

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL  
FACULDADE DE INFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**RECUPERAÇÃO DE POSES HUMANAS 3D  
A PARTIR DE IMAGENS BIDIMENSIONAIS**

LEANDRO LORENZETT DIHL

Tese apresentada como requisito à obtenção do grau de Doutor em Ciência da Computação na Pontifícia Universidade Católica do Rio Grande do Sul.

Orientador: Prof<sup>a</sup> Dr<sup>a</sup> Soraia Raupp Musse

**Porto Alegre  
2013**



D575r Dihl, Leandro Lorenzett  
Recuperação de poses humanas 3D a partir de imagens  
bidimensionais / Leandro Lorenzett Dihl. – Porto Alegre, 2013.  
83 f.

Tese (Doutorado) – Fac. de Informática, PUCRS.  
Orientador: Prof<sup>a</sup>. Dr<sup>a</sup>. Soraia Raupp Musse.

1. Informática. 2. Processamento de Imagens. 3. Movimento  
Humano (Fisiologia). I. Musse, Soraia Raupp. II. Título.

CDD 006.61

**Ficha Catalográfica elaborada pelo  
Setor de Tratamento da Informação da BC-PUCRS**



Pontifícia Universidade Católica do Rio Grande do Sul  
FACULDADE DE INFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

### TERMO DE APRESENTAÇÃO DE TESE DE DOUTORADO

Tese intitulada "Recuperação de Poses Humanas 3D a partir de Imagens Bidimensionais", apresentada por Leandro Lorenzetti Dihl, como parte dos requisitos para obtenção do grau de Doutor em Ciência da Computação, aprovada em 12/08/2013 pela Comissão Examinadora:

---

Profa. Dra. Soraia Raupp Musse  
Orientadora

PPGCC/PUCRS

---

Prof. Dr. Márcio Sarroglia Pinho

PPGCC/PUCRS

---

Prof. Dr. Léo Pini Magalhães

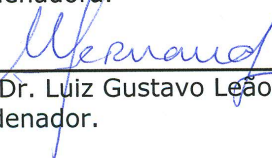
UNICAMP

---

Prof. Dr. Roberto Marcondes Cesar Junior

USP

Homologada em 27.08.2013, conforme Ata No. 15 pela Comissão Coordenadora.

  
Prof. Dr. Luiz Gustavo Leão Fernandes  
Coordenador.

**PUCRS**

**Campus Central**

Av. Ipiranga, 6681 – P. 32 – sala 507 – CEP: 90619-900

Fone: (51) 3320-3611 – Fax (51) 3320-3621

E-mail: [ppgcc@pucrs.br](mailto:ppgcc@pucrs.br)

[www.pucrs.br/facin/pos](http://www.pucrs.br/facin/pos)



## AGRADECIMENTOS

Quando traçamos nossos objetivos e vamos em busca deles, nosso êxito somente será completo se no final da conquista dividirmos os méritos conquistados e agradecermos as pessoas que estiveram presente ao nosso lado durante todo o percurso. Por essa razão, expresso aqui os meus mais sinceros agradecimentos.

À Prof<sup>a</sup> Dr<sup>a</sup> Soraia Raupp Musse, minha orientadora, pela competência científica e acompanhamento do trabalho, pela disponibilidade e generosidade reveladas ao longo destes anos de trabalho, assim como pelas críticas, correções e sugestões relevantes feitas durante a orientação. Mas não somente pelo lado profissional, também pela pessoa humana e compreensiva, justa e dedicada com todos os seus alunos.

Queria agradecer também a minha amiga e colega Adriana Braun pela sua ajuda, nos trabalhos desenvolvidos juntos, por seus "pitacos", e pelas nossas trocas de idéias durante estes 4 anos. Ao colega Dr<sup>o</sup> Julio Jacques, por suas idéias e os trabalhos desenvolvidos neste período.

Um agradecimento especial a todos os colegas e amigos do VhLab, que foram muitos nesses 4 anos, e que de alguma forma ajudaram-me na conclusão desta tese. Também queria mencionar o aluno Jorge Lucas pelo trabalho dedicado aos resultados da minha tese e agradecer à todos os colegas CPCA que de alguma forma contribuíram para a conclusão deste trabalho.

Por fim, agradeço a minha família, minha querida mãe, Terezinha Dihl, que sempre me incentivou desde criança na busca pelo estudo e conhecimento. Meus queridos filhos Yann, Yuri e Sophia que são uma das razões da minha busca e evolução do meu crescimento profissional. Minha esposa, Adriana Dihl, por tudo que ela representa para mim, minha companheira há 20 anos, sempre ao meu lado nos momentos difíceis e apoiando as minhas decisões.

Esse trabalho foi desenvolvido com apoio financeiro da Hewlett-Packard Brasil Ltda.

