

## RECONSTRUÇÃO PROBABILÍSTICA 3D APLICADA À ANÁLISE DE MOVIMENTO: UM ENFOQUE NA TELE-OPERAÇÃO

GISELE M. SIMAS\*, GUILHERME P. FICKEL\*, LUCAS NOVELO\*,  
RODRIGO A. DE BEM\*, SILVIA S. C. BOTELHO\*

\**Universidade Federal do Rio Grande - FURG*  
*Av Itália Km 8 - Rio Grande - RS - Brasil*

Emails: gisele\_simas@yahoo.com.br, guilhermefickel@gmail.com, lucasnovelo@gmail.com,  
rodrigo.bem@gmail.com, silviacb@furg.br

**Abstract**— This paper proposes a method for motion analysis of objects without a defined shape, it is based on a 3D probabilistic volumetric reconstruction obtained from multiple cameras. This method can be used in teleoperation, analyzing the movements of the human operator to robot and avatar control. Moreover, the volumetric reconstruction can be used to rebuild the workspace of the robot to be operated, favoring the interaction with the workplace and the collision-free trajectory planning. First, a volumetric reconstruction of the objects of interest is obtained by the 3D Probabilistic Occupancy Grid technique, which was recently proposed for to be applied in monitoring of environments using multiple cameras and there are few studies that employ this technique in the movement analysis. Next, Optical Flow and a similar method to the Expectation-Maximization (EM) algorithm are used to identify and track the body parts of the objects of interest.

**Keywords**— 3D probabilistic volumetric reconstruction, optical flow, motion tracking, teleoperation.

**Resumo**— Este artigo propõe um método para análise de movimento de objetos sem forma pré-definida, baseado em uma reconstrução volumétrica probabilística 3D obtida a partir de múltiplas câmeras. Este método pode ser empregado na tele-operação, analisando os movimentos do operador humano para controle de robôs e avatares. Além disso, a reconstrução volumétrica pode ser utilizada para reconstruir o espaço de trabalho do robô a ser operado, favorecendo a interação com o ambiente e a elaboração de trajetórias livres de colisão. Primeiramente, se obtém uma reconstrução volumétrica dos objetos de interesse pela técnica de Grid de Ocupação Probabilístico 3D, esta técnica foi recentemente proposta para ser aplicada no monitoramento de ambientes através de múltiplas câmeras e encontram-se poucos trabalhos que a empreguem na análise de movimento. Após, utiliza-se Fluxo Óptico e um método similar ao algoritmo Expectation-Maximization (EM) para identificar e rastrear as partes do corpo dos objetos de interesse.

**Palavras-chave**— reconstrução volumétrica probabilística 3D, fluxo óptico, rastreamento de movimento, tele-operação.

### 1 Introdução

A tele-operação é um meio de operar um robô utilizando a inteligência humana e requer uma interface adequada entre homem e máquina (Cui et al., 2003). Atualmente, vários sistemas de tele-operação vêm sendo desenvolvidos, a fim de permitir a realização de tarefas perigosas ou em ambientes remotos como, por exemplo, em ambientes subaquáticos, usinas nucleares, campos de batalha; manipulação de materiais tóxicos; tele-cirurgia; neutralização de material explosivo; construção civil; operações policiais.

Normalmente, os sistemas de tele-operação são baseados em uma arquitetura mestre-escravo, na qual o operador interage diretamente com o manipulador mestre, cujos sensores definem os valores de referência a serem aplicados no manipulador escravo (Soares, 2006). Porém, deve-se notar que, utilizando tal arquitetura, apenas uma quantidade limitada de movimentos do operador humano pode ser analisada, além disso, utilizam-se equipamentos especificamente projetados para a aplicação em questão.

Outro fato observado é que grande parte dos sistemas robóticos de tele-operação utilizados no Brasil são desenvolvidos no exterior (em especial

no campo de intervenções referentes à exploração de petróleo) e, portanto, apresentam arquiteturas fechadas que impossibilitam sua adaptação a determinadas tarefas (Soares, 2006).

Neste contexto, nós sugerimos a aplicação de métodos de rastreamento de movimento baseado em múltiplas câmeras e sem marcas, para promover o uso da tele-operação sob um menor custo e de forma a permitir a análise de movimentos mais diversificados.

Assim, neste artigo, expomos o nosso método de análise de movimento, cujos principais diferenciais são:

- I) A reconstrução volumétrica empregada que é obtida pela técnica de Grid de Ocupação Probabilístico (Franco and Boyer, 2005). Esta técnica foi pouco explorada para reconstrução de ambientes monitorados por múltiplas câmeras e fornece resultados mais robustos que os métodos tradicionais. Além disso, informações sobre o movimento dos objetos, obtidas através da técnica de fluxo óptico, são adicionadas à reconstrução volumétrica;
- II) O método de rastreamento é aplicável a objetos sem forma pré-definida, sendo realizado com o emprego de um modelo simplifi-

cado de representação destes objetos e utilizando Fluxo Óptico obtido através das imagens 2D das diferentes câmeras.

Nota-se que, uma vez que este método pode ser aplicado a objetos de diferentes formas, o rastreamento e a reconstrução podem favorecer tanto a análise de movimento do operador humano como do robô a ser operado. A reconstrução do espaço de trabalho do robô possibilita, também, a obtenção de uma maior percepção quanto à interação do mesmo no seu ambiente.

Este artigo, então, apresenta um método de rastreamento de movimento aplicável à teleoperação. A Seção 2, a seguir, faz uma contextualização do campo de pesquisa de análise de movimento. A Seção 3 apresenta um resumo da arquitetura proposta. As Seções posteriores 4, 5 e 6 são dedicadas aos componentes do sistema. Finalizando, a Seção 7 apresenta os resultados do sistema em análise de movimento humano e a Seção 8 conclui o trabalho.

