

Aquisição de Dados para um Sistema de Captura de Movimentos

Carolina A. S. Bigonha[†], David L. Flam (orientador)[†]

João Victor B. Gomide (orientador)[§], Arnaldo de A. Araújo (orientador)[†]

[†] Departamento de Ciência da Computação, [§]Escola de Belas Artes – Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627, Pampulha, Belo Horizonte, Minas Gerais, Brasil
{carolb, david, arnaldo}@dcc.ufmg.br, jvbgomide@gmail.com

Resumo

O presente trabalho aborda a localização e a identificação de marcadores em um software de aquisição de dados para um sistema de captura de movimentos. Utilizar-se-á a identificação, por meio de técnicas de detecção de quinas, de marcadores passivos que reflitam infravermelho. Uma vez realizada a captura de dados proposta e, ainda, a triangulação e o rastreamento dos pontos, será possível realizar a animação de personagens.

1. Introdução

1.1. Captura de Movimento

A captura e a análise do movimento humano são áreas de pesquisa relevantes devido ao vasto campo de aplicação e à complexidade inerente.

As potenciais aplicações dessa área são divididas em três conjuntos, segundo Moeslund *et al.* [5]: monitoramento – quando um ou mais alvos são rastreados e monitorados ao longo do tempo -, controle – correspondente à construção de interface para jogos, ambientes virtuais e, ainda, para fazer animação – e análise – que engloba a análise detalhada dos movimentos do ponto de vista clínico, permitindo realizar diagnósticos de enfermidades de pacientes ou melhorar a performance de atletas.

Pode-se dizer que o presente artigo aborda a segunda aplicação por ter como finalidade usar a captura de movimentos para fazer animação de personagens.

Os sistemas óticos de captura de movimento são divididos em, basicamente, dois subsistemas: o de detecção – hardware – e o de processamento – software.

A detecção pode ser ativa ou passiva. Na ativa, em um ambiente pré-definido e controlado, dispositivos são acoplados tanto ao alvo da detecção quanto aos arredores, que transmitem ou recebem sinais gerados, respectivamente.

Tais mecanismos de emissão e leitura de sinais facilitam a identificação e a localização dos pontos-chave em uma cena.

A detecção passiva, por sua vez, não requer vestimentas munidas de equipamentos com processamento interno: a captura pode ocorrer até mesmo a partir de seqüências de vídeo comuns. Os sensores passivos são baseados em fontes “naturais” de luz, e.g., o espectro visível ou outras ondas eletromagnéticas [3]. Portanto, quando se fala em marcadores passivos, refere-se a objetos ou marcações que, simplesmente, facilitam a segmentação e, logo, a identificação de pontos. Esses marcadores são diferentes dos citados no contexto ativo, já que os últimos respondem a estímulos externos.

Sistemas de captura do tipo passivo são preferidos por serem um boa alternativa em termos de custo, facilidade de uso e adaptação. Outro importante aspecto é que eles não são intrusivos, permitindo um comportamento humano natural perante câmeras.

Para este projeto, preferiu-se a detecção passiva feita por marcadores, da mesma categoria, que refletem infravermelho.

1.2. Os Softwares 3D

Em *softwares* para modelagem 3D, o personagem virtual é feito usando-se estruturas articuladas. Os modelos articulados, denominados *skeletons*, representam de maneira realista o movimento humano: são conjuntos de objetos rígidos, também conhecidos como *bones*, conectados por articulações – ou juntas -, que permitem o movimento.

Convém ressaltar que tais articulações, ou juntas, obedecem a uma hierarquia semelhante à de uma árvore. Dessa maneira, apenas a primeira articulação da estrutura precisa ser posicionada no espaço: as demais são posicionadas de acordo com os ângulos relativos entre os pontos definidos anteriormente.

A raiz da estrutura, conforme mostra a Fig. 1, é localizada no centro do personagem. Cada ponto relevante – e.g.

cotovelo, ombro, joelho – pertence a um nível da hierarquia e depende diretamente do ponto pertencente ao nível anterior.

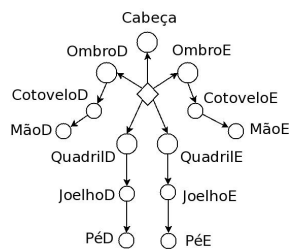


Figura 1. Hierarquia de um personagem.

