

Utilizando Propagação de Crença Não Paramétrica para Rastreamento de Movimento com Mínimo Uso de Informação a Priori

Gisele M. Simas, Rodrigo A. de Bem, Silvia S. C. Botelho
Centro de Ciência Computacionais - Programa de Pós Graduação em Modelagem Computacional
Universidade Federal do Rio Grande - FURG - Rio Grande - Brasil

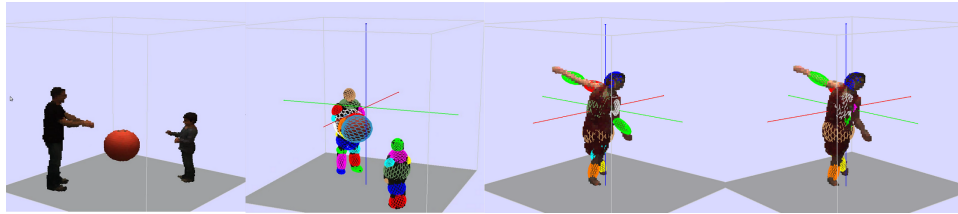


Fig. 1. Exemplos de resultados - esquerda para a direita: reconstrução volumétrica; inicialização automática; estimativa de pose empregando a solução proposta; estimativa de pose sem analisar a evidência global.

Abstract—Most existing tracking motion methods works in specific situations and requires the use of large amount of a priori information about: i) the object representation (appearance, kinematic structure); ii) possible moves; iii) physically valid poses. Thus, this work aims to investigate generic methods and reduce the need to use large amount of a priori information. This paper proposes a methodology capable of tracking a priori unknown objects and simultaneously build a representation model of these objects. This work investigates the use of the Nonparametric Belief Propagation technique, of the PAMPAS algorithm and of the Loose-limbed model. New algorithms are proposed to: i) decompose volume data into geometric shapes, ii) combine the information from bottom-up process of pose estimation, iii) use top-down strategy, in conjunction with bottom-up process, in order to resolve ambiguities.

Keywords-motion tracking; nonparametric belief propagation; dynamic programming.

Abstract—A maioria dos métodos de rastreamento de movimento existentes trabalha em situações específicas e requer a utilização de grande quantidade de informação a priori sobre: i) a representação dos objetos (aparência, estrutura cinemática); ii) movimentos possíveis de serem executados; iii) poses válidas fisicamente. Desta forma, este trabalho pretende contribuir no sentido de investigar métodos mais genéricos que diminuam a necessidade de utilização de grande quantidade de informação a priori. É proposta uma metodologia capaz de rastrear objetos desconhecidos a priori e, simultaneamente, construir um modelo de representação destes objetos. Investiga-se o emprego da técnica de Propagação de Crença Não Paramétrica, do algoritmo PAMPAS e do modelo Loose-Limbed. São propostos novos algoritmos para: i) decompor dados volumétricos em formas geométricas; ii) unir as informações do processo bottom-up de estimativa de pose; iii) utilizar estratégia top-down, em conjunto com o processo bottom-up, com intuito de resolver ambiguidades.

Keywords-rastreamento de movimento; propagação de crença não paramétrica; programação dinâmica.

I. INTRODUÇÃO

Este trabalho tem como objetivo elaborar um método de rastreamento visual sem marcas genérico, capaz de ser aplicado a diferentes tipos de objetos, sem a necessidade de utilização de grande quantidade de informação *a priori*. Dessa forma, pode se empregar o método proposto em situações diversificadas onde não se tem de antemão o conhecimento de quais objetos estão presentes na cena.

Para permitir tal tarefa, propõe-se que o modelo de representação dos objetos deva ser aprendido simultaneamente a tarefa de rastreamento de movimento. Como requisitos da metodologia citam-se: i) tratar as incertezas envolvidas; ii) permitir uma estimativa de pose flexível em relação ao modelo de representação dos objetos (pois este deve ser considerado impreciso, visto que deve ser aprendido simultaneamente ao rastreamento); e iii) permitir a modelagem de objetos de várias formas, incluindo dependências cíclicas entre as diferentes partes de um objeto.

Para possibilitar estas características, investiga-se a utilização da técnica de Propagação de Crença Não Paramétrica (NBP) [1], do algoritmo PAMPAS (*P*Article *M*essage *P*ASSing) [2] e do modelo Loose-Limbed [3]. Ressalta-se como diferencial da metodologia proposta o fato de possibilitar o emprego da técnica de Propagação de Crença Não Paramétrica (NBP) em uma aplicação na qual não se possuem padrões visuais fortes; e se utilize modelos de representação e movimento aprendidos simultaneamente ao rastreamento. Pelo conhecimento da autoria, até o momento, a NBP foi empregada apenas em conjunto de conhecimento a priori sobre: o modelo de representação do objeto; movimentos esperados; e modelagem de padrões visuais para cada membro do objeto (aprendidos a partir de treinamento prévio) [4].

