a necessidade do usuário permanecer imóvel para a detecção do piscar também é interessante.

8. References

[1] M. Betke, J. Gips, and P. Fleming. The camera mouse: Visual tracking of body features to provide computer access for people with severe disabilities. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 10:1, pages 1-10, March 2002.

[2] T.N. Bhaskar, *et al.* Blink detection and eye tracking for eye localization. *Proceedings of the Conference on Convergent technologies for Asia-Pacific Region (TENCON 2003)*, pages 821-824, Bangalore, India, October 2003.

[3] M. Chau and M. Betke. Real Time Eye Tracking and Blink Detection with USB Cameras. *Boston University Computer Science Technical Report No. 2005-12*, May 2005.

[4] S. Crampton and M. Betke. Counting fingers in real time: A webcam-based human-computer interface with game applications. *Proceedings of the Conference on Universal Access in Human-Computer Interaction (affiliated with HCI*

International 2003), pages 1357-1361, Crete, Greece, June 2003.

[5] K. Grauman, *et al.* Communiation via eye blinks and eyebrow raises: video-based human-computer interfaces. *Universal Access in the information society, Volume 2, Numer 4*, November 2003.

[6] Infrared Webcam Conversion
<http://www.youtube.com/watch?v=yC3kJ2zjcbc&feature=related>

[7] Intel Corp. OpenCV Library.
<http://opencvlibrary.sourceforge.net/>

[8] Multilaser, “*WEBCAM OLHO 350K – WC31102*”.
http://www.multilaser.com.br/produtos_perifericos12.php?acao=detalhes&id_prod=4519

[9] MercaloLivre: Canhão Infravermelho,
http://produto.mercadolivre.com.br/MLB-69783151-canho-infravermelho-led-infra-vermelho-infrared-12-leds-_JM

[10] Take Infrared Pictures With Your Digital Camera – AOL Video, <http://video.aol.com/video-detail/take-infrared-pictures-with-your-digital-camera/879829136>

[11] Wikipedia. Eye tracing.
http://en.wikipedia.org/wiki/Eye_tracking

[12] Webcam Microscope – Hacked Gadgets – DIY Tech Blog,

<http://hackedgadgets.com/2008/04/03/webcam-microscope/>