

Um Sistema de Captura de Movimentos de Baixo Custo para Animação de Personagens

Matheus Pereira, Agostinho Júnior

¹ Departamento de Engenharia da Computação e Automação
Universidade Federal do Rio Grande do Norte (UFRN) – Natal, RN – Brazil

matheus.petrovich@gmail.com, ambj@dca.ufrn.br

Resumo. *Este artigo descreve o desenvolvimento de um sistema de captura de movimentos portátil e de baixo custo para assistir o processo de animação de personagens, reduzindo o tempo gasto pelo animador e atribuindo maior realismo ao produto final. O sistema é baseado no processamento de imagem oriunda de uma câmera afixada em um aparato de captura que a mantém estável em relação ao rosto do ator. O sistema consegue rastrear com êxito 13 marca-dores ao redor da boca e sombrancelhas e ainda as pupilas do ator.*

Abstract. *This article describes the development of low cost and portable motion capture system to aid with the process of character animation. Such systems reduces the time spent by the animator and gives more realism to the final animation. The system is based on image processing using a camera fixed to a helmet that remains stable in relation to the actor's face. The system can successfully track 13 markers around the mouth and eyebrows and even the actor's pupils.*

Introdução

A sutileza das expressões faciais representa parte significativa do processo de comunicação humana. Pequenas nuances podem separar um sorriso verdadeiro que passe aceitação, de um forçado que passe desconforto. Buscando um nível cada vez maior de realismo, a indústria do entretenimento tem usado diversos artifícios para tornar seus produtos cada vez mais realistas e verossímeis, tanto nas formas e qualidade da imagem, quanto na animação. Esse realismo é normalmente obtido usando técnicas de captura de movimentos [Moeslund and Granum 2001].

São os pequenos detalhes que agregam realismo ao desempenho do personagem e acrescem ao individualismo de seu comportamento. Esses podem ser difíceis de se criar através de métodos mais tradicionais de animação como o *keyframe*, no qual todos os movimentos do modelo são definidos manualmente pelo animador [Fatih Erol 2007].

A técnica de captura de movimentos (*motion capture*, ou *mocap*) permite que conjuntos de movimentos, capturados de um ator no mundo real, sejam inseridos na animação de um personagem virtual. Aplicando diretamente os dados extraídos do ator ao personagem, são inseridos os macro movimentos conscientes, bem como os micro movimentos normalmente associados às emoções. De tal forma, o animador pode gerar personagens com maior grau de realismo, conferindo agilidade ao processo de animação [et al. 2013].

O método de animação por *keyframe* pode ser trabalhoso e demorado, podendo exigir mais de um dia para um animador completar 10 segundos do rascunho de animação

chamado de *blocking*. O uso de métodos baseados em captura de movimento pode reduzir esse tempo em aproximadamente 7 vezes. Autores como Shiratori, da *Disney Research*, reforçam que a agilidade no processo de *blocking* da animação é importante uma vez que nessa etapa de criação, a agilidade é essencial para que os animadores possam concretizar a cena que têm em mente. Essa economia de tempo ainda permite que eles explorem e experimentem muito mais possibilidades com menos esforço, construindo uma base de movimentos que podem ser aprovados pelos diretores e então refinados para uma versão final [Shiratori et al. 2013].

Existem diversas técnicas de captura de movimentos, com sensoriamento baseado em fibras ópticas, sensores eletromagnéticos, potenciômetros embarcados em uma roupa apropriada e processamento de imagem. Essas tecnologias podem ser separadas em duas grandes classes: com marcadores e sem marcadores, sendo a primeira a mais evidente no mercado atualmente [Sigal and Koltun 2012]. Atualmente, sistemas de *mocap* baseados em processamento de imagem estão se tornando cada vez mais comuns dada a evolução dos trabalhos científicos na área.

Métodos baseados em marcadores passivos são mais largamente utilizados na indústria pelo fato de ser mais simples rastrear regiões de cor ou formato conhecido (marcador), e permitir maior flexibilidade ao ator. Em contrapartida, métodos que dispensam o uso de marcadores tendem a apresentar limitações no rastreamento em tempo real e uma maior margem de erro [Vicon].

Dentre os sistemas de rastreamento de movimento, os de detecção de expressões faciais estão entre os mais difíceis de serem concebidos, posto que devem ter a precisão para captar as pequenas nuances de expressões faciais. Essas nuances são as principais responsáveis pelo realismo de um personagem, especialmente o movimento dos olhos.