Há uma série de formatos de arquivos que armazenam os dados extraídos de uma cena durante uma captura de movimento. No sistema proposto utilizar-se-á um arquivo com o padrão *Biovision Hierarchy* (BVH) que armazena os movimentos detectados, organizando os dados em duas grandes seções: hierarquia e movimento [4].

Na seção correspondente à hierarquia estabelece-se a relação entre os vértices, e, na correspondente aos movimentos, armazenam-se as transformações de cada ponto de acordo com o fluxo de tempo.

Escolheu-se esse formato dada sua simplicidade e compatibilidade com os mais conhecidos *softwares* de modelagem 3D, como o *3D Studio Max*, *Blender3D*, *Lightwave* e *Maya*.

2. Metodologia

A metodologia adotada pode ser dividida nas seguintes etapas: preparação das câmeras e do ambiente, localização e identificação dos marcadores, rastreamento dos pontos, reconstrução tridimensional, e, por último, manipulação dos dados a fim de se obter a saída esperada.

Cada uma dessas etapas será explicada adiante e está representada no diagrama da Fig. 2.

2.1. Decisões de Projeto

Para realizar a captura de dados, pretende-se trabalhar com câmeras digitais de vídeo preto e branco com resolução de 1280x1024 pixels. Ao trabalhar com mais de uma câmera, uma representação tridimensional pode ser obtida.

A fim de poupar o processamento de localização dos pontos, propôs-se que cada marcador passivo possuísse um desenho único, distinguível dos demais. Dessa maneira, tornar-se-ia fácil encontrar um mesmo ponto tanto em dois quadros quanto nos vídeos de duas câmeras distintas.

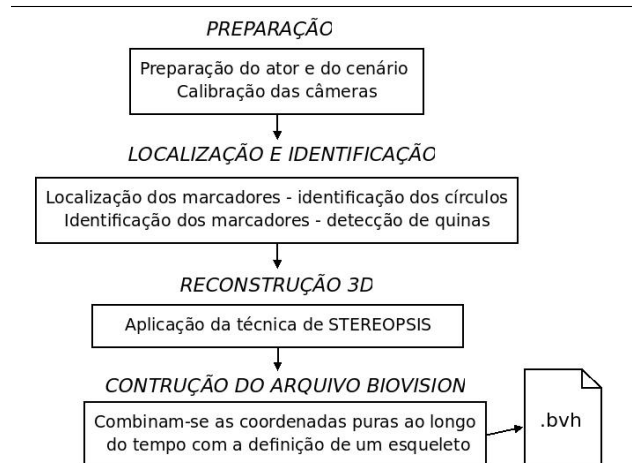


Figura 2. As etapas do sistema.

Aproveitando-se a alta resolução da câmera disponível para a captura, optou-se por um desenho que pudesse ser identificado por um procedimento de detecção de quinas. Essa técnica foi escolhida por ser relativamente simples e atender bem ao propósito.

Cada articulação possuirá um marcador com um certo número de quinas, permitindo-se, assim, a distinção entre cada ponto, conforme exemplifica a Fig. 3. Os marcadores passivos serão posicionados no corpo do ator de acordo com uma tabela, que indicaria qual marcador pertence a qual junta.

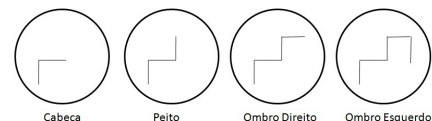


Figura 3. Identificação dos marcadores.

Para completar o material, têm-se disponíveis, ainda, emissores de luz na faixa do infravermelho, filtros para o mesmo comprimento de onda e marcadores passivos que refletem nessa mesma faixa. Tais marcadores, têm formato circular, possuem identificação e são posicionados estrategicamente no corpo do ator, em pontos-chave como cabeça, ombros e joelhos.

A implementação dos algoritmos propostos foi feita usando a biblioteca *Open Source Computer Vision Library*, ou OpenCV.

2.2. Preparação

Antes de realizar a aquisição de dados, há uma série de condições – que envolvem tanto o ambiente quanto o

ator – que devem ser satisfeitas para que se tenha uma interpretação correta da cena a ser capturada.

Primeiramente, o ambiente deve ter uma iluminação constante – no caso, infravermelha – e um fundo estático e uniforme. As câmeras devem permanecer paradas, posicionadas sobre tripés. O alvo, ou ator, deve usar uma vestimenta adequada e deve ter os marcadores descritos anteriormente, posicionados em certos pontos de seu corpo.

