

Análise de Movimento Humano por Visão Computacional: Uma Síntese

Raquel R. Pinho, João Manuel R. S. Tavares

INEGI – Instituto de Engenharia Mecânica e Gestão Industrial
LOME – Laboratório de Óptica e Mecânica Experimental
FEUP – Faculdade de Engenharia da Universidade do Porto
Rua Dr. Roberto Frias, s/n, 4200-465 Porto
{[rpinho](mailto:rpinho@fe.up.pt), [tavares](mailto:tavares@fe.up.pt)}@fe.up.pt

Miguel F. P. V. Correia

INEB – Instituto de Engenharia Biomédica
LSI – Laboratório de Sinal e Imagem
FEUP – Faculdade de Engenharia da Universidade do Porto
mcorreia@fe.up.pt

RESUMO

O movimento humano é complexo, não linear e varia com o tempo. Nos últimos tempos, inúmeros investigadores têm-se dedicado ao desenvolvimento de sistemas automáticos capazes de realizar o seguimento, a análise e o reconhecimento deste tipo de movimento, utilizando técnicas de Visão Computacional. Neste artigo, serão resumidamente enumeradas e descritas algumas das técnicas actualmente empregues neste domínio.

INTRODUÇÃO

A análise de movimento por Visão Computacional tem vindo a desenvolver-se ao longo dos últimos anos, nomeadamente no domínio da análise de movimento do corpo humano. Este trabalho tem vindo a ser motivado pela vantagem de melhorar a interacção homem/máquina em diversas aplicações, tais como: na análise do desempenho atlético, em circuitos de vigilância, em animações de realidade virtual, nos estudos e diagnósticos clínicos [1, 2]. O âmbito da maioria dos trabalhos realizados na análise de movimento humano é, normalmente, na área de seguimento de sujeitos e no reconhecimento, quer em termos de estimação de pose quer em reconhecimento do tipo de andar.

CAPTURA DE MOVIMENTO

A análise de movimento principia com a captura dos dados associados ao mesmo. Os sistemas utilizados para capturar o movimento humano, são geralmente compostos por etapas de detecção e processamento [2]. A detecção pode ser realizada de forma passiva (baseada em fontes “naturais” de sinal) ou activa (são colocados dispositivos nos sujeitos e nos espaços envolventes que emitem e recebem sinais). A detecção passiva pode ser obtida através da análise de luz visível ou de outros comprimentos de onda electromagnéticos, não sendo necessária a integração de dispositivos especiais (com a excepção da utilização de marcadores, que são colocados no sujeito de forma a facilitar a captura de movimento). A detecção activa, permite um processamento mais simples e é geralmente utilizada em ambientes “controlados”, no entanto, nos últimos tempos tem vindo a ser desafiada por sistemas automáticos de captura passiva que recorrem a Visão Computacional.

PROCESSAMENTO DE DADOS CAPTURADOS

No que concerne ao processamento dos dados associados ao movimento humano, para que um sistema computacional os possa considerar de forma adequada é, geralmente, necessário

inicializá-lo, realizando-se posteriormente o seguimento do movimento assim como o eventual reconhecimento.

Inicialização

A inicialização do procedimento de análise inclui as operações que asseguram que o sistema principia as suas tarefas com a interpretação correcta da cena inicial [1, 2]. Exemplos de procedimentos de inicialização são o estabelecimento de restrições, a escolha e ajuste de modelos para o indivíduo, a calibração da(s) câmara(s) envolvida(s), a adaptação das metodologias às características da cena, etc.

A inclusão de restrições tem por objectivo a simplificação do problema associado à análise do movimento, podendo ser considerada a distinção entre duas grandes classes de restrições: as baseadas no movimento e as baseadas na aparência [1, 2]. As restrições de movimento estão relacionadas com os movimentos dos indivíduos e/ou da(s) câmara(s) envolvida(s); por exemplo, se a pessoa permanece dentro do espaço que está a ser considerado na captura, se existe oclusão, se a câmara permanece estática ou com movimento constante, etc. Por outro lado, as restrições de aparência estão relacionadas com o ambiente e com os sujeitos; por exemplo, existência de iluminação constante, de fundos estáticos, o conhecimento prévio da posição/forma inicial dos indivíduos, etc.

Seguimento

Na etapa seguinte, é normalmente realizado o seguimento do movimento, o que implica que o indivíduo seja segmentado em cada imagem e que sejam estabelecidas correspondências entre características ao longo da sequência, tendo em consideração, por exemplo, a posição, a velocidade, a textura ou a cor.