Bons sistemas de rastreamento de características faciais são caros e normalmente devem ser adaptados a cada tipo de ambiente. Além disso, necessitam de aparato especial a ser colocado no ator para que o rastreamento da face possa ser feito *in loco* enquanto esta atua no espaço de gravações.

Sistemas de baixo custo podem ser interessantes para pequenos estúdios que não dispõem de muitos recursos mas necessitam de usar *mocap* para animar personagens em filmes ou jogos digitais.

Esse trabalho visa contribuir apresentando uma proposta de um sistema de captura de movimento de baixo custo para faces usando marcadores passivos. O sistema é composto por um dispositivo de captura ao qual está acoplada a câmera, e um *software* de processamento de imagens que recebe e trata o *stream* de vídeo em tempo real para uso na animação de personagens.

O presente artigo é organizado nas seguintes seções. A seção 2 descreve o modelo de aparato desenvolvido para realizar a captura e as técnicas de captura de movimento que foram empregadas. A seção 3 descreve como se dá o processo de rastreamento das regiões das sombrancelhas e da boca. A seção 4 mostra o processo usado no rastreamento das pupilas. A seção 5 conclui o trabalho, apresentando perspectivas de aprimoramento e expansão da utilidade da ferramenta¹.

¹Algumas imagens no trabalho foram premeditadamente modificadas na versão de apreciação pelos

Modelo do sistema de captura

O aparato de captura é composto de um capacete em resina de poliéster reforçada com fibra de vidro e uma câmera afixada em um braço de alumínio. O capacete deve ser confeccionado a partir de um molde da cabeça do usuário, obtido em um processo de moldagem com ataduras de gesso. Normalmente, a moldagem é necessária para o capacete fique bem firme na cabeça do seu utilizador, minimizando vibrações da câmera que será colocada em frente ao rosto.

Sobre o modelo foram laminadas camadas de resina de poliéster pigmentada reforçadas com manta de fibra de vidro. Cortes foram feitos na parte traseira do capacete para permitir flexibilidade e uma tira de velcro foi afixada para ajuste, de modo que o capacete fique justo na cabeça do usuário. No interior do capacete foram colados pedaços de feltro para melhorar o conforto e ajudar a deixar o capacete mais fixo, para que a câmera estivesse sempre estática em relação ao rosto do usuário. A estrutura do capacete é mostrado na Figura 1.



Figura 1. Aparato de captura completo.

Dois parafusos foram usados para afixar o braço de alumínio ao capacete. A barra de alumínio necessita ser medida e moldada de forma a permitir que a câmera tenha uma visão completa da face do usuário. Na ponta do braço de alumínio é presa uma câmera USB que fará a captura da imagem da face do usuário.

A fim de não limitar a mobilidade do ator durante a captura, o *software* foi embarcado em um Raspberry Pi afixado à parte traseira do capacete. Através de uma conexão à internet o sistema pode ser remotamente configurado e acompanhado. As imagens são capturadas na resolução de 480 pontos horizontais por 640 pontos verticais. Nesta resolução, o *hardware* do Raspberry consegue capturar imagens com velocidade em torno 20 quadros por segundo de uma câmera USB e processá-las.

Rastreio de marcadores

Os marcadores utilizados neste trabalho devem ser confeccionados bolas de isopor revestidas com pintura tinta amarela e colados na face do ator utilizando o adesivo Pros-Aid[®]. Este adesivo é produzido pela ADM Tronics[®] e utilizado largamente na indústria de entretenimento em maquiagem e efeitos especiais [ADMTronics]. Um exemplo de imagem com os marcadores colados na face do usuário é mostrada na Figura 2.

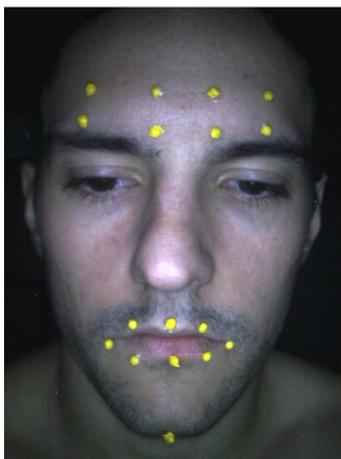


Figura 2. Marcadores Passivos Colados no Rosto do Usuário

Foram realizados testes com bolas isopor em cor natural branca e tingidas com tintas azul, vermelha e verde. Entretanto, a cor amarela foi a que produziu melhores condições de separação do restante dos elementos da imagem.