Satisfeitas tais condições, deve-se ajustar o sistema de acordo com as características particulares do cenário e calibrar as câmeras.

A chamada 'inicialização do modelo' é mais uma etapa crucial na preparação do sistema [5]. A maioria dos sistemas existentes adota a pose inicial do objeto como uma pose específica e pré-definida. Outras alternativas podem ser o operador do sistema especificar a pose inicial como parâmetro ou verificar o posicionamento do alvo em todas ou alguns quadros. Nesse último caso, o sistema simplesmente faz a relação entre eles.

O presente trabalho procura reduzir ou até eliminar essa fase de inicialização, já que os marcadores serão identificáveis a qualquer momento. O objetivo é que apenas ajustes opcionais precisem ser feitos pelo operador, deixando, assim, o programa mais robusto.

2.3. Localização e Identificação dos Marcadores

A reflexão do infravermelho pelos marcadores é captada pela câmera na forma de um borrão branco. Em vista disso, a realização de um pré-processamento agiliza a detecção significativamente.

Aplica-se, em cada quadro, uma limiarização binária – fazendo com que todos os pixels com valor menor que o limiar sejam transformados em pretos, e os pixels com valor maior ou igual ao limiar sejam transformados em brancos. Em seguida, realiza-se uma suavização de Gauss para que pequenos detalhes e distorções sejam removidos e para que se obtenham círculos mais bem delimitados. Por fim, aplica-se um algoritmo para detecção de bordas utilizando-se a técnica de Canny, que, embora tenha maior complexidade, produz resultados melhores [2].

Nesse ponto do processamento, tem-se uma imagem que contém somente o contorno dos círculos que representam os marcadores. Aplica-se, portanto, o algoritmo de detecção de círculos proposto em [1]. Em poucas palavras, tal algoritmo constrói um vetor com todos os pixels da borda e avalia, para cada conjunto de quatro pixels, se eles formam um círculo. Caso afirmativo, verifica-se se o círculo obedece aos parâmetros estabelecidos, como, por exemplo, diâmetro máximo. A saída desse algoritmo são as coordenadas do centro dos marcadores e seus respectivos raios.

Uma vez localizados os marcadores, passa-se para a segunda fase, que consiste na identificação dos mesmos.

Realiza-se a busca por quinas somente nas regiões que correspondem aos pontos encontrados. Pode-se, assim, identificar cada marcador.

Após o término dessa etapa, obtém-se as coordenadas do centróide dos marcadores nos planos que representam a imagem de cada câmera. Assim, bastará fazer a relação de posição dos marcadores entre os quadros do vídeo. Técnicas avançadas de *tracking* são dispensadas pelo fato de sempre se conhecer as coordenadas de cada um dos marcadores, mas serão utilizadas para predição do movimento em casos que ocorram oclusão de pontos.

2.4. Reconstrução 3D

Unir as informações gravadas por mais de uma câmera, explorando as disparidades entre elas, permite que se obtenha um senso de profundidade [6]. Para transformar os pontos independentes de cada câmera do sistema em coordenadas tridimensionais, a chamada *stereopsis* realiza dois processos: a fusão das características observadas de duas ou mais câmeras e a reconstrução da imagem tridimensional.

O uso dessa técnica permitirá que se encontrem as coordenadas x , y e z de cada marcador.

2.5. Preparação dos Dados para Animação

Os dados puros, extraídos da cena, até então, são chamados de dados de translação global [4]. Eles representam a posição de cada marcador em relação à origem global, sem incluir nenhuma hierarquia ou definição de esqueleto.

Embora trabalhar com os próprios arquivos translacionais globais seja possível e, ainda, ofereça uma maior liberdade quanto a complexidade estrutural dos personagens, para o projeto proposto, é interessante transformar as coordenadas cartesianas coletadas em um formato hierárquico. Além do fato de que trabalhar-se-á com personagens humanóides, as formas de armazenamento que mantêm uma relação dos pontos com um esqueleto deixam a configuração da estrutura do personagem mais independente do modelador 3D e mais próximo do *software* que gera os dados.

Uma vez realizada a tradução das coordenadas cartesianas puras para o formato *.bvh*, descrito anteriormente, basta importar o arquivo em um *software* 3D e aplicá-lo a um personagem.

3. Resultados Parciais

Até o presente momento, desenvolveu-se a parte de localização e identificação dos marcadores.