Para proceder o seguimento pode ser considerada a modelação do corpo humano. A utilização de modelos do sujeito, pode ter a desvantagem da perda de generalidade do algoritmo implementado; contudo, poderá ter a vantagem de facilitar a correspondência de características entre quadros (imagens) consecutivos [3, 4].

Alguns sistemas utilizam um modelo genérico construído a partir da média de vários indivíduos [4]; outros sistemas, consideram o sujeito actual e geram um modelo a partir dos dados observados, podendo refiná-lo à medida que se vai obtendo mais informação acerca do sujeito em causa [5], ou então pode ser construído um modelo personalizado que adequa um modelo genérico preexistente aos dados actuais [6].

Os modelos utilizados no seguimento de movimento humano, podem considerar ou não a informação prévia sobre a forma do indivíduo em causa. Se apenas for considerada informação acerca da textura ou da cor do indivíduo, então são utilizados modelos de aparência. No entanto, caso sejam incluídas informações prévias acerca das formas, os modelos poderão ser cinemáticos, dinâmicos, esqueléticos, deformáveis ou volumétricos [1, 2] (figura 1).

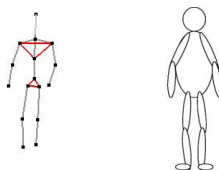


Figura 1 – Dois exemplos de modelos utilizados para a análise de movimento humano: modelo esquelético (esquerda) e modelo de contornos (direita). (Exemplo de [1].)

As abordagens baseadas em modelos de aparência, estabelecem a correspondência entre imagens por predição ou estimativa das características relacionadas com a posição, a velocidade, a textura ou a cor; recorrendo, por exemplo, a modelos Gaussianos, filtros de Kalman e de partículas, diferenciação temporal, agrupamentos de regiões ou contornos activos [1, 2]. Os modelos Gaussianos são utilizados na detecção e seguimento de movimento considerando um problema de propagação para a frente, em que a diferença entre imagens pode ser modelada por uma mistura de duas distribuições normais. Ainda que muito rápido, usualmente este modelo não é capaz de gerir os casos em que o fundo da cena tem textura significativa junto aos contornos dos objectos. Por sua vez, os filtros de Kalman e suas

variantes, assim como os filtros de partículas, têm sido utilizados no seguimento eficaz do corpo humano em condições controladas, podendo resolver problemas relacionados com o desaparecimento de alguns dos dados devido a situações de oclusão. Em [7] é exemplificado o seguimento pelo método de condensação (um filtro de partículas) que se revelou um estimador interessante e que suporta por exemplo distribuições multimodais.

Por outro lado, os modelos de diferenciação temporal têm vindo a ser desenvolvidos para superar a necessidade da utilização de um filtro de predição, como os filtros supracitados, para obter o seguimento de forma robusta. Por exemplo, as derivadas temporais e a correspondência entre contornos podem ser combinadas para a segmentação de regiões de objectos em movimento [1].

Quanto ao agrupamento de regiões pode ser considerado na segmentação da imagem e na anotação de regiões prováveis do indivíduo. Posteriormente, essas regiões podem ser seguidas ao longo do tempo através do seu centróide. Por exemplo, em [3, 8] a utilização de *voxels* revelou-se como sendo uma forma robusta de seguimento, uma vez que permite grandes deslocamentos entre imagens.

Nos casos em que a informação sobre a cor é considerada no seguimento, é geralmente assumido que a iluminação é homogénea e que existe contraste significativo entre o indivíduo e o fundo da cena. Esta última restrição também pode ser considerada quando se utilizam contornos para fazer o seguimento. O recurso a contornos activos é bastante comum em Visão Computacional, e esta técnica tem sido evidenciada em vários trabalhos como sendo um bom método de seguimento em tempo real [1-3].

Por outro lado, quando se consideram informações prévias acerca do indivíduo e/ou do seu movimento, a correspondência entre características é automaticamente estabelecida através do emparelhamento entre as imagens e os dados do modelo [3]. Nestes casos, comparando os modelos cinemáticos com os modelos dinâmicos, estes últimos têm merecido pouca atenção [1]. De facto, a maioria das metodologias de seguimento que utilizam modelos dinâmicos, são simples e genéricas ou altamente específicas mas controladas e afinadas de forma manual. Por sua vez, as abordagens biomecânicas têm sido criticadas pelas dificuldades inerentes à medição das dinâmicas de objectos complexos que envolvem um número elevado de massas, binários e forças; sendo assim difícil a redução da complexidade inerente ao modelo. Pelo contrário, os modelos cinemáticos revelaram ser adequados na detecção de movimento humano [3].