II. TRABALHOS RELACIONADOS

Dadas as várias aplicações do rastreamento de movimento, tem-se desenvolvido uma série de trabalhos que, especificamente, no rastreamento de movimento humano, alcançam resultados bem realísticos [5]. No entanto, estes métodos apresentam algumas restrições delimitadas [5]. Com o intuito de diminuir tais restrições, nos últimos anos, vem sendo dedicada maior atenção à obtenção de métodos mais genéricos que possibilitem: análise de movimento em imagens monoculares [6]; utilização de câmeras em movimento e não sincronizadas [7]; dispensa de inicializações manuais [8]; adaptação a diferentes formas de um mesmo objeto [9]; tratamento de modificações online [10]; análise do movimento de objetos variados [11]; diminuição da necessidade de informação *a priori* [12].

A maioria dos trabalhos existentes em rastreamento emprega modelos de representação pré-definidos ou adaptáveis a diferentes formas de um mesmo tipo de objeto [9]. Normalmente, associa-se ao modelo de aparência de um objeto o seu modelo cinemático, que descreve as características dos movimentos e poses possíveis [13]. Alguns poucos trabalhos são dedicados à construção automática e não supervisionada de Modelos de Representação. Os trabalhos existentes constroem modelos através do estabelecimento de correspondências em uma sequência de imagens dos objetos em diferentes poses, empregando: padrões pontuais [14] e marcas ópticas [15]; correspondência entre vértices de uma malha de triângulos conhecida [16]; ou utilizando clusterização de elementos em partes rígidas e empregando heurísticas para se fazer a correspondência das partes no decorrer do tempo [17]. Em alguns trabalhos, pode-se interpretar que o modelo aprendido possui certa coerência temporal, obtida pelas restrições definidas no estabelecimento das correspondências [17]. Poucos trabalhos realizam alguma estimativa de movimento junto com a estimativa da representação dos objetos [14].

III. SOLUÇÃO PROPOSTA

A Fig. 2 apresenta a arquitetura proposta para a realização do rastreamento e atualização simultânea do modelo de representação dos objetos. A arquitetura é composta por três módulos principais responsáveis pelo Modelo de Observação, Modelo de Representação e Rastreamento.

Primeiramente, no módulo de Observação (verde), o sistema constrói um modelo de fundo da cena (1). Após, o sistema passa a receber imagens com os objetos de interesse em movimento e calcula a reconstrução volumétrica do primeiro instante de tempo (2), que é passada para o módulo de Representação (rosa). No módulo de Representação, analisa-se a reconstrução volumétrica calculada, para obter uma inicialização automática do modelo de representação dos objetos (3). O modelo de representação obtido é passado para o módulo de Rastreamento (azul) (4). O módulo de Rastreamento recebe, também, a reconstrução volumétrica de um próximo instante de tempo $t + 1$. Assim, tendo o modelo de representação dos objetos calculado com a informação até o tempo t e a observação do tempo $t + 1$, o módulo de Rastreamento estima a pose dos objetos no tempo $t + 1$. A

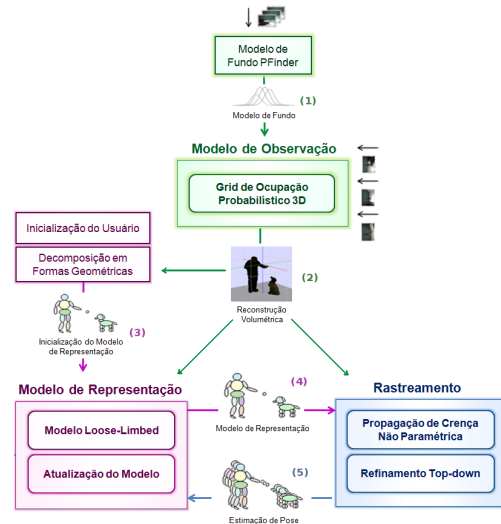


Fig. 2. Arquitetura Proposta.

pose estimada é enviada para o módulo de Representação (5). O módulo de Representação analisa a compatibilidade entre a observação e a pose estimada para adaptar (melhorar) o modelo de representação dos objetos. As etapas de obtenção da reconstrução volumétrica (3), estimativa de pose (5) e adaptação do modelo de representação (4) são, então, repetidas para todos os instantes de tempo.

IV. MODELO DE OBSERVAÇÃO

O módulo de observação utiliza uma reconstrução volumétrica, baseada na técnica de Grid de Ocupação Probabilístico 3D [18]. Nesta técnica, primeiramente, realiza-se a modelagem do fundo da cena; neste trabalho, empregou-se o modelo de fundo do sistema PFinder [19]. Após, passa-se a analisar as imagens das múltiplas câmeras com o objeto de interesse em cena e se obtém a reconstrução volumétrica 3D, para cada instante de tempo: o espaço da cena é discretizado em elementos de volume (voxels); para cada voxel, é associada uma probabilidade de este estar ocupado por algum objeto, através de inferência Bayesiana [18].