# RECUPERAÇÃO DE POSES HUMANAS 3D A PARTIR DE IMAGENS BIDIMENSIONAIS

## RESUMO

Esta tese apresenta um novo modelo para identificar poses humanas 3D a partir de fotos. Dada uma única imagem de entrada e um novo modelo de caracterização de posturas baseado no seu “conforto”, este trabalho visa resolver desambiguidades e gerar posturas que sejam humanamente possíveis em 3D. Este modelo de recuperação de pose humana 3D é baseado no método de Taylor [Tay00] e inclui melhorias em termos de restrições biomecânicas com o objetivo de reduzir o espaço de busca de possíveis posturas 3D que possam representar a pose na imagem 2D. Além disso, propõe-se um sistema de classificação que pode ser utilizado para sugerir as melhores posições geradas de acordo com as características de conforto e da luminosidade da imagem que também são exploradas. O critério de conforto adota premissas em termos de equilíbrio da postura, enquanto o critério de luminosidade elimina as ambiguidades de posturas, levando em conta a iluminação da imagem. Deve-se enfatizar que a recuperação de poses em 3D, que correspondem a uma única imagem bidimensional, é o foco principal deste trabalho. O trabalho também propôs uma aplicação para minimização de poses ambíguas baseado em auto-occlusão. Foram realizadas análises a fim de verificar a validade dos modelos propostos nesta tese.

**Palavras-chave:** Processamento de Imagens, Análise Visual do Movimento Humano, Obtenção de Poses Humanas.

# RECOVERING 3D HUMAN POSE FROM TWO-DIMENSIONAL IMAGES

## ABSTRACT

This thesis presents a new model to identify 3D human poses from photos. Given a single input image and a new model for characterization of postures based on their "comfort", this work aims to solve ambiguous postures and to generate the possible postures in 3D. The 3D human pose recovery model is based on the method of Taylor [Tay00] and includes improvements with biomechanical constraints in order to reduce the search space of possible 3D postures that can represent the pose in the 2D image. Furthermore, it is proposed a classification system which can be used to suggest the best positions generated according to the features of comfort and lightness of the image which are also used. The comfort criteria is related to assumptions in terms of the posture balance, while the shade criterion eliminates the ambiguities of postures, taking into account the illumination in the image. It should be emphasized that the recovery poses in 3D from just a single two-dimensional image is the main focus of this work. The paper also proposes one application for minimizing ambiguous poses based on self-occlusion. Result analysis was performed to check the validity of the models proposed in this thesis.

**Keywords:** Image Processing, Visual Analysis of Human Movement, Recovery of human poses.



## LISTA DE FIGURAS

Figura 1.1	Problema da ambiguidade citado por [AT04], onde a obtenção desta pose através da silhueta não é possível distinguir qual perna está dobrada. . . . .	23
Figura 1.2	Outro problema relativo à ambiguidade: as posições $x$ e $y$ no 2D das articulações da pessoa nas duas imagens (a) e (b) são semelhantes. Mas as poses no 3D são diferentes, pois os braços estão em posições opostas. A imagem (c) é um esqueleto 2D que poderia corresponder para as duas imagens (a) e (b). . . . .	23
Figura 2.1	Fluxograma do modelo proposto por Hu et al [HWLY09]. . . . .	27
Figura 2.2	Visão geral do modelo proposto por Ferrari, Jiménez e Zisserman [FMJZ08], <b>1. Detecção da parte superior do corpo:</b> (a) A pessoa detectada (retângulo interno) e a janela ampliada onde o processamento é aplicado (retângulo externo). <b>2. Foreground:</b> (b) sub-regiões para inicialização do Grabcut [RKB04]. (c) saída da região do primeiro plano ( <i>foreground</i> ) pelo Grabcut. <b>3. Análise:</b> (d) área $F$ a ser analisada (dilatada a partir de (c)) e bordas (e) dentro $F$ . (f) <i>Pictorial Structures</i> da posição das partes do corpo após a inferência baseada nas bordas. (g) <i>Pictorial Structures</i> após uma segunda inferência baseado nas bordas e na aparência. Figura obtida em [FMJZ08]. . . . .	28
Figura 2.3	Alguns resultados obtidos por Jiang [Jia11] . . . . .	29
Figura 2.4	Fluxograma do algoritmo apresentado por Malik et al. [MREM04] . . . . .	30
Figura 2.5	Alguns resultados apresentados por Malik et al, Figura obtida em [MREM04]. As colunas da imagem representam respectivamente: as imagens de entrada; os " <i>superpixels</i> " obtidos; os esqueletos encontrados; e as máscaras de segmentação. . . . .	30
Figura 2.6	Fluxograma do modelo proposto por Lee e Cohen. [LC06] . . . . .	31
Figura 2.7	Figura obtida em [LC06]. A primeira linha são as imagens de entrada e a segunda e terceira são os resultados obtidos pelo modelo de Lee [LC06]. O problema da ambiguidade devido a profundidade é percebido nas imagens das colunas (b) e (c). . . . .	32
Figura 2.8	Alguns resultados apresentados por [AT04]. . . . .	33
Figura 2.9	Resultados apresentados por Hua et al [HYW05]. . . . .	33
Figura 2.10	Trabalho baseado em uma base de dados de milhares de exemplares [Jia10].	34
Figura 2.11	Figuras processadas manualmente para o trabalho de Hua et al [HYW05]. . .	35
Figura 2.12	Processo em 3 partes proposto por Simo-Serra et al [SSRA*12]. . . . .	36
Figura 3.1	Fluxograma do modelo proposto. . . . .	38
Figura 3.2	O modelo do esqueleto utilizado na proposta. . . . .	40