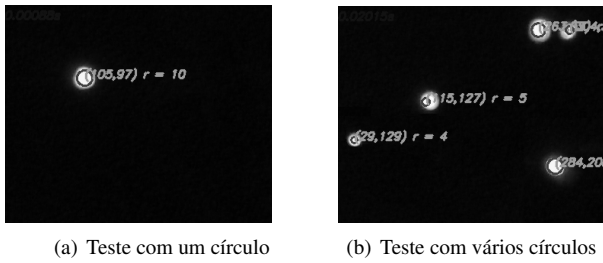


Figura 4. Localização dos marcadores.

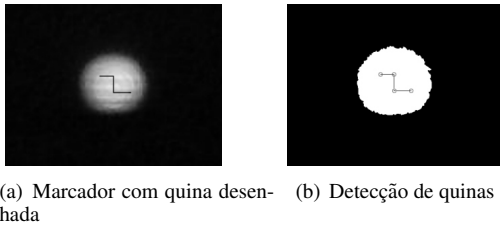


Figura 5. Identificação dos marcadores.

Os testes apresentados a seguir, foram feitos via simulação da recepção de infravermelho por uma câmera por meio de sinal de controle remoto.

As Figuras 4(a) e 4(b) mostram o centro e o raio dos marcadores fictícios – os borrões brancos da imagem – encontrados como resultado da detecção de círculos aplicada.

As imagens da Fig. 5 ilustram o desempenho da detecção de quinas. Em 5(a) tem-se uma imagem sintética que simula o desenho de um dos marcadores. Em 5(b) tem-se a saída do algoritmo empregado.

Por último, pode-se observar, por meio dos gráficos a seguir, o desempenho dos algoritmos de detecção de círculos – Fig. 6 – e de quinas – Fig. 7 – em relação ao tempo médio de processamento gasto. Para cada um dos algoritmos, variou-se os seguintes parâmetros: no primeiro, o número de pontos e, no segundo, o número de quinas.

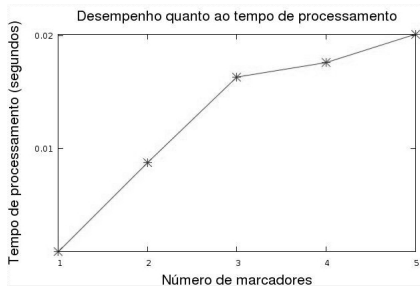


Figura 6. Tempo x Número de marcadores.

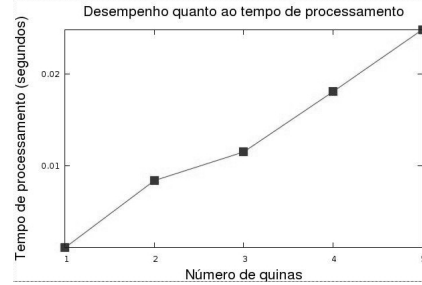


Figura 7. Tempo x Número de quinas.

4. Conclusão e Trabalhos Futuros

Inicialmente, o objetivo é implementar todo o *software* proposto e testá-lo para a animação de personagens. Uma vez alcançado esse objetivo pretende-se comparar o sistema desenvolvido com outros disponíveis no mercado. A comparação em termos de velocidade, qualidade de resultado e custo será de suma importância para a melhoria da qualidade do sistema.

Acredita-se que a utilização de marcadores passivos identificáveis será de suma importância para a construção de um sistema de captura rápido e de baixo custo.

Pode-se dizer que o aumento da eficiência se relaciona diretamente com a diminuição do processamento gasto com o rastreamento dos pontos, que é uma das etapas mais custosas do processo de captura. Se cada articulação é identificável em qualquer quadro, não há necessidade de se fazer cálculos complicados a cerca do movimento dos marcadores.

Os autores agradecem ao CNPq e à FAPEMIG pelo suporte financeiro deste trabalho.

Referências

- [1] T. C. Chen and K. L. Chung. An efficient randomized algorithm for detecting circles. *Computer Vision and Image Understanding*, (83):172–191, April 2001.
- [2] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Pearson Prentice Hall, 3rd edition, 2008.
- [3] Kakadiaris and C. Barrón. *The Handbook of Mathematical Models in Computer Vision*. Springer, 2008.
- [4] A. Menache. *Understanding Motion Capture for Computer Animation and Video Games*. Morgan Kaufmann; 1st edition, 1999.
- [5] T. B. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, (104):90–126, 2006.
- [6] J. Ponce and D. A. Forsyth. *Computer Vision - A Modern Approach*. Prentice Hall, 2002.