V. MODELO DE REPRESENTAÇÃO

Propõe-se o emprego de um modelo de representação constituído de um modelo de aparência e um modelo cinemático. i) *Modelo de aparência* - representa as dimensões e formas dos membros. Nesse trabalho, adota-se um conjunto de elipsoides por simplicidade dos cálculos. ii) *Modelo cinemático* - para possibilitar que o rastreamento seja feito com um modelo de representação impreciso, sugere-se a adoção do modelo *Loose-Limbed* [3]. O objeto é representado como um modelo gráfico probabilístico, no qual os nodos modelam os parâmetros de pose (posição e orientação) a serem estimados (Fig. 3). [3] comparam este modelo com um brinquedo "toy push puppet" com juntas elásticas: uma parte do objeto puxa a parte adjacente, mas ambas não precisam estar exatamente

coladas. Assim, tem-se certa flexibilidade: mesmo que duas partes sejam erroneamente consideradas dependentes, em um primeiro *frame*; o sistema poderá encontrá-las mais distantes, em um *frame* seguinte; e, então, poderá perceber que tais membros, na realidade, não estão ligados fisicamente, corrigindo o modelo de representação. O modelo *Loose-Limbed*, também, permite a utilização da técnica de Propagação de Crença Não Paramétrica (NBP). A NBP permite ciclos nas relações de dependências entre os membros do objeto, diferentemente das técnicas comumente utilizadas nos métodos de rastreamento. Tal fato é importante, especialmente, no caso deste trabalho, em que não se conhece de antemão a posição verdadeira das juntas e, portanto, o sistema deve tratar situações que se visualizem ciclos (exemplo, uma pessoa com o braço na cintura).

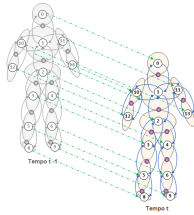


Fig. 3. Modelo Gráfico Probabilístico - as relações temporais estão apresentadas em verde e as espaciais em azul.

a) *Juntas*: são definidas a cada dois elipsoides E_1 e E_2 que aparecem conectados - a posição J da junta é obtida pelo centroide dos voxels de E_1 que possuem como vizinhos um voxel de E_2 e dos voxels de E_2 vizinhos de voxels de E_1 .

b) *Atualização do Modelo de Representação*: ao final da estimação de pose de cada novo instante de tempo, deve-se atualizar o modelo de representação utilizando a nova observação. Propõe-se o seguinte esquema: i) a cada novo instante de tempo, estima-se a pose dos objetos, empregando o modelo de representação do tempo anterior; ii) então, excluem-se os elipsoides cuja porcentagem do volume de voxels ocupados no interior da forma for abaixo de certo limiar; iii) após, realiza-se o método de Decomposição Geométrica do Volume (descrito a seguir), considerando apenas os voxels ocupados não atribuídos a qualquer elipsoide. Assim, ter-se-á um novo modelo de representação que deve ser empregado para estimar a pose do objeto do próximo instante de tempo.

A. Decomposição Geométrica do Volume

Este trabalho propõe um método para dividir o conjunto de voxels ocupados por objetos em formas geométricas, de modo que, no interior de cada forma, sobre uma quantidade mínima de espaço vazio. Este método é empregado como uma inicialização automática e, também, para adaptar o modelo de representação, a cada novo instante de tempo. Primeiramente, determina-se uma componente conexa (grupos de voxels conectados) através de uma busca em largura; calcula-se a moda da posição dos voxels desta componente; e, então, expande um elipsoide a partir do voxel mais próximo à

moda de forma que a porcentagem de voxels livres (sem estar ocupado por objetos), no interior do elipsoide, seja pequena. Se o elipsoide formado possuir um número mínimo de voxels ocupados, considera-se este como uma parte válida do objeto; caso contrário, este é desconsiderado. O processo é repetido até que todos os voxels tenham sido analisados.