## 2 Sistemas de Análise de Movimento

O problema de captura de movimento é definido como a determinação de parâmetros da postura de uma pessoa (ou outro objeto) no espaço 3D. Este problema complexo pode ser decomposto nos seguintes estágios: inicialização, rastreamento, estimação de pose e reconhecimento do movimento. Dessa forma, o rastreamento pode ser empregado em captura de movimento, preparando os dados para os estágios posteriores (estimação de pose e reconhecimento) (Moeslund and Granum, 2001).

Existem muitos métodos para rastreamento em sistemas de captura de movimento. Estes métodos são baseados, principalmente, em técnicas magnéticas, mecânicas, acústicas ou visuais (Allen et al., 2001) (Zhoua and Hu, 2008).

Neste escopo, o rastreamento visual sem marcas tem recebido muita atenção dentro da comunidade científica, devido ao fato de que esta técnica apresenta algumas vantagens importantes em relação aos demais métodos, tais como: a) é uma técnica não intrusiva (não requer equipamentos extras anexados aos objetos de interesse); b) não requer equipamentos redundantes para o rastreamento de múltiplos objetos de interesse; c) permite a extração de informações extras como características físicas dos objetos (cor e textura); d) apresenta menores custos que os demais métodos.

O rastreamento visual pode ser definido como a detecção e localização recursiva de objetos ou, mais geralmente, de padrões visuais em seqüências de imagens (vídeos) (Pérez et al., 2004). Esta é uma importante tarefa em visão computacional (Kobayashi et al., 2006) e tem sido estudada extensivamente na literatura (Poppe, 2009), (Huang et al., 2009), (Canton-Ferrer et al., 2009). Este

interesse advém do grande número de aplicações para este problema, tal como: interface homem-máquina, tele-operação, estudos antropológicos, realidade virtual, entretenimento e vigilância.

Entretanto, o rastreamento de pessoas ou outros objetos utilizando imagens 2D é um problema difícil de ser tratado devido: a natureza complexa do movimento 3D; a perda de informações nas imagens devido à restrição do espaço 2D; alterações de cor devido a variações de luminosidade; existência de outros objetos em movimento na cena.

Assim, uma forma de minimizar tais problemas é a utilização de múltiplas câmeras dispostas ao redor do ambiente no qual o objeto de interesse se movimenta. A partir das imagens capturadas pelas diferentes câmeras, se obtém uma reconstrução volumétrica 3D, que consiste na localização e determinação do volume do objeto de interesse. A reconstrução volumétrica é freqüentemente empregada em conjunto com o rastreamento de movimento (Caillette, 2006), (Huang et al., 2009), (Canton-Ferrer et al., 2009), bem como, em outras aplicações, tais como na construção de animações realísticas, reconstrução de ambientes e visualização científica.

Em um contexto mais geral, podemos dizer que um sistema de rastreamento visual é formado por três componentes: I) **modelo de observação** - forma pelo qual os objetos de interesse são observados; II) **modelo de representação** - forma pelo qual os objetos são representados no sistema de rastreamento; III) **algoritmo de rastreamento** - procedimento que, utilizando o modelo de observação correspondente a um determinado instante de tempo, atualiza o estado do modelo de representação obtido no tempo anterior.

Na seção a seguir, então, apresentamos as definições feitas para cada um destes componentes.

## 3 Arquitetura do Método Proposto

Neste trabalho, o modelo de observação utilizado (Seção 4) é composto de: Fluxo Óptico, obtido pelo método Lucas Kanade; e reconstrução volumétrica 3D, obtida através da técnica de Grid de Ocupação Probabilístico. O modelo de representação (Seção 5) utiliza blobs gaussianas para representar as partes do corpo dos objetos de interesse. Então, estes objetos são rastreados por um algoritmo similar ao Expectation-Maximization (EM) (Seção 6).

A Figura 1 apresenta a arquitetura do método proposto. O sistema utiliza, como entrada, uma seqüência de imagens 2D capturadas por múltiplas câmeras sincronizadas e dispostas em torno do ambiente onde os objetos se movimentam.

**Modelo de Observação:** Primeiramente, antes do sistema iniciar a rastrear o movimento dos objetos, um conjunto de imagens das diferen-

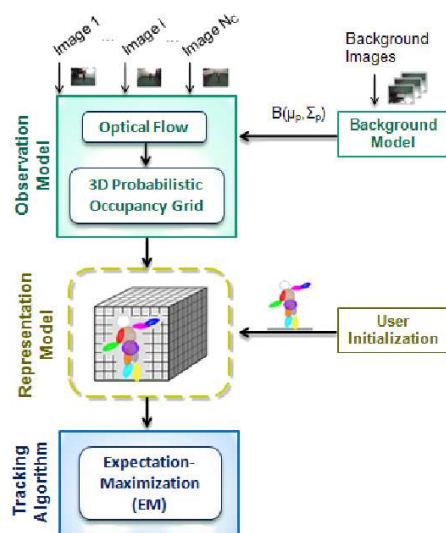


Figura 1: Arquitetura do método proposto.

tes câmeras é utilizado para construir um modelo de fundo da cena. Este modelo de fundo serve como entrada para o método de Grid de Ocupação Probabilístico.

Então, a cada novo instante de tempo, o método de Grid de Ocupação Probabilístico recebe uma imagem para cada câmera e constrói um grid de ocupação. Através deste grid, pode-se verificar quais elementos de volume (voxels) possuem maior probabilidade de estarem ocupados por um objeto de interesse. Além disso, nós adicionamos informações de fluxo óptico ao grid: cada voxel é associado a um vetor de velocidade 3D.