A segmentação das imagens para extração dos marcadores é feito no espaço de cores HSV, utilizando a componente de matiz (H). A seleção da faixa de matizes adequada para o marcador amarelo deve ser feita conforme as condições de iluminação estabelecidas para o ambiente de captura. Uma ferramenta permite ao utilizador escolher os limiares inferior e superior de limitam a região de matizes amarelas dos marcadores. O uso do model HSV é importante para desvincular a informação da cor das condições de iluminação da cena. Para diversos testes realizados em ambientes com condições de iluminação variada, a faixa $H_0 \in [90, 110]$, considerando uma faixa de varredura para $H \in [0, 180]$, mostrou os melhores resultados para separar os marcadores do restante da cena. A Figura 3 mostra um exemplo de imagem segmentada com os marcadores isolados.

Os marcadores são fixados em pontos estratégicos com áreas de maior movimentação e expressividade e maior relevância para a percepção de emoções como a boca e sombrancelhas [Neth and Martinez 2009]. Para acompanhamento ao longo da captura, cada marcador recebe um rótulo identificador. A quantidade e posição dos marcadores devem seguir a distribuição vista na Figura 2.

A imagem segmentada é então varrida, buscando pelas regiões brancas. Quando uma região é encontrada, assume-se que ali existe um círculo e seu diâmetro é medido. Para evitar a influência do ruído que naturalmente ocorrem durante a segmentação da cena, uma região só será considerada uma região de marcador se tiver diâmetro maior que 7 pixels (valor estabelecido experimentalmente). Caso a região seja aceita, as coordenadas

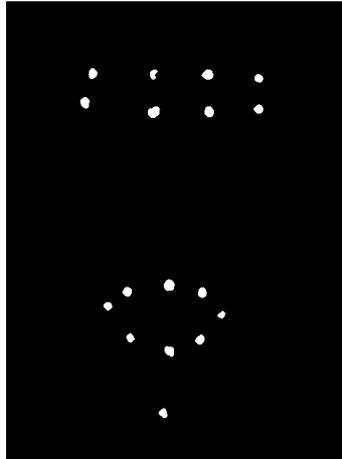


Figura 3. Exemplo de imagem segmentada mostrando os marcadores isolados do restante da cena.

do seu centro de massa serão rotuladas e guardadas para rastreamento. A região é removida para que não seja encontrada múltiplas vezes na mesma cena.

A identificação dos pontos e sua associação com os rótulos é feita considerando-se tanto sua posição na imagem, quanto seu posicionamento em relação aos demais pontos. Para facilitar essa identificação, os pontos da região da boca e os da sombrancelha são separados em conjuntos distintos, de forma que sejam comparados apenas com os demais pontos de sua própria região. Uma vez que a varredura seja finalizada, os pontos são ordenados conforme a ordem crescente de suas coordenadas horizontais. Caso seja a primeira varredura da imagem, os pontos de referência terão de ser criados, caso contrário, eles receberão apenas a informação do deslocamento. Em ambos os casos, a identificação dos rótulos é feita usando esse mesmo algoritmo.

Rastreamento das sombrancelhas

A distribuição dos marcadores na região das sobrancelhas é sempre realizado conforme mostra a Figura 4. A distribuição dos marcadores nessa região é feita em quatro colunas com dois marcadores cada. Dessa forma haverá 4 pares de pontos (dois em cada coluna), onde cada par possui valores próximos para suas coordenadas horizontais. Dessa forma, ordenando os pontos conforme a ordem crescente de suas coordenadas horizontais, os pares de pontos de cada coluna serão sempre subsequentes nesse conjunto ordenado.

Cada par de pontos será comparado quanto às suas coordenadas verticais. O ponto mais próximo ao topo da imagem será classificado como ponto superior e o mais próximo ao rodapé será classificado como ponto inferior. A Figura 5 ilustra a rotulação realizada, destacando os pontos superiores $s1s$, $s2s$, $s3s$ e $s4s$ e os pontos inferiores $s1i$, $s2i$, $s3i$ e $s4i$. O resultado da segmentação é mostrado na Figura 5.