Para se expandir um novo elipsoide, define-se que, primeiramente, apenas o voxel ocupado mais próximo da moda pertença ao elipsoide. Então, executa-se uma Busca em Largura, a partir deste voxel, acrescentando um nível da busca por vez. A cada novo nível, calcula-se o elipsoide que engloba os voxels ocupados adicionados ao elipsoide, a partir do método 'Atualiza parâmetros do elipsoide'. Após, insere-se todos os voxels ocupados que estejam no interior da forma obtida e que ainda não tenham sido designados a qualquer elipsoide como pertencentes a esta forma. Então, determina-se o novo voxel central do elipsoide. Repete-se por um número pré-definido de interações: a 'Atualização dos parâmetros do elipsoide'; adição de voxels ocupados internos à forma; e determinação do novo voxel central. Este *loop* é empregado para migrar o centro do elipsoide para a posição mais favorável ao seu crescimento (assim, a forma se desloca em direção ao volume do objeto e pode englobar um número maior de voxels). Então, calcula-se a porcentagem de voxels vazios no interior do mesmo. Considera-se que o elipsoide é válido se esta porcentagem for menor que certo limiar: caso a forma seja considerada válida, repete-se o processo inserindo novos voxels ao elipsoide; caso contrário, deve-se adotar a forma e o conjunto de voxels obtidos num passo anterior.

a) *Atualiza Parâmetros do Elipsoide*: este método recebe um conjunto de voxels V_E ocupados pertencentes a um elipsoide E e calcula a forma de E . Para isto, primeiramente, se calcula a média de posição dos voxels de V_E , este valor corresponderá ao centro do elipsoide; então, se obtém a covariância de posição destes voxels. Após, emprega-se a Decomposição em Valores Singulares da matriz de covariância para obter os principais eixos do elipsoide (\vec{a} , \vec{b} e \vec{c}): os auto-vetores da matriz de covariância fornecem as direções destes eixos; os auto-valores fornecem os módulos [20]. Após, deve-se 'Verificar o tamanho mínimo do elipsoide' (descrito a seguir) obtido, para corrigir formas nas quais a covariância tenha uma mínima variação no sentido de um dos principais eixos. Ao final, realiza-se algumas operações adicionais: computa-se a matriz de rotação necessária para rotacionar o elipsoide de forma que os seus principais eixos fiquem paralelos aos eixos coordenados. Então, aplica-se a mesma rotação aos vetores de posição \vec{P}_v de cada um dos voxels v internos ao elipsoide E , obtendo os vetores \vec{P}_{vr} ; e computa-se as constantes a , b e c : $a = (\text{maximo}(P_{vr_x}) - \text{minimo}(P_{vr_x}))/2$, $b = (\text{maximo}(P_{vr_y}) - \text{minimo}(P_{vr_y}))/2$, $c = (\text{maximo}(P_{vr_z}) - \text{minimo}(P_{vr_z}))/2$, onde o valor *maximo* e *minimo* consideram todos os voxels. Finalizando, define-se o módulo dos novos principais eixos do elipsoide (\vec{a}_n , \vec{b}_n e \vec{c}_n): $\vec{a}_n = a * \vec{a} / |\vec{a}|$, $\vec{b}_n = b * \vec{b} / |\vec{b}|$, $\vec{c}_n = c * \vec{c} / |\vec{c}|$.

b) *Verifica tamanho mínimo do elipsoide*: este método atribui novos valores aos eixos considerados inapropriados. Dependendo do número de eixos que possuem módulo menor

do que um limiar, definem-se três casos: i) Três eixos possuem módulos menor do que o limiar - atribui-se um elipsoide padrão mínimo; ii) Dois eixos possuem módulos menor do que o limiar - conhece-se apenas a direção e o módulo de um dos eixos, denominando-o de \vec{A} , sabe-se que os outros dois eixos devem ser ortogonais a \vec{A} e entre si. Calcula-se um vetor \vec{ncol}_A qualquer não colinear a \vec{A} . Então, executa-se o produto vetorial de \vec{A} e \vec{ncol}_A , obtendo um vetor \vec{B} ortogonal a \vec{A} ($\vec{B} = \vec{A} \times \vec{ncol}_A$). Após, obtém-se o terceiro vetor $\vec{C} = \vec{A} \times \vec{B}$; iii) Um eixo possui módulo menor do que o limiar - denominando os eixos conhecidos de \vec{A} e \vec{B} , a direção do eixo desconhecida é dada por $\vec{C} = \vec{A} \times \vec{B}$.

VI. RASTREAMENTO

O rastreamento é realizado empregando-se a técnica NBP, o algoritmo PAMPAS e uma etapa de refinamento top-down. A seguir são especificados os passos executados, iterativamente, para se obter a estimação de pose de cada instante de tempo. 1) Inicializações: de parâmetros empregados nos cálculos das mensagens e dos modelos de movimento; define-se a crença de cada nodo do tempo t como sendo igual a crença deste nodo no tempo $t - 1$ mais o modelo de movimento. 2) Processo *bottom-up*: a) Troca de mensagens e avaliação da evidência individual - cada nodo calcula as mensagens a serem enviadas para os nodos dependentes. Para a obtenção das mensagens, avalia-se a observação de cada nodo individualmente. b) Produto das mensagens recebidas - cada nodo realiza o produto das mensagens que recebeu, de forma a obter uma estimativa de densidade que informe as poses mais prováveis deste nodo. 3) Processo *top-down*: a) Refinamento da crença e avaliação da evidência global da pose inteira do objeto - analisa-se a evidência global (pose inteira do objeto) e o produto das mensagens recebidas por cada nodo (calculado no processo *bottom-up*), para resolver problemas de sobreposição das crenças e refinar a pose de cada membro do objeto. b) Atualização de parâmetros empregados nos cálculos das mensagens - emprega-se as crenças atualizadas de todos os nodos, para recalcular os parâmetros utilizados no cálculo das mensagens a serem trocadas pelos diferentes nodos.