**Modelo de Representação:** No início do sistema, um usuário deve inicializar o modelo de representação dos objetos, utilizando o grid de ocupação correspondente ao primeiro instante de tempo: as blobs são posicionadas corretamente de acordo com as partes do corpo que elas representam. Nos instantes de tempo seguintes, este modelo de representação é atualizado pelo algoritmo de rastreamento.

**Algoritmo de Rastreamento:** O loop de rastreamento é definido por um método similar ao algoritmo Expectation-Maximization (EM) (Caillette, 2006), composto por dois passos. No primeiro, passo Expectation, cada um dos voxels do grid é associado a uma blob do modelo de representação. Após, no passo Maximization, cada uma das blobs é atualizada de acordo com as associações realizadas no passo Expectation. O modelo de representação atualizado serve, então, como entrada para o passo Expectation da próxima iteração do loop de rastreamento.

#### 4 Modelo de Observação

O problema tratado, nesta seção, é a reconstrução volumétrica probabilística 3D obtida a partir de

múltiplas câmeras e o Fluxo Óptico obtido a partir das imagens 2D e estimado no espaço 3D.

##### 4.1 Reconstrução Volumétrica

Na reconstrução volumétrica, em ambientes monitorados por múltiplas câmeras, freqüentemente se utiliza uma subtração binária do fundo, analisando cada imagem em separado, tal como realizado nos métodos tradicionais de Shape-from-Silhouette. No entanto, este tratamento pode dramaticamente alterar a percepção 3D que se teria, caso se observasse todas as imagens em conjunto, pois, intuitivamente, o conhecimento de todas as imagens simultâneas traz mais informações que o conhecimento de apenas uma imagem (Franco and Boyer, 2005).

Analisando este contexto, Franco (Franco and Boyer, 2005) propôs calcular a fusão das informações de todas as imagens no espaço 3D, antes de realizar avaliações sobre cada imagem individualmente. Para isto, Franco utilizou a técnica de Grid de Ocupação Probabilístico. Esta técnica tem sido amplamente utilizada na comunidade robótica para representar o ambiente de navegação de robôs monitorado por sensores de profundidade e de medidas de orientação (Elfes, 1989). Franco propôs, então, estender o conceito do Grid de Ocupação para sensores baseados em imagens.

Esta técnica surge, então, como uma forma de obter uma reconstrução mais robusta, contornando problemas de variações de luminosidade e semelhança entre as cores de fundo e dos objetos de interesse. Dessa forma, evita-se a ocorrência de ruídos e reconstruções incompletas, que ocorrem com maior freqüência nos métodos tradicionais de Shape-from-Silhouette.

No método de Grid de Ocupação Probabilístico 3D, as imagens obtidas pelas múltiplas câmeras sincronizadas são unificadas em um grid de ocupação. Cada pixel da câmera é tratado como um sensor estatístico suscetível a incertezas. O problema é, então, tratado como uma estimativa Bayesiana (Bishop, 2006). O espaço 3D é discretizado em elementos de volumes, nomeados de voxels, como ilustrado na Figura 2, e para cada voxel se calcula a probabilidade deste estar ocupado por um objeto de interesse.

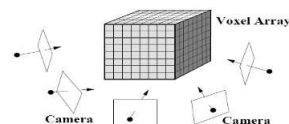


Figura 2: Espaço 3D discretizado em voxels

**Modelo de fundo.** Primeiramente, é necessária a construção de um modelo de fundo da cena livre de objetos em movimento. Assim, é adquirido um conjunto de imagens da cena livre de ob-

jetos. A partir destas, se constrói um modelo estatístico de fundo (Wren et al., 1997): cada pixel  $p$  de cada câmera  $i$  é modelado por uma distribuição gaussiana representada por um vetor média  $\mu_p^i$  de cor e por uma matriz de covariância  $\Sigma_p^i$ . Sendo  $D = 3$  a dimensão da gaussiana, cada pixel terá a seguinte probabilidade de pertencer ao fundo:

$$N(\mu, \Sigma) \cong \frac{1}{(2\pi)^{\frac{D}{2}} \sqrt{|\Sigma|}} \exp \left\{ -\frac{1}{2}(x - \mu)\Sigma^{-1}(x - \mu)^T \right\}$$

No entanto, um grande número de multiplicações de pequenos valores pode sofrer underflow de precisão numérica; assim, a expressão acima é transformada em uma soma de logaritmos:

$$\log N(\mu, \Sigma) = -\frac{3}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma|) - \frac{1}{2}(x - \mu)\Sigma^{-1}(x - \mu)^T$$

**Inferência Bayesiana.** As ocupações dos voxels, por questões de custo computacional, são consideradas estatisticamente independentes. E, portanto, se calcula a probabilidade de ocupação de cada voxel independentemente.

Como entrada do sistema, são utilizadas as seguintes informações:  $I_p^i$  (cor relativa ao píxel  $p$  da câmera  $i$ );  $B^i$  (modelo estatístico de fundo da câmera  $i$ );  $P^i$  (matriz de projeção da câmera  $i$ );  $\tau$  (simboliza o conhecimento a priori que se tem sobre a cena ou sobre as características da câmera, neste trabalho, não se utilizou qualquer conhecimento a priori).

Como saída, o sistema de reconstrução deve fornecer:  $G$ , o grid de ocupação (Figura 2). Para cada voxel  $X$  do grid é associada uma variável binária de ocupação  $G_X \in \{0, 1\}$ , o valor 0 informa que o voxel está livre de qualquer objeto de interesse e o valor 1, que está ocupado.