Rastreamento da boca

Apesar da ausência da uniformidade vista na região da sombrancelha, os marcadores da boca são identificados de forma similar. O ponto mais próximo à lateral esquerda é denominado **canto 1**. Os dois pontos seguintes serão comparados usando sua posição vertical. Aquele situado mais próximo à região superior do conjunto será o ponto superior. O

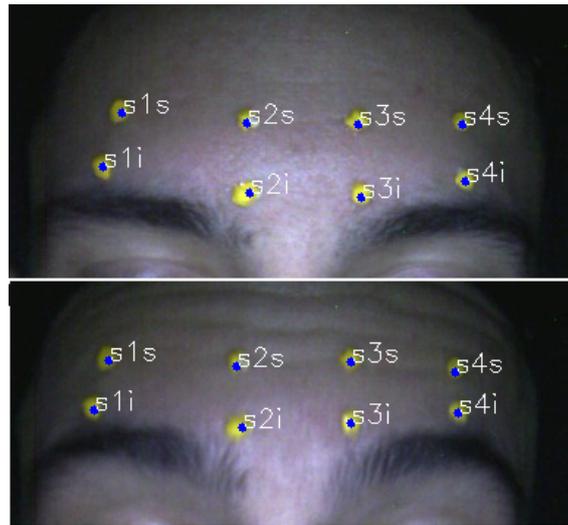


Figura 4. Rastreamento dos marcadores na sombrancelhas.

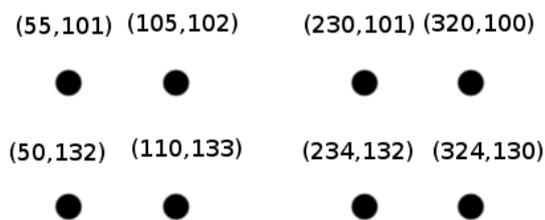


Figura 5. Resultado da segmentação dos pontos na região das sombrancelhas

mais próximo à região inferior do conjunto será o ponto inferior. Em seguida, serão comparados 3 pontos, que se encontra mais abaixo será o ponto do queixo, o que se encontra mais acima será o superior e o outro ponto, será o ponto inferior da boca. Os próximos dois são comparados de forma semelhante ao primeiro par. O último ponto será sempre o **canto 2**.

A Figura 6 ilustra o processo de marcação automática dos pontos da boca em duas situações distintas.

Rastreamento dos olhos

O rastreamento dos olhos do usuário é semi-assistido. Requer mínima intervenção do usuário apenas para criar uma marcação inicial a pupila de cada olho em uma imagem capturada no sistema interativo desenvolvido. Uma vez marcada a pupila aparecerão retângulos sobre os olhos indicando a região processada para o rastreamento, como pode ser visto na Figura 7. Como o aparato mantém a câmera estática em relação ao rosto, a posição dos retângulos não precisa ser corrigida no decorrer da captura.

O rastreamento se dá em duas fases: primeiramente encontra-se a posição horizontal da pupila e então a posição vertical. Cada olho é isolado em uma nova imagem de dimensões iguais a 60 pixels verticais por 110 Pixels horizontais. As imagens são obtidas em tons de cinza a partir do isolamento de um dos canais de cor da imagem original em RGB. Em seguida, as imagens são limiarizadas, assumindo para os pixels valor zero (preto) se for

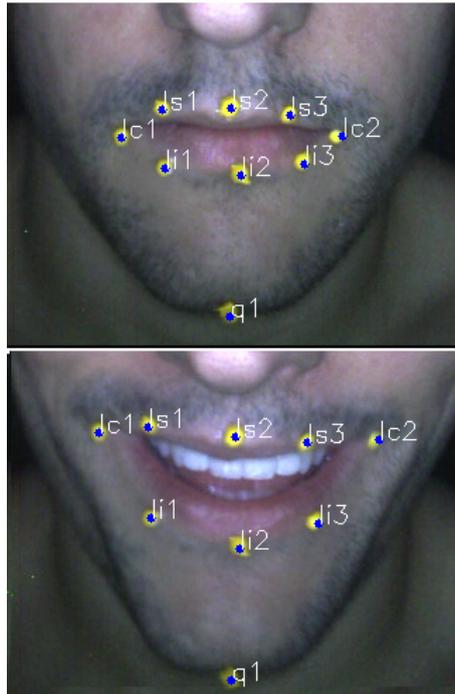


Figura 6. Rastreamento dos marcadores na Boca

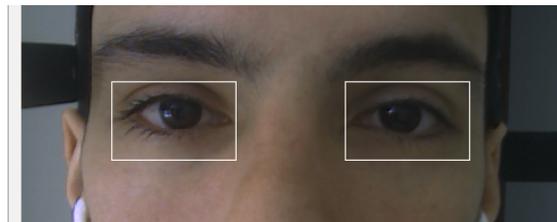


Figura 7. Regiões de rastreamento dos olhos

menor ou igual ao limiar estabelecido, e 255 (branco) caso seja maior que tal limiar. O limiar adequado é definido por meio do software de captura e deve ser ajustado conforme as condições de iluminação da cena.