a) *Modelo de Movimento*: para cada membro do objeto, calcula-se um modelo de movimento que informa como a pose deste membro deve evoluir dos tempos $t - 1$ para t . Neste trabalho, o modelo de movimento consiste em uma mistura de gaussianas calculada com base nos k últimos instantes de tempo, utilizando os deslocamentos estimados entre pares de tempos consecutivos. b) *Evidência*: neste trabalho, adota-se, como evidência, apenas a ocupação dos voxels.

A. Mensagens

As mensagens consistem em uma estimativa de densidade (uma mistura de gaussianas) que cada nodo deve enviar ao seu adjacente, informando como 'pensa' que seu adjacente deve estar posicionado. Neste trabalho, adotam-se dois tipos de mensagens: espaciais e temporais (Fig. 3). Na formulação proposta por este trabalho, as mensagens temporais são calculadas em função da crença do nodo remetente e de um deslocamento

estimado. As mensagens espaciais são calculadas em função da crença do remetente, da crença do destinatário e de uma estimativa de um vetor de 'flexibilidade' da posição de encontro dos elipsoides envolvidos. Especificam-se os seguintes passos para a obtenção da mensagem que um nodo n_1 deve enviar para o seu adjacente n_2 : i) Obter um conjunto de amostras das poses prováveis de n_2 (poses de como n_1 'pensa' que n_2 está posicionado); ii) Avaliar a evidência, calculando um peso de compatibilidade de cada amostra em relação à observação; iii) Empregar algum método para se obter uma estimativa de densidade que represente o conjunto de amostras que atingiram certo limiar de evidência. Os cálculos empregam quatérnios.

a) *Mensagens Temporais*: para o cálculo das mensagens temporais, define-se uma estimativa de densidade de *Deslocamento* para cada membro do objeto. Esta estimativa é atualizada a cada iteração do *loop* da NBP. Cada amostra da mensagem temporal que o nodo N_{t-1} envia para o nodo N_t é obtida da seguinte forma: calcula-se uma amostra de pose da crença de N_{t-1} ; calcula-se uma amostra da estimativa de deslocamento de N_t ; calcula-se uma pose possível para N_t dada por: $\vec{Posicao}_{N_t} = \vec{Posicao}_{Crenca_{N_{t-1}}} + \vec{Posicao}_{Deslocamento_{N_t}}$, $\vec{Orientacao}_{N_t} = \vec{Orientacao}_{Crenca_{N_{t-1}}} * \vec{Rotacao}_{Deslocamento_{N_t}}$; avalia-se a evidência local de N_t na pose calculada. Após se ter obtido um conjunto de amostras, utiliza-se um método para calcular a estimativa de densidade da mensagem. Após cada nova iteração do *loop* da NBP, atualiza-se a estimativa do *Deslocamento* de cada nodo. Para isto, cria-se uma nova estimativa de densidade, para representar o deslocamento observado (transformação entre a pose do nodo no tempo $t - 1$ e t). O novo *Deslocamento* é obtido pelo produto das estimativas de densidade do deslocamento observado e deslocamento antigo.

b) *Mensagens Espaciais*: entre dois membros fisicamente conectados, define-se uma junta J e dois vetores $\vec{v}_{e_{1t}}$ e $\vec{v}_{e_{2t}}$ que ligam a posição da junta J aos centroides dos elipsoides associados a esta junta. Adicionalmente a tais vetores, estima-se um terceiro vetor, denominado de *vetor_flexivel*, que permite um certo deslocamento da posição original de encontro dos elipsoides envolvidos. A possibilidade de se aceitar certo deslocamento na posição de encontro das formas favorece a realização da estimação de pose quando o modelo de representação dos objetos é impreciso. Esta imprecisão pode ser devido ao fato de que: i) os modelos de representação devem ser aprendidos no decorrer do tempo e não devem ser informados *a priori*; ii) ou, ainda, elipsoides podem não conseguir representar adequadamente o corpo de alguns tipos de objetos. Cada amostra da mensagem espacial que o nodo e_{1t} envia para o nodo e_{2t} é obtida da seguinte forma: calcula-se uma amostra de pose da crença de e_{1t} ; calcula-se uma amostra de pose da crença de e_{2t} ; calcula-se uma amostra da estimativa de densidade representativa do *vetor_flexivel*; calcula-se uma pose possível para e_{2t} dada por: $\vec{Posicao}_{e_{2t}} = \vec{Posicao}_{Crenca_{e_{1t}}} + \vec{v}_{e_{1t}} \cdot \vec{Orientacao}_{Crenca_{e_{1t}}} + \vec{v}_{flexivel}_{e_{1t}, e_{2t}} - \vec{v}_{e_{2t}} \cdot \vec{Orientacao}_{Crenca_{e_{1t}}}$, $\vec{Orientacao}_{e_{2t}} = \vec{Orientacao}_{Crenca_{e_{2t}}}$; avalia-se a evidência local do nodo e_{2t} na pose calculada. Após se ter obtido um conjunto de amostras, utiliza-se um método