Além disso, se define, para cada pixel  $p$  de cada câmera  $i$ , uma variável binária  $F_p^i$  que representa a detecção da silhueta de um objeto de interesse.  $F_p^i = 1$  se a câmera  $i$  informa que o pixel  $p$  pertence a um objeto de interesse.

As relações de dependências entre as variáveis foram modeladas de acordo com (Franco and Boyer, 2005):

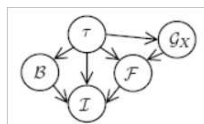


Figura 3: Dependência entre as variáveis

Assim, para inferir a probabilidade de ocupação  $G_X$ , deve-se computar a probabilidade conjunta  $p(G_X, I, B, F, \tau)$ :

$$p(G_X, I, B, F, \tau) = p(\tau)p(B|\tau)p(G_X|\tau)p(F|G_X, \tau)p(I, F, B, \tau)$$

$p(\tau), p(B|\tau), p(G_X|\tau)$  são, respectivamente, a probabilidade a priori do conjunto de parâmetros;

a probabilidade prévia do fundo da cena; e a probabilidade prévia de ocupação do voxel  $X$ . Estes parâmetros foram definidos como distribuições uniformes.

$p(F|G_X, \tau)$  é o termo de probabilidade da silhueta, depende do conhecimento prévio do sistema e da ocupação do voxel relacionado ao pixel  $(p, i)$ , pois a ocupação de um voxel que é projetado sobre  $p$  na imagem da câmera  $i$  explica a detecção de um objeto nesta mesma imagem.

$p(I, F, B, \tau)$  é o termo de probabilidade da imagem. As cores  $I$  nas imagens são condicionadas pela detecção  $F$  de objetos nestas imagens e pelo modelo de fundo da cena  $B$ .

Desconsiderando as distribuições uniformes, e tratando as observações de cor  $I$  de cada pixel como independentes uma das outras, pode-se simplificar a equação acima da seguinte forma:

$$p(G_X, I, B, F, \tau) = p(F|G_X, \tau)p(I|F, B, \tau)$$

*I - Termo de Formação da Silhueta.*  $p(F|G_X, \tau)$  modela a resposta da detecção da silhueta do pixel  $(i, p)$  para o estado de ocupação  $G_X$  do voxel  $X$ .

a) Se  $G_X = 1$ , o voxel  $X$  está ocupado por um objeto de interesse e, então, a detecção da silhueta é regida pela distribuição de probabilidade  $P_d(F_p^i)$ . Na prática, esta distribuição é definida utilizando uma constante  $P_D \in [0, 1]$ , que é parâmetro do sistema: i.  $P_d([F_p^i = 1]) = P_D$  é a taxa de detecção correta de silhueta em um pixel, e ii.  $P_d([F_p^i = 0]) = 1 - P_D$  é a taxa de falha nesta detecção. Falhas na detecção ocorrem quando o pixel relata que não existe qualquer objeto sobre a sua linha de visão quando na verdade existe. Esta consideração é útil para este problema, pois algumas vezes a extração da silhueta falha localmente e a consideração de incertezas permite que o sistema recupere a informação correta do voxel.

b) Se  $G_X = 0$ , o voxel  $X$  não está ocupado e, portanto, não se pode tirar conclusões sobre o seu estado de detecção. Logo, a detecção é regida pela distribuição  $P_f(F_p^i)$ . Esta distribuição é definida utilizando uma constante  $P_{FA} \in [0, 1]$ , parâmetro do sistema: i.  $P_f([F_p^i = 1]) = P_{FA}$  é a taxa de falsos alarmes de um pixel. Falsos alarmes ocorrem quando erroneamente um pixel demonstra a existência de algum objeto sobre a sua linha de visão, quando na verdade não existe. ii.  $P_f([F_p^i = 0]) = 1 - P_{FA}$  é a taxa com a qual esperamos que o pixel  $(i, p)$  informe corretamente a não-detecção de um objeto.

*II - Termo de Formação da Imagem.*  $p(I_p^i|F_p^i, B_p^i, \tau)$  pode ser interpretado da seguinte forma:

a) se  $F_p^i = 1$ , um objeto é encontrado no pixel  $(i, p)$  e, portanto, o conhecimento sobre o fundo é irrelevante para definir a cor esperada para este pixel (o fundo é obstruído por um objeto, cuja cor não foi definida previamente). Neste caso, este

termo é considerado como uma distribuição uniforme:  $p(I_p^i|F_p^i, B_p^i, \tau) = U(I_p^i)$ ;

b) se  $F_p^i = 0$ , nenhum objeto se encontra no pixel  $(i, p)$ , então, a cor observada neste pixel deve ser similar à cor do modelo de fundo. Neste caso, o valor de probabilidade é calculado utilizando a distribuição gaussiana do modelo de fundo:  $p(I_p^i|[F_p^i = 0], [B_p^i = \mu_p^i, \Sigma^i], \tau) = N(I_p^i|\mu_p^i, \Sigma^i)$ .

#### Inferência da Ocupação de cada Voxel.

Uma vez que as probabilidades conjuntas tenham sido determinadas, é possível utilizar o teorema de Bayes para inferir a probabilidade da variável buscada  $G_X$ :

$$p(G_X|I, B, \tau) = \frac{\sum_F p(G_X, I, B, F, \tau)}{\sum_{G_X, F} p(G_X, I, B, F, \tau)} = \frac{\sum_{F_p^i} p(F_p^i|G_X, \tau) p(I_p^i|F_p^i, B_p^i, \tau)}{\sum_{G_X} \sum_{F_p^i} p(F_p^i|G_X, \tau) p(I_p^i|F_p^i, B_p^i, \tau)}$$

Após, define-se um limiar para a probabilidade  $p(G_X|I, B, \tau)$ : todos os voxels cuja probabilidade passar deste limiar serão utilizados no rastreamento.