As regiões escuras dos cílios provocam o aparecimento de ruídos nas laterais dos olhos próximas ao lado externo da face. Para evitar o processamento desnecessário desse ruído, optou-se por realizar a busca pela pupila do centro da imagem em direção às laterais da face, coluna após coluna.

Para cada pixel em uma linha de uma coluna, procura-se pela primeira sequência contínua de cinco pixels pretos que ocorrem na direção vertical. Quando esse evento ocorre, o contorno externo da pupila é encontrado. Para chegar à coordenada horizontal do centro da pupila, caminha-se um total de 10 pixels em direção à lateral da face.

Para determinar a coordenada vertical do centro, caminha-se da borda inferior da janela em direção à superior. A posição do contorno externo é encontrada quando um conjunto consecutivo de 5 pixels pretos é percorrido. O primeiro pixel determina a posição do contorno. Caminha-se um total de 10 pixels em direção à borda superior da imagem e ficam assim determinadas as coordenadas horizontal e vertical do centro da



Figura 8. Regiões de rastreo dos olhos segmentadas

pupila.

A decisão por caminhar de baixo para cima em busca da coordenada vertical do centro evita a presença incômoda dos cílios nessa etapa de reconhecimento.

A contagem de cinco pontos mostrou-se eficiente para eliminar ruído proveniente da etapa de segmentação e encontrar a pupila com sucesso. Como o diâmetro da pupila é constante, um pequeno *offset* de 10 pixels foi suficiente para determinar as coordenadas dos centros. Esses parâmetros podem ser facilmente ajustados para imagens tomadas em resoluções alternativas.



Figura 9. Rastreo das pupilas

Considerações Finais

O presente trabalho apresentou um sistema para captura de movimento de pontos da face usando *hardware de baixo custo*. Foram mostrados os procedimentos para preparação do sistema de captura e como se dá o rastreo de pontos estratégicos na face e nos olhos do usuário.

É importante notar que os algoritmos utilizados são todos de baixo custo computacional, posto que executam em equipamentos com capacidade de computação limitada. Todos os algoritmos utilizados baseiam-se em estratégias simples, tais como verificar faixas de valores para componentes de cor e rotular de pixels em regiões pré-determinadas.

A configuração do sistema é feita utilizando uma aplicação que pode ser acessível diretamente no Raspberry Pi. Uma vez configurada, poderá guardar o arquivo de captura de movimento para uso posterior em algum processo de animação.

A documentação completa do sistema e a disponibilização dos códigos-fonte desenvolvidos estão sendo realizadas sob licença de código livre para a comunidade.

O sistema está sendo aperfeiçoado para que toda a configuração se dê de forma remota usando tecnologias de baixo custo computacional. Versões futuras do sistema também terão incorporados modelos computacionais usando ferramentas de animação de código livre como o Blender para facilitar a tarefa da equipe de animação.

O suporte para animação baseada em robôs animatrônicos também está em fase de testes com o sistema criado. Versões futuras dessa plataforma deverão contemplar também o processo de construção de um animatrônico.

Referências

- ADMTronics. Pros-aide adhesive. <http://www.pros-aide.com/original.html>. Accessed: 2016-09-28.
- et al., A. H. B. (2013). Facial performance enhancement using dynamic shape space analysis.
- Fatih Erol, U. G. (2007). An interactive facial animation system.
- Moeslund, T. B. and Granum, E. (2001). A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*, 81(3):231–268.
- Neth, D. and Martinez, A. (2009). Emotion perception in emotionless face images suggests a norm-based representation. *Journal of Vision*.
- Shiratori, T., Mahler, M., Trezevant, W., and Hodgins, J. K. (2013). Expressing animated performances through puppeteering.
- Sigal, L. and Koltun, V. (2012). (marker-based) motion capture. University Lecture.
- Vicon. What is motion capture. <https://www.vicon.com/what-is-motion-capture>. Accessed: 2016-10-15.