Figura 3.3	Suposição 1: O topo da cabeça está sempre no plano da imagem. . . . .	43
Figura 3.4	Suposição 2: Os ossos da coluna devem estar com o sinal de delta Z em acordo. . . . .	44
Figura 3.5	Restrição 3: Define o ângulo entre os dois ombros, de acordo com os vetores: $\vec{v} = J_5 - J_2$ e $\vec{u} = J_9 - J_2$ . Se o ângulo interno entre $\vec{v}$ e $\vec{u}$ é menor que $165^\circ$ (de acordo com [NH94]), então o sinal de $\Delta Z_{J_9}$ é oposto a $\Delta Z_{J_5}$ . . . . .	45
Figura 3.6	Resultado após a aplicação da restrição dos joelhos, (a) Imagem de entrada, (b) Resultado sem restrição do ângulo entre coxa e a parte inferior da perna. (c) Resultado com restrição: evita a pose que a pessoa estaria com a coxa para trás e a parte inferior da perna para frente. . . . .	45
Figura 3.7	Cada osso do corpo foi modelado como um tronco de cone, a fim de calcular seu centro de massa. . . . .	46
Figura 3.8	Os pontos azuis representam o centro de massa individual de cada osso, o ponto magenta indica o ponto de apoio do corpo e o ponto verde mostra o centro de massa global do corpo. . . . .	46
Figura 3.9	Exemplo das imagens de entrada e as poses ordenadas. As poses são mostradas em pontos diferentes de visão a fim de ter uma noção melhor da postura 3D. Abaixo de cada pose, aparece a sua posição na classificação de conforto e a sua distância $dC$ . . . . .	48
Figura 3.10	Devido à semelhança das posições dos braços, a distância $dC$ é a mesma nas duas posturas. No primeiro caso, o braço direito do humano virtual está a frente do corpo e o braço esquerdo está para trás do corpo. E o oposto acontece na Figura da direita. Como consequência, as posturas de ambos têm a mesma medida de conforto. . . . .	49
Figura 3.11	A região $A$ , mostrada na imagem (b), é calculada pelas articulações $J_{15}$ , $J_{16}$ , $J_{19}$ e $J_{20}$ clicadas pelo usuário no processo de inicialização com a entrada da imagem (a). A imagem (c) mostra como a região $A$ é então dividida em dois quadrantes ( $Q_{sup}$ e $Q_{inf}$ ). Para cada quadrante é realizado o somatório dos pixels considerados sombras pela abordagem proposta por Guo, Daim e Hoiem [GDH11] ( <i>pixels</i> em vermelho e verde) da imagem (b). Caso $Q_{sup}$ tenha um resultado superior a $Q_{inf}$ é definido que a fonte de luz está na parte frontal da pessoa na imagem. . . . .	50
Figura 3.12	A região de interesse (retângulo em preto) é calculada usando a posição dos pontos clicados pelo usuário e as medidas antropométricas. . . . .	50
Figura 3.13	Fluxograma do algoritmo para a obtenção da posição dos braços baseado na luminosidade da imagem. . . . .	52

Figura 3.14	Resultado após a aplicação do modelo de luminosidade sobre a classificação da posturas baseadas no conforto. As imagens superiores apresentam a classificação das poses obtidas somente com o critério do conforto, onde a pose correta está na segunda posição. Nas imagens inferiores é apresentada a classificação com a aplicação do modelo de luminosidade onde a pose gerada correta passou para a primeira posição devido a sua menor penalização. . . .	54
Figura 4.1	Resultado da quantidade de conjuntos de possíveis poses obtidos pelo modelo.	56
Figura 4.2	Resultado onde a pose real não foi definida corretamente devido a influência da perspectiva sobre o modelo de Taylor [Tay00]. A imagem (a) é a imagem de entrada, vista de um ângulo de cima e do lado direito da pessoa na foto. A imagem (b) é um dos resultados gerados pelo modelo do Taylor que é suscetível à perspectiva. . . . .	57
Figura 4.3	O Histograma mostra para cada posição da classificação quantas poses 3D geradas pelo modelo estão de acordo com a pose real na fotografia. . . . .	57
Figura 4.4	A curva cumulativa apresenta as posições onde foram encontradas as posturas de acordo com o <i>ground truth</i> . . . . .	58
Figura 4.5	Resultados do esqueleto 3D obtido pelo modelo proposto classificado em 1º lugar. A primeira coluna representa a imagem de entrada, a segunda coluna é uma visão frontal do esqueleto gerado e a terceira coluna é a imagem de um ponto de vista lateral do esqueleto. A quarta coluna é um humano virtual gerado na mesma postura obtida pelo modelo. . . . .	59
Figura 4.6	Outros resultados obtidos pelo modelo proposto. As colunas estão na mesma ordem da Figura 4.5. . . . .	60
Figura 4.7	Resultado dos cliques realizados nas articulações de quatro pessoas a fim de verificar a sensibilidade do processo de inicialização do modelo proposto. . .	62
Figura 4.8	Exemplo das imagens de entrada e as poses ordenadas. As poses são mostradas em pontos diferentes de visão a fim de ter uma noção melhor da postura 3D. . . . .	63
Figura 4.9	Software fornecido pelo <i>Microsoft Kinect</i> SDK fazendo o rastreamento de usuários pelo seu sensor. . . . .	64
Figura 4.10	Resultados da localização das posturas na classificação de acordo com o <i>ground truth</i> obtido pela <i>Microsoft Kinect</i> ®. . . . .	65
Figura 5.1	Visão geral do modelo proposto. (a) Modelo de esqueleto proposto. (b) O resultado da segmentação. (c) As intersecções entre as partes do corpo e os pontos de intersecção (em vermelho). (d) Ilustração do resultado da estimação de auto-occlusão. (e) A pose 3D estimada. . . . .	69