#### 4.2 Fluxo Óptico

O Fluxo óptico é a distribuição 2D da velocidade aparente de movimento de padrões no plano da imagem (Horn and Schunck, 1981). Ou seja, o campo de fluxo óptico consiste em um campo denso de velocidades, onde cada pixel, no plano da imagem, está associado a um único vetor de velocidade (Barbosa et al., 2005).

Métodos baseados em fluxo óptico têm sido frequentemente utilizados, na literatura, para a análise de movimento, tanto no domínio 2D (Barbosa et al., 2005), quanto no domínio 3D (Barron and Thacker, 2005). No entanto, até o momento, nenhum trabalho foi encontrado utilizando fluxo óptico em conjunto com a técnica de Grid de Ocupação Probabilístico 3D.

Partindo do pressuposto que a agregação de mais informações ao grid probabilístico 3D (além da cor) é possível e pode favorecer a obtenção de maiores informações na reconstrução 3D (Franco and Boyer, 2005), consideramos que o fluxo óptico possa ser uma importante fonte de informação sobre o movimento a ser analisado. Assim, propomos obter o fluxo óptico individualmente nas imagens 2D das diferentes câmeras e estimar um vetor de fluxo para cada voxel do espaço 3D, este último é, então, utilizado como informação no algoritmo de rastreamento.

**Espaço 2D.** Primeiramente, o fluxo óptico é obtido nas imagens 2D, através do algoritmo Lucas-Kanade (Lucas and Kanade, 1981).

**Espaço 3D.** Após, para cada voxel no Grid 3D, projeta-se a coordenada de seu centro para o espaço 2D das imagens de todas as câmeras, de forma a obter um pixel correspondente a este

voxel, para cada imagem. Como cada pixel apresenta um vetor de fluxo, ao final do processo, cada voxel será relacionado com  $N_C$  vetores de fluxo (sendo  $N_C$  o número de câmeras utilizadas). Então, projeta-se estes  $N_C$  vetores de fluxo 2D para o espaço 3D do Grid e calcula-se uma média. O resultado desta média será o valor de fluxo 3D do voxel em questão.

**Pós-processamento.** Tendo-se, então, um vetor de fluxo para cada voxel, utiliza-se um pós-processamento de forma a diminuir erros de ruído. Para cada voxel, se analisa os vetores de fluxo de uma janela de voxels ao redor do mesmo. Ordena-se as coordenadas  $x, y$  e  $z$  destes vetores de fluxo dos voxels contidos na janela separadamente e se emprega as coordenadas medianas  $x_M, y_M$  e  $z_M$  como as coordenadas do vetor de fluxo final deste voxel.

## 5 Modelo de Representação

Objetos articulados, como o corpo humano, são, em sua maioria, formados por partes rígidas, que individualmente não apresentam modificações significativas em suas formas. Partindo desta suposição, neste trabalho, cada parte do objeto é representada por um modelo gaussiano, nomeado de blob.

**Posição e Cor.** Com referência a informação espacial, uma blob é frequentemente representada por uma forma elipsoidal, tendo a sua superfície definida pelo desvio padrão em torno do valor médio de posição. Semelhantemente, a informação de cor é modelada por uma cor média e a sua covariância. Portanto, uma blob é uma distribuição gaussiana com seis dimensões (as gaussianas D-dimensionais são representadas pela equação apresentada na Seção 4).

Analisando o conceito de gaussiana, pode-se determinar o quão distante cada voxel está em relação a cada blob. Considerando um voxel  $V$  na posição  $X_V$  com as cores  $C_V = \{C_V^1, \dots, C_V^m, \dots, C_V^{N_C}\}$  vistas pelas  $N_C$  câmeras, a distância de Mahalanobis  $D_M(V, B) = (V - \mu)\Sigma^{-1}(V - \mu)^T$  mede a distância deste voxel a blob  $B(\mu, \Sigma)$ .

Supondo que não existe dependência entre as informações de cor e posição, pode-se simplificar  $D_M(V, B)$  ( $\mu_X, \Sigma_X$  referem-se à posição e  $\mu_C, \Sigma_C$ , à cor):

$$D_M(V, B) = (X_V - \mu_X)\Sigma_X^{-1}(X_V - \mu_X)^T + (C_V^{mc} - \mu_C)\Sigma_C^{-1}(C_V^{mc} - \mu_C)^T$$

Para o cálculo de  $D_M(V, B)$ , utiliza-se apenas o vetor de cor  $C_V^{mc}$  que minimiza a distância ao modelo da blob, ou seja:

$$mc = \arg_{i=1 \dots N_C} \min(C_V^i - \mu_C)\Sigma_C^{-1}(C_V^i - \mu_C)^T$$

**Inicialização do modelo.** Um procedimento de inicialização é necessário, a fim de fornecer uma estimativa inicial dos parâmetros das blobs para o algoritmo de rastreamento. Neste projeto, foi utilizada uma inicialização manual, na qual um usuário informa, para cada blob: a) dois pontos extremos do principal eixo de variância da blob em questão. Através destes pontos pode-se estimar o valor médio de posição da blob e a direção do principal eixo de variação da gaussiana. b) três valores  $\sigma_x, \sigma_y, \sigma_z$ , que informarão os desvios padrões em cada eixo de direção.