para calcular a estimativa de densidade da mensagem. Após cada nova iteração do *loop* da NBP, atualiza-se a estimativa do *vetor_flexível* de cada junta J que associa dois nodos e_{1t} e e_{2t} . Para tal tarefa, cria-se uma nova estimativa de densidade, para representar o vetor flexível observado (ou seja, o deslocamento do ponto de encontro dos elipsoides calculado a partir das crenças de e_{1t} e e_{2t}). O novo vetor flexível é obtido pelo produto das estimativas de densidade do vetor flexível observado pelo vetor flexível antigo.

c) *Produto das Mensagens*: após a troca de mensagens, cada nodo deve realizar a fusão das mensagens recebidas. Na formulação original da NBP, a fusão das mensagens é obtida por um produto aproximado das misturas de gaussianas de forma a obter uma nova mistura com um número pré-definido de componentes [1]. O cálculo deste produto é baseado no método de Amostragem de Gibbs. No entanto, a Amostragem de Gibbs tem uma desvantagem: amostras aleatórias raramente caem em torno das modas da distribuição a ser estimada, o método pode se perder explorando áreas sem interesse [21]. Além disso, a Amostragem de Gibbs falha em duas situações quando: i) existem modas de alta massa de probabilidade com um caminho de probabilidade nula entre tais modas; ii) todos estados possuem probabilidade não nula e existe uma única região de grande massa de probabilidade. Adicionalmente, o método empregado por [1] tem um alto custo computacional, pois é necessário um número razoável de iterações.

Analisando tais aspectos, este trabalho propõe um algoritmo, baseado em programação dinâmica, para se obter as componentes mais adequadas do produto resultante das mensagens. O produto de N_M misturas de gaussianas, tendo cada uma N_C componentes, resulta em uma nova mistura com $(N_C^{N_M})$ componentes. Desta forma, o número de componentes da mistura resultante cresce exponencialmente com o número de mensagens, tornando os cálculos complexos. O algoritmo proposto seleciona as N_C melhores componentes do produto de duas misturas. Após, combina os resultados, para obter as N_C melhores componentes do produto de cada três misturas. Segue-se o mesmo raciocínio até que se obtenham as componentes mais adequadas do produto de todas as mensagens recebidas. Definem-se, como as componentes mais adequadas, aquelas que são resultantes do produto de componentes com menor distância entre suas médias. Tal critério é estabelecido pelo seguinte raciocínio: se um nodo e_i recebe mensagens dos nodos e_j e e_k , as poses mais prováveis para este nodo e_i são aquelas que estão presentes em ambas as mensagens de e_j e e_k ; desta forma, o produto das mensagens que e_i recebe deve ser resultante das componentes de e_j e e_k que possuem menor distância entre suas médias. Observa-se que a ordem no qual se selecionam as melhores componentes do produto das misturas influencia nos resultados. Então, denominando-se $S_{i\dots j}$ como as melhores componentes resultantes do produto das misturas de i a j , define-se a seguinte relação de recorrência:

$$S_{i,j\dots NroMisturas} = \begin{cases} S_{j\dots NroMisturas} \text{ pela Mistura}_i \\ S_{i,j\dots k-1,k+1\dots NroNodos} \text{ pela Mistura}_k \end{cases} \quad (1)$$

B. Crença

A partir do produto das mensagens, deve-se calcular a crença de cada nodo. A crença de um nodo pode ser definida como uma estimativa igual ao produto das mensagens. No entanto, a evidência individual dos nodos, empregada no cálculo das mensagens, pode não ser suficiente para a estimação correta da pose. Isto ocorre, principalmente, nos casos em que não se possuem padrões visuais fortes distinguíveis para cada membro do objeto. Nestes casos, as crenças resultantes dos diferentes nodos podem, até mesmo, serem calculadas como sobrepostas. No sentido de se diminuir problemas deste tipo, [1] comenta a possibilidade de se empregar um Amostrador de Gibbs, para refinar a pose estimada. No Amostrador de Gibbs, deve-se analisar a evidência global da pose inteira do objeto e, portanto, pode-se denominar esta etapa como um processo *top-down*. No entanto, a quantidade elevada de parâmetros envolvidos na pose inteira do objeto dificultam a convergência do método que requer uma grande quantidade de iterações. Isto, somado ao alto custo computacional de se computar a evidência global da pose inteira do objeto, tornam este método de refinamento da crença inviável.