Figura 5.2	(a) Auto-oclusão detectada. (b) Uma das possíveis poses incorretas gerada apenas com as restrições biomecânicas, pois o braço esquerdo da pessoa está por trás do torso. (c) A única pose obtida explorando a abordagem de auto-oclusão. . . . .	69
Figura 5.3	Número de poses 3D geradas por imagem, usando somente restrições biomecânicas (azul) e incluindo o modelo de auto-oclusão (vermelho). . . . .	70
Figura 5.4	A detecção de auto-oclusão e os resultados da recuperação da pose 3D. . . . .	71
Figura 6.1	Limitação do modelo: Mesmo que o centro de apoio da pessoa não esteja de acordo com a realidade (a), o sistema gera a pose correta (b). . . . .	76
Figura 6.2	Pose obtida através de uma perspectiva sobre a pessoa (a). A Figura (b) é o resultado gerado pelo modelo. . . . .	77
Figura 6.3	Limitações do modelo apresentado. A imagem obtida em perspectiva afeta o tamanho das pernas. . . . .	77
Figura 6.4	Outra limitação do modelo é quando a fonte de luz não está em frente a pessoa, o que gera informação incorreta para o modelo de luminosidade. . . . .	78
Figura 6.5	Limitação do modelo: A imagem foi invertida verticalmente e possui o mesma distância do conforto, o qual não é consistente com a realidade. Esquerda: Imagem original. Direita: Imagem invertida verticalmente. . . . .	78



## LISTA DE TABELAS

Tabela 3.1	Os ossos, articulações e os fatores de proporcionalidade $f_{lk}$ , $f_{wi}$ and $f_{wj}$ . . .	41
Tabela 4.1	O impacto obtido pela abordagem da luminosidade na classificação. A primeira coluna indica a imagem de entrada, a segunda coluna mostra a posição da postura de acordo com o <i>ground truth</i> na classificação gerada pelo modelo usando apenas abordagem de conforto. Na terceira coluna são as posições na classificação gerados pelo modelo com a abordagem de conforto junto com a abordagem de luminosidade. A quarta coluna é a situação comparativa de melhora ou não no posicionamento. . . . .	61
Tabela 4.2	Resultado da análise dos cliques no processo de inicialização do modelo por cinco usuários voluntários. As colunas QT_G de cada usuário é a quantidade de poses possíveis geradas pelo modelo através do seu modo de inicialização. As colunas Pos indicam a posição da pose correta na classificação de acordo com o <i>ground truth</i> . . . . .	62
Tabela 5.1	Resultados obtidos utilizando a abordagem de auto-oclusão. . . . .	73
Tabela 5.2	Mais resultados obtidos utilizando a abordagem de auto-oclusão. A última imagem não obteve a postura correta por motivo da projeção errada. . . . .	74



## LISTA DE SIGLAS

HOG	<i>Histogram of Oriented Gradients</i>
kd-tree	<i>k-dimensional tree</i>
MCMC	<i>Markov chain Monte Carlo</i>
PCA	<i>Principal Component Analysis</i>
PDM	<i>Point Distribution Model</i>
PS	<i>Pictorial Structures</i>
ROI	<i>Region of Interest</i>
RVM	<i>Relevance Vector Machine</i>
SDK	<i>Software Development Kit</i>



# SUMÁRIO

<b>Lista de Figuras</b>	9
<b>Lista de Tabelas</b>	13
<b>Lista de Siglas</b>	15
<b>1. INTRODUÇÃO</b>	19
1.1 Motivação . . . . .	20
1.2 Escopo da Tese . . . . .	21
1.3 Objetivos . . . . .	21
1.3.1 Objetivo geral . . . . .	21
1.3.2 Objetivos específicos . . . . .	21
1.4 Contribuições . . . . .	21
1.5 Considerações Iniciais - O problema da ambiguidade nas posturas 3D a partir do 2D	22
1.6 Metodologia . . . . .	22
1.7 Organização deste trabalho . . . . .	23
<b>2. REFERENCIAL TEÓRICO</b>	25
<b>3. MODELO DESENVOLVIDO</b>	37
3.1 Inicialização do Esqueleto . . . . .	38
3.2 Identificação da pose 3D . . . . .	40
3.2.1 Restrições Biomecânicas . . . . .	42
3.2.2 Ordenação das posturas geradas baseado na medida de conforto . . . . .	44
3.2.3 Abordagem através da Luminosidade . . . . .	48
<b>4. RESULTADOS</b>	55
4.1 Análise sobre a recuperação da pose 3D . . . . .	56
4.2 Análise da abordagem de Luminosidade . . . . .	58
4.3 Análise do método de inicialização do modelo . . . . .	58
4.4 Análise através de pesquisa . . . . .	63
4.5 Análise utilizando o sensor <i>Kinect</i> © para obter o <i>Ground Truth</i> . . . . .	64