Denominando-se  $R$  a matriz de rotação do elipsóide (blob) em relação aos eixos  $X, Y, Z$ , a matriz de covariância  $\Sigma_X$  pode ser calculada por:

$$\Sigma_X = R \cdot \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix} \cdot R^T$$

Para a obtenção da matriz  $R$ , pode-se utilizar o principal eixo de variação da blob que permite calcular os ângulos desse eixo com relação aos eixos  $X, Y, Z$  da cena (ângulos de Euler), seguindo a Figura 4.

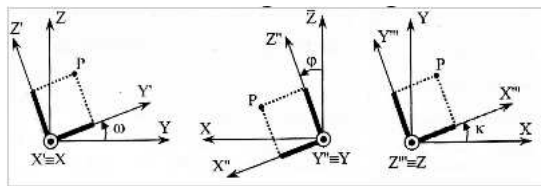


Figura 4: Ângulos de Euler - para a determinação da matriz de rotação  $R$

A partir destes ângulos e considerando que a rotação inicial é dada por  $R_X(\omega)$ , uma secundária por  $R_Y(\varphi)$  e uma final dada por  $R_Z(\kappa)$ , a matriz de rotação resultante  $R$  pode ser calculada da seguinte maneira (Galo and Tozzi, 2001) ( $s$  representa o seno e  $c$ , o cosseno):

$$\omega = \arctg\left(\frac{z_b - z_a}{y_b - y_a}\right) \quad \varphi = \arctg\left(\frac{x_b - x_a}{z_b - z_a}\right)$$

$$\kappa = \arctg\left(\frac{y_b - y_a}{x_b - x_a}\right)$$

$$R = \begin{pmatrix} c\varphi c\kappa & s\omega s\varphi c\kappa + c\omega s\kappa & -c\omega s\varphi c\kappa + s\omega s\kappa \\ -c\varphi s\kappa & -s\omega s\varphi s\kappa + c\omega c\kappa & c\omega s\varphi s\kappa + s\omega c\kappa \\ s\varphi & -s\omega c\varphi & c\omega c\varphi \end{pmatrix}$$

**Agregando informação de movimento ao modelo de representação - Fluxo Óptico.** Até este momento, temos associado a cada voxel  $V$ , no instante  $t$ , uma posição no espaço 3D  $X = \{x, y, z\}$  e o seu vetor de fluxo neste instante  $F = \{dx, dy, dz\}$  (calculado através dos frames referentes ao tempo  $t - 1$  e  $t$ ).

Então, pode-se estimar a próxima posição que este voxel  $V$  deva ocupar, caso seu movimento não se altere entre os instantes de tempo de  $t - 1$  a  $t + 1$ . Dessa forma, se calcula  $\dot{X} = \{\dot{x}, \dot{y}, \dot{z}\}$  que indica a posição que este voxel deve ocupar no tempo  $t + 1$ .

Além disso, sabendo que todos os voxels de uma determinada parte rígida do corpo do objeto de interesse (referentes a uma determinada blob) devem apresentar um mesmo movimento, pode-se estimar uma matriz de transformação  $H$  (Chum et al., 2005) para cada blob, que mapeie uma transformação linear de rotação e translação do conjunto de voxels desta blob, em um dado instante de tempo  $t$ , para o próximo instante  $t + 1$ .

Denominando-se  $A$  uma matriz com as posições  $X$  em coordenadas homogêneas de todos os voxels de uma determinada blob e  $\dot{A}$ , uma matriz com as posições  $\dot{X}$  estimadas para o próximo instante de tempo, a matriz  $H$  pode ser calculada da seguinte forma:

$$H \cdot A = \dot{A} \rightarrow H = \dot{A} \cdot pseudo\_inversa(A)$$

Sendo  $pseudo\_inversa(A)$  a matriz pseudo-inversa de  $A$ , pois  $A$  não é uma matriz quadrada.

## 6 Rastreamento

O Rastreamento é executado por um método similar ao algoritmo Expectation-Maximization (EM) (Caillette, 2006): o algoritmo EM tradicional foi modificado para tratar informações de movimento e incorporar a informação de distância euclidiana. O método empregado é composto pelos dois passos seguintes (executados repetidamente):

**Expectation:** para cada elemento de volume da cena discretizada (voxel), é determinado a qual parte dos objetos este elemento pertence, ou seja, a qual blob. Para tal tarefa, se utiliza os parâmetros das blobs estimadas nas iterações anteriores do EM ou inicializados pelo usuário (caso seja a primeira iteração do EM).

Assim, para cada voxel  $V_i$ , se computa a distância deste a cada blob  $B_j$  e, então, se designa este voxel a blob com menor distância. O valor de distância utilizado é uma ponderação de três parâmetros:

$$D(V_i, B_j) = k_1 \cdot D_E + k_2 \cdot D_P + k_3 \cdot D_M$$

Onde  $k_1, k_2$  e  $k_3$  são constantes de ponderação dos diferentes parâmetros:

*I - Distância Euclidiana*  $D_E$  é a distância euclidiana de posição entre o voxel  $V_i$  e o centro da blob  $B_j$ .

*II - Distância probabilística*  $D_P$  é um parâmetro de distância que considera a probabilidade do voxel  $V_i$  pertencer à gaussiana referente à blob  $B_j$ .

Definindo que a blob  $B_j$  é representada pelo vetor médio  $\mu_j$  e pela matriz de covariância  $\Sigma_j$ , a probabilidade  $p(V_i, B_j)$  que  $V_i$  tem de pertencer a blob  $B_j$  é calculada por uma gaussiana de 6 dimensões:

$$P(V_i|B_j) = \frac{1}{(2\pi)^3 \sqrt{|\Sigma_j|}} \cdot e^{-\frac{1}{2} D_M(V_i, B_j)}$$



Onde  $D_M(V_i, B_j)$  é a distância de Mahalanobis (calculada na Seção 5) e  $|\Sigma_j|$  é o determinante da matriz de covariância de  $B_j$ . Supondo que não existe dependência entre as informações de cor e posição,  $|\Sigma_j|$  pode ser reduzida por um produto dos determinantes das matrizes de covariância de posição e cor.