Na modelação do movimento humano também podem ser utilizados modelos esqueletizados (figura 1), que consistem num conjunto de segmentos de recta unidos por juntas, permitindo assim estimar o movimento e reconhecer o comportamento de figura inteira (corpo inteiro). A utilização destes modelos, é estável em diversas circunstâncias e o custo computacional envolvido permite o seguimento em tempo real.

A utilização de modelos deformáveis articulados é também uma técnica *standard* de análise de movimento humano que considera frequentemente o fluxo óptico [4]. Por exemplo, um modelo deformável baseado em regiões pode ser construído a partir de um modelo deformável baseado em contornos, e a sua utilização conjunta com o fluxo óptico poderá realizar satisfatoriamente o seguimento do movimento humano. A delimitação da região de interesse é normalmente inicializada por um algoritmo de segmentação baseada na detecção de movimento, e é seguida por um modelo deformável que explora os dados obtidos da textura da região. Este procedimento permite o seguimento, mesmo quando existem grandes deslocamentos ou fundos complexos, e é robusto à oclusão parcial [1, 2].

O uso de contornos para representar o corpo humano, está directamente relacionado com a sua projecção nas imagens. Por vezes, o contorno de uma pessoa pode ser suficiente para fazer o seu seguimento; mas só funcionará correctamente, caso as poses e as perspectivas envolvidas estejam suficientemente bem representadas na sequência considerada. Os contornos, sendo características de nível mais elevado do que os pontos, reduzem a possibilidade de falsos emparelhamentos existente nos modelos esqueletizados. A maioria das abordagens existentes requerem um modelo esqueletizado do corpo humano, mas também

poderá ser considerada a combinação destes modelos com modelos volumétricos ou com primitivas 3D [1, 2].

Por sua vez, os modelos volumétricos requerem um elevado número de parâmetros de modelação, sendo uma das técnicas mais comuns, a utilização de cilindros elípticos para a modelação 3D do corpo humano [1, 2].

Outras abordagens utilizam modelos hierárquicos, marcadores, e a combinação entre informações da cor e do movimento. Considere-se por exemplo, o modelo hierárquico treinado a partir de exemplos reais usando uma mistura de Gaussianas para traduzir a geometria e a cinemática, e modelos de Markov não observáveis para analisar a dinâmica envolvida [1]. Outra abordagem comum, consiste na utilização de filtros de Kalman ou de partículas com modelos de Markov não observáveis. Os modelos de Markov não observáveis são utilizados para determinar a forma da pessoa numa imagem, e os filtros de Kalman ou de partículas utilizam os dados obtidos pelos modelos de Markov não observáveis para seguir o movimento, ao estimar uma caixa que limita a sua trajectória e prevendo a localização da mesma ao longo da sequência. Usualmente, a conjugação de técnicas parece ser eficiente [2].

De um modo geral, quando o modelo tem menos parâmetros é mais fácil a correspondência entre as características e o modelo, mas é mais difícil a extracção dessas características [3].

Reconhecimento

A fase seguinte de um sistema de seguimento e análise de movimento humano, poderá consistir na classificação/reconhecimento da acção associada ao movimento capturado [9].

O trabalho de reconhecimento de gestos, de actividades e de indivíduos é muito vasto (conforme exemplificado em [1], [2]): Pode ser utilizado na estimativa da postura tridimensional, por exemplo no sentido de compreender o comportamento do sujeito observado (ver por exemplo [10], figura 2), na análise do andar, por exemplo enquanto assinatura biométrica para identificação de pessoas ou para distinção de sexos, ou em medicina para a detecção e diagnóstico de anomalias da marcha.

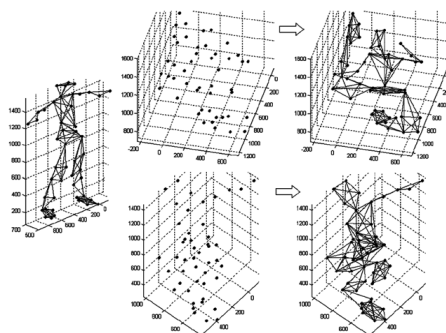


Figura 2 - Estimação de pose pelo método proposto em [10] utilizando pontos que compõe o modelo (esquerda), as poses não identificadas (ao centro) e as mesmas após identificação (direita). (Exemplo de [10].)