5. APLICAÇÃO DO MODELO PROPOSTO	67
5.1 Estimativa da pose 3D utilizando informação de auto-oclusão . . . . .	67
5.1.1 Auto-oclusão . . . . .	67
5.1.2 Minimização da ambiguidade do modelo de Taylor usando a detecção de auto-oclusão . . . . .	68
5.1.3 Análise dos resultados utilizando a técnica de auto-oclusão . . . . .	71
6. CONSIDERAÇÕES FINAIS E PERSPECTIVAS	75
6.1 Deficiências do modelo . . . . .	75
6.2 Trabalhos Futuros . . . . .	77
Bibliografia	79

# 1. INTRODUÇÃO

Atualmente, com a popularização das câmeras de vídeo e das câmeras digitais, a necessidade de sistemas automáticos ou semi-automáticos de identificação e manipulação de imagens para consumidores leigos ou especialistas têm se tornado notório e imprescindível. O uso dessas câmeras ocorre nas mais variadas situações e vai desde a observação de multidões até a sua utilização em videoconferências, passando por um amplo uso em entretenimento. A área de estudo que engloba todas essas aplicações é denominada *Análise Visual do Movimento Humano* que utiliza conceitos de *Processamento Digital de Imagens* e *Visão Computacional* para a busca de soluções nos inúmeros desafios que ela apresenta. Diversos trabalhos do estado da arte na área ([MG01], [Gav99], [AC99]) estão disponíveis na bibliografia e proporcionam bons resumos dos modelos desenvolvidos, servindo como ponto de partida no desenvolvimento desta tese.

A Análise Visual do Movimento Humano é algo que se faz extremamente útil e necessário. É atualmente um importante campo de pesquisa e pode ser dividido em várias subáreas diferentes, que vão desde a detecção da pessoa na imagem, a descrição das suas atitudes em frente às câmeras, até a recuperação das suas poses ou posturas. Todas estas subáreas possuem seus próprios desafios e problemas em aberto. A simples recuperação de poses humanas possui um campo de estudos bem vasto, ela pode ser feita a partir de uma única imagem, onde não existe a informação de profundidade, ou então de sequências de vídeos, onde se tem informações de movimento, facilitando algumas vezes o processo. Os resultados também podem variar, obtendo-se a pose em duas ou três dimensões, dependendo da necessidade da solução do problema. Esta tese é focada na recuperação da pose humana em 3D a partir de uma única imagem bidimensional.

A obtenção da pose 3D a partir de uma imagem apresenta grandes desafios devido à perda intrínseca de informações durante o processo de criação da imagem 2D. No estudo do estado da arte ficam claras as dificuldades que podem ser encontradas pelas técnicas e abordagens para a solução deste problema. Há ainda diversas situações que contribuem para aumentar a sua dificuldade e precisam ser tratadas simultaneamente.

Provavelmente um dos desafios mais importantes na recuperação de uma pose 3D a partir de uma imagem bidimensional é a ambiguidade inerente do próprio 3D quando se possui somente um ponto de observação, o que se denomina visão monocular. Dada a relevância deste problema, na Seção 1.5 apresenta-se uma breve análise do problema da ambiguidade. Outro problema ainda mais evidente e desafiador é quando a pessoa está inserida em cenas do mundo real com várias outras pessoas, ocorrendo oclusões parciais ou sofrendo a influência de luminosidade e sombreamento ou até mesmo de ruídos dos meios que criaram a imagem. Outro desafio importante para a recuperação da pose a partir de imagens 2D é a complexidade das articulações humana e a sua aparência, ocorrendo problemas de ambiguidade e poses improváveis mas não impossíveis no mundo real. Além disso, os fundos complexos das cenas realistas podem dificultar e confundir as informações que auxiliem na obtenção da pose.

No estudo dos trabalhos atuais da área é possível verificar uma grande quantidade de abordagens dirigidas a obter a pose 3D a partir de uma única imagem bidimensional, cada uma com as suas vantagens e desvantagens e normalmente focadas em uma situação particular, em que o modelo obtém resultados para aquele contexto específico. Alguns modelos são aplicados somente para uma determinada situação, por exemplo, Malik et al [MREM04], desenvolveu uma abordagem para obtenção de poses de jogadores de *baseball*. Cada modelo tenta fazer uso de uma situação mais controlada possível a fim de facilitar o seu funcionamento e obter resultados melhores e mais expressivos.

Dito isto, é relevante e desafiador o desenvolvimento de um modelo mais genérico que a partir de uma imagem 2D permita a obtenção da pose humana tridimensional. Como desenvolver um modelo, que supere os desafios citados até o momento, e apresente resultados expressivos e coerentes, é a pergunta de pesquisa, bem como a principal aplicação desta tese.

## 1.1 Motivação

A recuperação da pose humana 3D a partir de uma única imagem estática é um problema desafiador e um campo de pesquisa ainda em aberto, apesar das inúmeras pesquisas existentes no estado da arte. A partir de uma imagem estática, as informações que podem ser usadas para inferir a pose completa e correta de uma pessoa na imagem estão sujeitas a diversas situações que influenciam e dificultam a solução do problema. Dentre algumas destas situações citam-se:

- Um espaço de busca extenso e complexo devida a articulação do corpo humano;
- A forma do corpo humano quando em situações incomuns;
- As oclusões parciais de partes do corpo humano;
- A perspectiva da imagem;
- As diferenças de iluminação e sombreamento; e
- Os vestuários de diversos tipos.