Dessa forma, este trabalho propõe um método, para realizar o refinamento da estimativa de pose, baseado em programação dinâmica, semelhante ao método apresentado para o produto de mensagens. O método proposto pressupõe que, a cada dois nodos que possuem crenças sobrepostas, ao menos um dos dois consegue definir corretamente a sua pose como uma das poses de maior evidência local (evidência analisando apenas o nodo em questão). Tal suposição é válida para a maioria dos casos, conduzindo a uma solução aproximadamente ótima para o problema. Nota-se que o estado de um membro exclui o estado dos demais (os membros não devem estar sobrepostos). Assim, observa-se que a ordem no qual se definem as crenças dos nodos influencia nos resultados.

Para cada nodo, tenta-se ajustar iterativamente as médias das componentes do produto de mensagens deste nodo de forma a obter um conjunto S com um número pré-definido de poses de maior evidência local (analisando apenas o nodo em questão). Então, parte-se para selecionar as melhores poses de cada grupo de dois nodos. Definindo-se dois nodos A e B , com os conjuntos de melhores poses S_A e S_B , as poses mais adequadas olhando ambos os nodos A e B (AB) são: i) alguma pose contida em S_A e, após, iteragindo-se as médias das componentes do produto de mensagens de B de forma que B não se sobreponha à pose definida para A ; ii) ou alguma pose contida em S_B e, após, iteragindo-se as médias das componentes do produto de mensagens de A . De forma semelhante, combinam-se os resultados de cada dois nodos com um terceiro e segue-se este raciocínio até se obterem as poses mais adequadas para o corpo inteiro do objeto. Assim, denominando-se $S_{i\dots j}$ como as melhores poses considerando os nodos de i a j , o método proposto utiliza a seguinte relação de recorrência:

$$S_{i,j\dots NroNodos} = \begin{cases} S_{j\dots NroNodos} + \text{iterage}(\text{nodo}_i) \\ S_{i,j\dots k-1,k+1\dots NroNodos} + \text{iterage}(\text{nodo}_k) \end{cases} \quad (2)$$

VII. RESULTADOS E DISCUSSÕES

A Fig. 4 apresenta resultados do método do rastreamento simultaneamente à atualização do modelo de representação e empregando uma inicialização automática. A sequência adotada possui oito câmeras e encontra-se disponível em 'http://4drepository.inrialpes.fr/data-4d/dancer/'. Adotou-se como tamanho mínimo dos elipsoides considerados como válidos igual a 40 voxels. Empregou-se o limiar de porcentagem de voxels livres no interior de elipsoide válido como igual a 0.2. Foram avaliadas 100 amostras na composição de cada mensagem, empregou-se 10 iterações no *loop* da NBP e as misturas utilizadas possuem cada uma 5 componentes. Mais resultados estão disponíveis no endereço <https://sites.google.com/site/motiontrackingfurg/>.

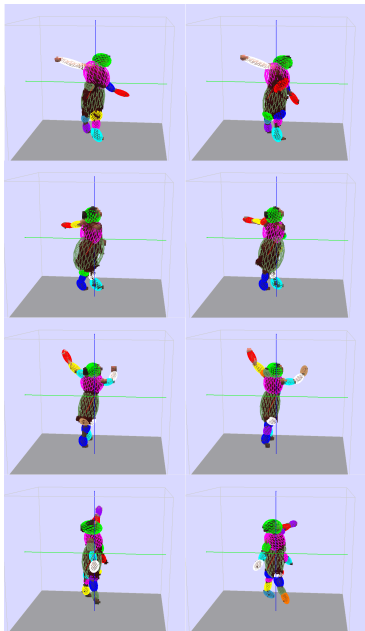


Fig. 4. Resultados do rastreamento com inicialização automática e atualização do modelo representação simultaneamente ao rastreamento.