Existe, assim, uma diversidade de abordagens utilizadas como por exemplo as baseadas em técnicas estatísticas ou as baseadas em modelos [1, 2]. As estatísticas incluem a utilização de espaços próprios, de informação espacio-temporal, de séries de tempo e de silhuetas. Por outro lado, as abordagens baseadas em modelos, incluem o uso de modelos de Markov não observáveis, contornos activos, modelos esqueletizados, modelos de movimento recorrendo por exemplo ao emparelhamento de *templates* ou a marcadores.

Os diferentes movimentos humanos podem ser representados usando *templates* temporais, isto é, vectores de imagens estáticas, onde o valor do vector em cada ponto é função das propriedades do movimento nas localizações espaciais correspondentes ao longo de uma sequência de imagens. Contudo, esta abordagem tem a desvantagem de ser demasiado sensível à variação do comprimento da sequência. Esse problema é evitado se se definir cada postura estática como um estado de um espaço de estados, em que a sequência de

movimento é traduzida por uma sequência de estados e a transição entre estes é definida por probabilidades [1].

O reconhecimento pode também ser obtido através de marcadores colocados nas articulações, como por exemplo nos cotovelos ou nos ombros, o que poderá fornecer informação vital acerca do movimento em causa [9].

Os modelos articulados conseguem modelar de uma forma realista o movimento humano, e têm sido explorados para estimar a postura. O corpo pode ser modelado utilizando um enquadramento detalhado, e a informação extraída de cada imagem é interpretada por uma rede baseada na estrutura do modelo humano.

Os modelos deformáveis baseiam-se no estudo espacio-temporal da silhueta do sujeito, a partir de sequências de imagens geralmente adquiridas por várias câmaras em simultâneo. Estes métodos recorrem usualmente a contornos ocultos e dispensam a utilização de marcadores ou outros dispositivos, simplificando as dificuldades inerentes à existência de oclusões [2].

Consoante a aplicação, também podem ser utilizadas algumas ferramentas estatísticas que permitem a caracterização o movimento. As abordagens estatísticas incluem a utilização de espaços próprios, de informação espacio-temporal, de séries de tempo e de silhuetas. As abordagens baseadas em modelos, incluem o uso de modelos de Markov não observáveis, contornos activos, modelos esqueléticos, modelos de movimento.

O reconhecimento tem evoluído consideravelmente, podendo não ser necessária a recuperação de propriedades tridimensionais das pessoas em causa, ou o seguimento bidimensional dos seus membros para que o reconhecimento seja conseguido com êxito [2].

REFERÊNCIAS

- [1] Wang, J., Singh, S., 2003. Video Analysis of Human Dynamics - A Survey. *Real-time Imaging Journal*, vol. 9, pp. 320/345.
- [2] Pinho, R., Correia, M., Tavares, J., 2004. *Introdução à Análise de Movimento usando Visão Computacional*. FEUP/LOME.
- [3] Aggarwal, J., Cai, Q., 1999. Human Motion Analysis: A Review. *Computer Vision and Image Understanding*, vol. 73, pp. 428/440.
- [4] Baumberg, A., 1995. *Tese de Doutoramento: Learning Deformable Models for Tracking Human Motion*. School of Computer Studies. UK, The University of Leeds.
- [5] Wren, C., Azarbayejani, A., Darrell, T., Pentland, A., 1997. Pfinder: Real-Time Tracking of the Human Body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19.
- [6] Zheng, J., Suezaki, S., 1998. A Model Based Approach in Extracting and Generating Human Motion. *International Conference in Pattern Recognition*.
- [7] Isard, M., Blake, A., 1998. Condensation - Conditional Density Propagation of Visual Tracking. *International Journal on Computer Vision*, pp. 5/28.
- [8] Mikic, I., Trivedi, M., Hunter, E., Cosman, P., 2003. Human Body Model Acquisition and Tracking using Voxel Data. *International Journal of Computer Vision*, vol. 53, pp. 199/223.
- [9] Moeslund, T., Granum, E., 2001. A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*, vol. 81, pp. 231/268.
- [10] Li, B., Meng, Q., Holstein, H., 2004. Articulated Pose Identification with Sparse Point Features. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 34.