Uma otimização padrão é comparar os logaritmos das probabilidades ao invés da própria probabilidade:  $\log P(V_i|B_j) = -3\log(2\pi) - \frac{1}{2}\log|\Sigma_j| - \frac{1}{2}D_M(V_i, B_j)$ .

Desconsiderando os termos e fatores multiplicativos constantes, obtém-se o valor que denominamos de  $D_P$ , cuja minimização é equivalente à maximização da função de probabilidade original:

$$D_P = \log|\Sigma_j| + D_M(V_i, B_j)$$

Minimizar  $D_P$  depende principalmente da distância de Mahalanobis  $D_M()$ . O termo  $\log|\Sigma_j|$  é constante para cada blob e utiliza a variância codificada na matriz  $\Sigma_j$  para favorecer as blobs menores, fornecendo a estas maiores chances de serem escolhidas pelos voxels. Este termo é útil quando duas blobs são muito semelhantes (em posição e cor), pois, se apenas  $D_M()$  fosse utilizada, a blob com maior variância sempre teria maior vantagem em ser escolhida.

*III - Distância de Movimento  $D_M$*  é um parâmetro que analisa o vetor de fluxo estimado para o voxel  $V_i$  em relação à matriz de transformação linear  $H$  calculada para a blob  $B_j$ . Primeiramente, sendo  $X$  a posição do voxel  $V_i$ , se calcula  $\dot{X}_{B_j}$  que indica a próxima posição que o voxel  $V_i$  deveria ocupar se este pertencesse a blob  $B_j$  ( $\dot{X}_{B_j} = H \cdot X$ ). Após, se estima  $\dot{X}$  que indica a próxima posição que o voxel  $V_i$  deve ocupar através da análise do vetor de fluxo  $F = \{dx, dy, dz\}$  estimado para este voxel ( $\dot{X} = X + F$ ). Então, pode-se calcular o parâmetro de distância  $D_M$  entre os movimentos esperados para a blob  $B_j$  e para o voxel  $V_i$  dado pela da distância euclidiana entre  $\dot{X}_{B_j}$  e  $\dot{X}$ , ou seja,  $D_M = \dot{X}_{B_j} - \dot{X}$ .

**Maximization:** novos valores para o vetor média  $\mu_j$  e para a matriz de covariância  $\Sigma_j$  de cada blob  $B_j$  são calculados, utilizando, para isto, as informações referentes aos voxels designados à blob  $B_j$  (durante o passo Expectation). Após, os parâmetros atualizados são, então, utilizados como estimativa inicial para o passo Expectation do próximo instante de tempo da seqüência de vídeo.

## 7 Resultados

Um software foi implementado, para permitir o uso dos métodos apresentados, na análise das seqüências de vídeo adquiridas no repositório público (Perception-Group, 2008). Um exemplo das imagens da seqüência analisada é demonstrado na Figura 5-a.

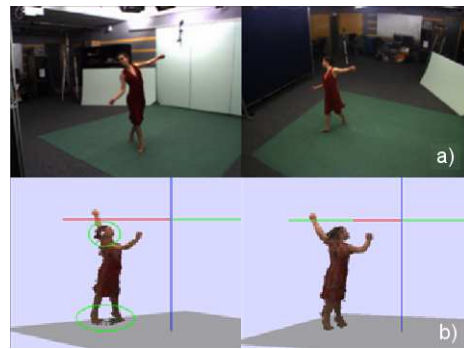


Figura 5: a) Amostra de imagens da seqüência analisada de duas câmeras diferentes. b) Reconstrução volumétrica. Esq: método tradicional Shape-from-silhouette (falhas em verde). Dir: técnica Grid de Ocupação Probabilístico 3D

Testes foram realizados, para permitir a comparação entre as reconstruções volumétricas 3D obtidas pelo método tradicional Shape-from-Silhouette e pelo método de Grid de Ocupação Probabilístico. Exemplos destes resultados são demonstrados na Figura 5-b.

A Figura 6 apresenta a reconstrução 3D obtida em três instantes de tempo consecutivos com os vetores de Fluxo Óptico correspondentes.

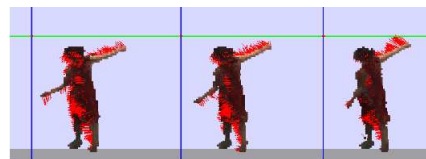


Figura 6: Reconstrução com Fluxo Óptico 3D

Os resultados do rastreamento são mostrados na Figura 7. As cores representam diferentes grupos de voxels, cada conjunto de voxels de uma única cor pertence a uma única blob do modelo de representação.

## 8 Conclusão

Este artigo apresentou um método para rastreamento de movimento sem marcas, em ambientes monitorados por múltiplas câmeras, que é aplicável à tele-operação. O método proposto utiliza, como observação, Fluxo Óptico em conjunto com uma reconstrução volumétrica 3D, obtida através da técnica de Grid de Ocupação Probabilístico 3D. Esta técnica foi recentemente apresentada, para uso em ambientes monitorados por múltiplas câmeras e encontram-se poucos trabalhos que a empreguem no rastreamento de movimento.