VIII. CONCLUSÃO

Neste trabalho, investigaram-se técnicas que visem diminuir a necessidade do uso de informação a priori em sistemas de rastreamento. Focou-se, principalmente, na tarefa de fornecer uma estimação de pose flexível para se tratar com modelos de representação imprecisos (tais modelos devem ser aprendidos no decorrer do tempo); que não requeira pistas visuais fortes; e que utilize um modelo de movimento impreciso. Quanto a critérios de implementação, citam-se como contribuições: o desenvolvimento de um esquema de decomposição de volumes para a inicialização/adaptação do sistema; a elaboração de uma modelagem diferenciada para as relações de dependência entre as partes do objeto; a proposição de um algoritmo alternativo para a fusão das informações bottom-up e um semelhante para a etapa top-down. O emprego da técnica de NBP sem a utilização de treinamento prévio ou informações

sobre os objetos em cena foi possível: i) pela formulação das mensagens espaciais, no qual se estima um vetor flexível, possibilitando certa flexibilidade na estimação de pose; ii) pela formulação da mensagem temporal que realiza uma estimação do deslocamento, facilitando a estimação de pose com modelo de movimento impreciso; iii) pela adoção de uma etapa top-down que analisa a pose, considerando todos os membros em conjunto.

AGRADECIMENTOS

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq.

REFERENCES

- [1] E. B. Sudderth, E. T. Ihler, W. T. Freeman, and C. Science, "Nonparametric belief propagation," in *In CVPR*, 2003, pp. 605–612.
- [2] M. Isard, "Pampas: Real-valued graphical models for computer vision," *Computer Vision and Pattern Recognition*, pp. 613–620, 2003.
- [3] L. Sigal, M. Isard, B. H. Sigelman, and M. J. Black, "Attractive people: Assembling loose-limbed models using non-parametric belief propagation," in *in NIPS*. MIT Press, 2003, pp. 1539–1546.
- [4] E. B. Sudderth, A. T. Ihler, M. Isard, W. T. Freeman, and A. S. Willsky, "Nonparametric belief propagation," *Communications of the ACM*, vol. 53, no. 10, pp. 95–103, 2010.
- [5] L. Sigal and M. J. Black, "Guest editorial: State of the art in image- and video-based human pose and motion estimation," *IJCV*, 2010.
- [6] A. Fossati, M. Salzmann, and P. Fua, "Observable subspaces for 3d human motion recovery," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 0, pp. 1137–1144, 2009.
- [7] N. Hasler, B. Rosenhahn, T. Thormählen, M. Wand, J. Gall, and H. Peter Seidel, "Markerless motion capture with unsynchronized moving cameras," in *Computer Vision and Pattern Recognition*, 2009, pp. 224–231.
- [8] A. Sundaresan and R. Chellappa, "Multicamera tracking of articulated human motion using shape and motion cues," *Trans. Img. Proc.*, 2009.
- [9] I. Mikić, M. Trivedi, E. Hunter, and P. Cosman, "Human body model acquisition and tracking using voxel data," *IJCV*, 2003.
- [10] D. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," in *the International Journal of Computer Vision, Special Issue: Learning for Vision*, 2007.
- [11] N. Ukita, M. Hirai, and M. Kidode, "Complex volume and pose tracking with probabilistic dynamical models and visual hull constraints," in *ICCV'09*, 2009, pp. 1405–1412.
- [12] J. Gall, B. Rosenhahn, T. Brox, and H. Seidel, "Optimization and filtering for human motion capture - a multi-layer framework," *IJCV*, 2010.
- [13] C. Canton-Ferrer, J. R. Casas, and M. Pardàs, "Voxel based annealed particle filtering for markerless 3d articulated motion capture," *3DTV Conference*, pp. 1–4, 2009.
- [14] D. A. Ross, D. Tarlow, and R. S. Zemel, "Learning articulated structure and motion," *International Journal of Computer Vision*, 2010.
- [15] E. D. Aguiar, C. Theobalt, and H. Peter Seidel, "Automatic learning of articulated skeletons from 3d marker trajectories," in *In Proc. Intl. Symposium on Visual Computing (ISVC 2006)*, 2006, pp. 485–494.
- [16] E. de Aguiar, C. Theobalt, S. Thrun, and H.-P. Seidel, "Automatic conversion of mesh animations into skeleton-based animations," *Comput. Graph. Forum*, vol. 27, no. 2, pp. 389–397, 2008.
- [17] C. Theobalt, E. D. Aguiar, M. A. Magnor, H. Theisel, and H.-P. Seidel, "Marker-free kinematic skeleton estimation from sequences of volume data," *Proceedings of the ACM symposium on VRST*, 2004.
- [18] J. S. Franco and E. Boyer, "Fusion of multi-view silhouette cues using a space occupancy grid," *ICCV 05*, 2005.
- [19] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, 1997.
- [20] F. Banégas and M. Jaeger, "The ellipsoidal skeleton in medical applications," in *In Proceedings of the Sixth ACM Symposium on Solid Modeling and Applications*, pages 30–38, Ann Arbor. ACM Press, 2001.
- [21] Y.-K. Wang and K.-Y. Cheng, "A two-stage bayesian network method for 3d human pose estimation from monocular image sequences," *EURASIP Journal on Advances in Signal Processing*, 2010.