As informações obtidas da imagem podem ainda sofrer distorções devido a ruídos, detecções falsas e ambiguidades, além de não fornecer diretamente dados sobre a profundidade. Outro problema, são as articulações do corpo humano que tem um grande número de graus de liberdade e, portanto, a estimativa da pose envolve um espaço de busca de estados de grande complexidade e valor dimensional. Além disso, alguns desses graus de liberdade resultam em diversas poses ambíguas. Baseado nestes problemas, a apresentação de uma solução inovadora que minimize seus efeitos e que se consiga resultados coerentes e satisfatórios na obtenção da pose 3D a partir de uma imagem 2D é a principal motivação desta tese.



## 1.2 Escopo da Tese

O escopo principal desta tese é a recuperação de posturas humanas 3D a partir de imagens 2D com pessoas em poses frontais. Este escopo foi assim definido, a fim de delimitar o conjunto dos principais problemas a serem resolvidos para a recuperação das poses. Estes problemas serão descritos nas próximas seções.

## 1.3 Objetivos

### 1.3.1 Objetivo geral

O objetivo principal desta tese é o desenvolvimento de um novo modelo para a recuperação semi-automática das poses humanas em 3D a partir de imagens bidimensionais estáticas. Nesse estudo visa-se obter resultados de pesquisa que tornem possível inferir a pose correta e coerente de pessoas em imagens possibilitando o seu uso nas mais diferentes situações como por exemplo, em sistemas de segurança e vigilância, na obtenção de avatares personalizados para uso em jogos ou mesmo em videoconferências, dentre outros.

### 1.3.2 Objetivos específicos

A fim de alcançar o objetivo principal, os seguintes objetivos específicos foram propostos:

- Criar um modelo computacional que obtenha poses tridimensionais a partir de uma imagem 2D;
- Desenvolver técnicas a fim de minimizar o espaço de busca na geração da pose 3D;
- Desenvolver técnicas a fim de minimizar o problema da ambiguidade na geração da pose 3D;
- Desenvolver um protótipo para a coleta de resultados e prova de conceitos;

## 1.4 Contribuições

Esta tese apresentou as seguintes contribuições na área da visão computacional durante o seu desenvolvimento:

- Melhoramento da proposta apresentada por Taylor [Tay00] através da aplicação de restrições biomecânicas, conforto da postura e análise da luminosidade minimizando a ambiguidade de posturas geradas pelo modelo e reduzindo o espaço de busca para a solução da postura correta;
- Desenvolvimento de um conjunto de novas características baseadas no conforto da postura humana. Estas características podem especificar poses e serem aplicadas na classificação e organização das posturas;

- Desenvolvimento de uma aplicação para determinar poses humanas minimizando o problema de posturas ambíguas através de um detector de auto-occlusão.

### **1.5 Considerações Iniciais - O problema da ambiguidade nas posturas 3D a partir do 2D**

Um dos principais problemas na recuperação de poses humanas 3D é a ambiguidade. A ambiguidade gerada na obtenção de posturas tridimensionais a partir da imagem 2D é inerente à perda das informações de profundidade que este tipo de imagem possui [HYW05]. Muitos autores citam este problema [AT04, HYW05, LC06, WC09, Jia10, PJA\*12] como sendo uma das principais dificuldades na obtenção da postura 3D. No trabalho de Agarwal e Triggs [AT04] os autores definem como um problema intrínseco a estimação de poses 3D e na Figura 1.1 eles o caracterizam através das imagens (a) e (b) onde não é possível distinguir qual das pernas está dobrada se a pose fosse obtida através da silhueta da pessoa. O tratamento da ambiguidade pode se dar de várias formas. Wei et al [WC09], utilizam um conjunto de restrições biomecânicas aplicadas nos ângulos das articulações do corpo. Já os autores, Chen e Lee [CL92] usam limitações fisiológicas específicas para obter um pequeno conjunto de posturas corporais possíveis a partir de uma única imagem. Esses autores também afirmam que uma aplicação com várias fontes de conhecimento irá reduzir a ambiguidade e pode levar a um pequeno conjunto de prováveis candidatos da postura. No trabalho de Lee e Cohen [LC06], a ambiguidade é tratada através de uma abordagem que emprega o conhecimento adquirido através de cadeias de Markov. Moeslund et al [MHK06] citam técnicas de restrições cinemáticas e de movimentos para tratar a ambiguidade em trabalhos baseados na captura de movimento.

As Figuras 1.2(a) e 1.2(b) mostram duas imagens de uma criança em poses claramente distintas. Entretanto, a projeção bidimensional de um esqueleto sobre ambas as imagens seria praticamente equivalente pois as posições x e y das articulações serão muito semelhantes, como pode ser visto na Figura 1.2(c). Um sistema com objetivo de recuperar automaticamente a postura 3D da criança somente a partir da projeção do esqueleto da figura 1.2(c) fatalmente retornaria ambas as posturas como resultados possíveis, restando ao usuário decidir qual a postura de origem. Nesta tese o problema da ambiguidade é minimizado através de um conjunto de abordagens, como restrições biomecânicas, o conforto postural e a luminosidade da imagem. Estas abordagens serão tratadas no Capítulo 3.