A reconstrução volumétrica, executada com o método de Grid de Ocupação Probabilístico, apresentou vantagens consideráveis em relação ao método tradicional de Shape-from-Silhouette, conforme demonstrado na seção de resultados. Este

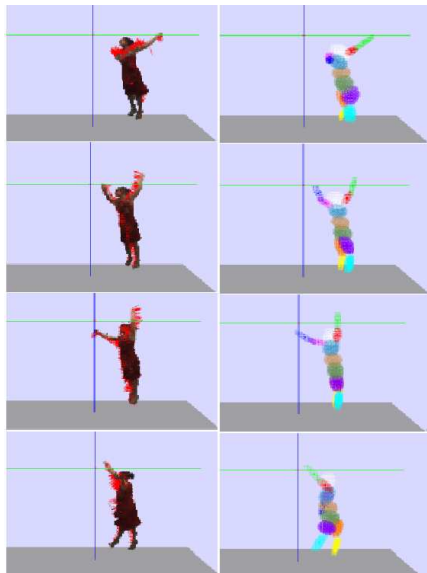


Figura 7: Resultados do rastreamento

fato deve-se à vantagem do método de reconstrução probabilística considerar as informações de todas as imagens em conjunto, através de fusão obtida por inferência Bayesiana. Portanto, o Grid Probabilístico não realiza avaliações prematuras da análise de cada câmera em separado, como o método tradicional.

Um diferencial da técnica de reconstrução empregada se deve a capacidade desta superar problemas de reconstruções incompletas (comuns nos métodos tradicionais). Nos resultados demonstrados, os métodos tradicionais de reconstrução falharam na análise de partes do objeto de interesse cuja cor se assemelhava com a cor de fundo da cena, além disso, os métodos tradicionais apresentaram-se mais suscetíveis a variações de luminosidades (sombras).

O fluxo óptico demonstrou ser uma importante fonte de informação sobre o movimento, possibilitando que o rastreamento fosse executado corretamente (ver Figura 7), mesmo sem a consideração de posições de juntas ou quaisquer definições de posições relativas entre as blobs e, dessa forma, o método proposto pode ser aplicado a objetos de diferentes formas.

Finalmente, pode-se concluir que este estudo mostrou resultados promissores, estabelecendo uma base para a investigação de métodos mais robustos. Como trabalho futuro, nós pretendemos avaliar o emprego de algoritmos de rastreamento mais fortes, como o Filtro de Partículas e o Filtro de Kalman.

### Agradecimentos

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior CAPES e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq.

### Referências

- Allen, B. D., Bishop, G. and Welch, G. (2001). Course 11 - tracking: Beyond 15 minutes of thought, *SIGGRAPH 2001 Courses*.
- Barbosa, R. L., Gallis, R. B. A., Silva, J. F. C. and Júnior, M. M. (2005). A computação do fluxo óptico em imagens obtidas por um sistema móvel de mapeamento terrestre, *Revista Brasileira de Cartografia*.
- Barron, J. L. and Thacker, N. A. (2005). Tutorial: Computing 2d and 3d optical flow, *Tina Memo Internal*, University of Manchester.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, Springer.
- Caillette, F. (2006). *Real-Time Markerless 3D Human Body Tracking*, Phd thesis, University of Manchester.
- Canton-Ferrer, C., Casas, J. and Pardàs, M. (2009). Voxel-based annealed particle filtering for markerless 3d articulated motion capture, *3DTV Conference*.
- Chum, O., Pajdla, T. and Sturm, P. (2005). The geometric error for homographies, *Computer Vision and Image Understanding*.
- Cui, J., Tosunoglu, S., Roberts, R., Moore, C. and Repperger, D. (2003). A review of teleoperation system control, *Florida Conference on Recent Advances in Robotics, FCRAR 2003*.
- Elfes, A. (1989). Using occupancy grids for mobile robot perception and navigation, *IEEE Computer, Special Issue on Autonomous Intelligent Machines*.
- Franco, J. S. and Boyer, E. (2005). Fusion of multi-view silhouette cues using a space occupancy grid, *ICCV 05*.
- Galo, M. and Tozzi, C. L. (2001). A representação de matrizes de rotação e o uso de quatérnions em ciências geodésicas, *Série em Ciências Geodésicas*.
- Horn, B. K. P. and Schunck, B. G. (1981). Determining optical flow, *Artificial Intelligence* **17**: 185–203.
- Huang, P., Hilton, A. and Starck, J. (2009). Human motion synthesis from 3d video, *CVPR*.
- Kobayashi, Y., Sugimura, D., Hirasawa, K., Suzuki, N., Kage, H., Sato, Y. and Sugimoto, A. (2006). 3d head tracking using the particle filter with cascaded classifiers, *British Machine Vision Conference* pp. 1–37.
- Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision, *Proceedings of Imaging understanding workshop*.
- Moeslund, T. B. and Granum, E. (2001). A survey of computer vision based human motion capture, *CVIU*.
- Perception-Group (2008). Perception's website, <http://perception.inrialpes.fr>.
- Poppe, R. W. (2009). *Discriminative Vision-Based Recovery and Recognition of Human Motion*, Phd thesis, University of Twente.
- Pérez, P., Vermaak, J. and Blake, A. (2004). Data fusion for visual tracking with particles, *Proc. IEEE*.
- Soares, B. F. F. (2006). Controle servo-bilateral de um manipulador robótico teleoperado, *Technical report*, Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio, Rio de Janeiro - RJ.
- Wren, C., Azarbayejani, A., Darrell, T. and Pentland, A. (1997). Pfunder: Real-time tracking of the human body, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**.
- Zhou, H. and Hu, H. (2008). Human motion tracking for rehabilitation - a survey, *Biomedical Signal Processing and Control*.