### **1.6 Metodologia**

Os trabalhos para o desenvolvimento desta tese foram baseados inicialmente em um estudo do estado da arte para a recuperação de posturas humanas a partir de imagens bidimensionais, normalmente fotografias. Com este estudo procurou-se definir um escopo a fim de atingir os objetivos propostos. Inicialmente foi implementado um protótipo baseado no modelo de Taylor [Tay00], e a partir desse modelo, verificando suas deficiências e os problemas em aberto obtidos pelo estudo do estado da arte, procurou-se propor soluções para o seu aprimoramento. Estas soluções foram



Figura 1.1: Problema da ambiguidade citado por [AT04], onde a obtenção desta pose através da silhueta não é possível distinguir qual perna está dobrada.

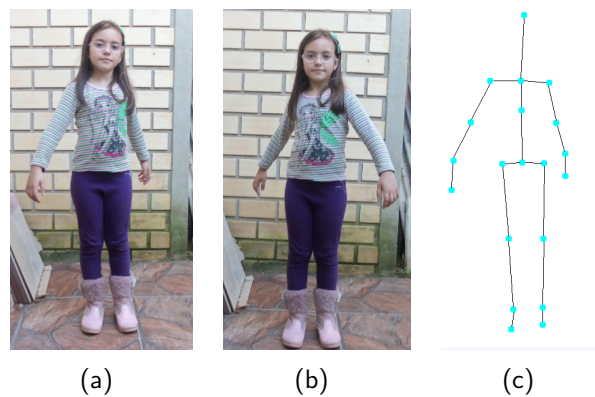


Figura 1.2: Outro problema relativo à ambiguidade: as posições  $x$  e  $y$  no 2D das articulações da pessoa nas duas imagens (a) e (b) são semelhantes. Mas as poses no 3D são diferentes, pois os braços estão em posições opostas. A imagem (c) é um esqueleto 2D que poderia corresponder para as duas imagens (a) e (b).

implementadas no protótipo inicial e através de testes empíricos foi verificado o resultado dessas soluções aos problemas existentes. As abordagens propostas foram descritas e os resultados dos experimentos realizados foram apresentados nesta tese.

## 1.7 Organização deste trabalho

Esta tese está organizada da seguinte forma: O Capítulo 2 apresenta o referencial teórico contendo o estado da arte das principais técnicas desenvolvidas atualmente para a recuperação de poses humanas além de apresentar duas taxonomias utilizadas pelos trabalhos de Hu et al [HWLY09] e pelos autores Agarwal e Triggs [AT06b], para a classificação dos modelos de recuperação de posturas humanas.

O Capítulo 3, apresenta o modelo científico desenvolvido. Neste capítulo são apresentadas as abordagens para a geração de possíveis poses humanas 3D baseadas em imagens bidimensionais, além de técnicas para a minimização do espaço de busca das poses e a diminuição do problema das

poses ambíguas.

No Capítulo 4, são apresentadas as análises desenvolvidas a fim de verificar a efetividade assim como as deficiências do modelo. Este capítulo também mostra os resultados obtidos.

O Capítulo 5 mostra aplicações desenvolvidas durante o período desta tese empregando o modelo proposto. A primeira aplicação é uma abordagem para a recuperação da pose humana 3D a partir de imagens 2D, minimizando o problema de posturas ambíguas através de uma abordagem que explora o conceito de auto-occlusão entre as partes do corpo da pessoa na imagem. A segunda aplicação é uma proposta para a recuperação de um pequeno conjunto de imagens (fotografias) a partir de uma base de dados de imagens muito extensa empregando informações de alto nível em relação a postura, como por exemplo, informações se a pessoa está com os braços erguidos, correndo ou dançando, etc. E por fim, o Capítulo 6 apresenta as conclusões e as perspectivas futuras desta tese.

## 2. REFERENCIAL TEÓRICO

A obtenção da pose de pessoas em imagens por meios computacionais é uma área de pesquisa bastante relevante devido à grande variedade de aplicações possíveis. Elgammal e Lee [EL04] citam como algumas dessas aplicações: a captura de movimentos, interfaces visuais, a vigilância visual e o reconhecimento de gestos. Desta forma, muitos esforços tem sido empregados para a obtenção das poses a partir de imagens ou vídeos, mas é possível perceber que ainda não existe uma taxonomia bem definida na literatura.

Hu et al [HWLY09] classificam a área inicialmente em duas categorias principais: Os modelos que obtém a postura humana a partir de sequências de vídeo e os modelos que recuperam a pose a partir de imagens estáticas, como por exemplo, fotografias. Essa classificação é importante porque os modelos baseados em sequências de vídeos normalmente conseguem obter determinados dados, como por exemplo, informações de movimento e informações sobre o *background* e *foreground* da imagem que são informações que facilitam o processo de obtenção da pose e que em imagens estáticas são difíceis, senão impossíveis de serem obtidas.

Os trabalhos [AT06a], [WL06], [LN07] e [MBR06] enquadram-se bem na categoria baseada em vídeos, que não é o escopo deste trabalho. Já a categoria que utiliza os modelos baseados em imagens estáticas é dividida em três tipos de abordagens:

1. Abordagens baseadas em casamento de padrões - *Matching-based* - ( [AT06b], [MM06]), esses métodos estimam a pose humana pela comparação das características do corpo humano obtidas na imagem de entrada com uma grande base de imagens classificadas. Esses métodos normalmente apresentam dificuldades quando se tem um fundo (*background*) muito ruidoso e complexo. Outra grande desvantagem que estes métodos apresentam é que eles necessitam de um grande conjunto de imagens (base de dados) para treinamento, com muitas poses, fundos e características diferentes;
2. Abordagens baseadas em partes - *Part-based* - ( [MREM04], [HYW05], [FMJZ08], [ARS09]), esses métodos tem por finalidade superar as deficiências apresentadas pelos métodos baseados em casamento de padrões. Eles procuram obter a postura humana em imagens pela detecção de possíveis candidatos de cada parte do corpo (tais como face, torso e membros) e fazer uma inferência da melhor montagem a partir de uma configuração de restrições pré-determinadas. De acordo com Hu et al [HWLY09], estes modelos necessitam bem menos dados de treinamento e são menos suscetíveis a variações de posturas na imagem. Seu tempo de execução depende dos detectores de partes e eles também podem tratar do problema de estilos de roupas das pessoas; e
3. Abordagens baseadas em Modelos - *Model-based* - ( [Tay00], [PC04], [LC06], [JDJ\*10]), esse tipo de abordagem, diferentemente das outras já citadas, procuram gerar inicialmente um

grande número de poses hipotéticas, através da mudança de parâmetros do modelo humano. Então a partir desse conjunto gerado, tenta-se obter a postura correta pela minimização de erros de projeção entre as poses hipotéticas e a imagem. Tais métodos são menos sensíveis aos detectores de partes e podem alcançar poses precisas com uma inicialização adequada. Nesta classificação é onde melhor se enquadra o modelo apresentado nesta tese.

Esta classificação proposta por Hu et al [HWLY09] não faz distinção quanto aos resultados obtidos pelos modelos, sendo que alguns obtêm como resultados somente posturas em duas dimensões ( [FMJZ08], [ARS09]) e outros modelos obtêm as posturas em três dimensões ( [AT06b], [MM06]).

Também na literatura é possível encontrar outras formas de classificação, como a apresentada por Agarwal e Triggs [AT06b]. De acordo com eles, existem duas escolas principais para obter as poses de pessoas em imagens: as *abordagens baseadas em modelos* e as *abordagens baseadas em aprendizagem*. Na primeira escola, os autores pressupõem um conhecimento explícito de um modelo paramétrico do corpo humano e as poses são estimadas diretamente através da cinemática inversa que fornece muitas soluções possíveis. Como exemplos, citam-se os trabalhos de [Tay00], [FMJZ08], [HWLY09], [EF09] e [Jia11].

Em [Tay00], Taylor desenvolveu um modelo parametrizado baseado em projeção ortográfica escalada. Este modelo é fundamentado em três premissas. Inicialmente, as articulações do esqueleto são fornecidas ao modelo. Segundo, o tamanho relativo dos segmentos que formam o esqueleto (ossos) também são conhecidos a priori. Em terceiro lugar, a relação entre as posições das articulações no espaço e suas projeções para a imagem pode ser modelada como uma projeção ortográfica escalada. É importante salientar que o modelo apresentado nesta proposta é baseado neste trabalho [Tay00], e que maiores detalhes sobre este modelo serão vistos no Capítulo 3.

No trabalho apresentado por Hu et al [HWLY09], as poses são obtidas somente para a parte superior do corpo. O modelo proposto por estes autores é dividido em três estágios: detecção das partes observáveis, inicialização das articulações, e inferência da pose. No primeiro estágio, a face, a pele do rosto e o torso são inicialmente detectados conforme ilustram as Figuras 2.1(a), (b) e (c). A face é obtida pelo detector de face baseado no AdaBoost [VJ04] e a pele é segmentada usando um método randômico de Markov. Em seguida é detectado o torso, que é uma região retangular obtida a partir da face. No segundo estágio é feita a inicialização das articulações de acordo com as observações e algumas configurações pré-estabelecidas, como por exemplo: os ombros, ancas e o pescoço são obtidos diretamente na região do torso, as mãos, são obtidas pela cor da pele e os cotovelos através de restrições heurísticas, Figura 2.1 (g) e (h). Finalmente, o processo de Monte Carlo via Cadeias de Markov (MCMC) é empregado para determinar a postura final. Este modelo obtém somente a pose 2D da pessoa na imagem, Figura 2.1 (i). A Figura 2.1 apresenta o fluxograma da abordagem proposta por Hu et al [HWLY09].

Em [EF09] é apresentada uma nova abordagem para estimar a aparência das partes do corpo usando estruturas pictóricas (*Pictorial Structures - PS*). PS são modelos probabilísticos onde os objetos são feitos de partes arranjadas por potenciais pares obtidos por informações a priori sobre suas relações espaciais (por exemplo, restrições cinemáticas). As abordagens utilizando técnicas

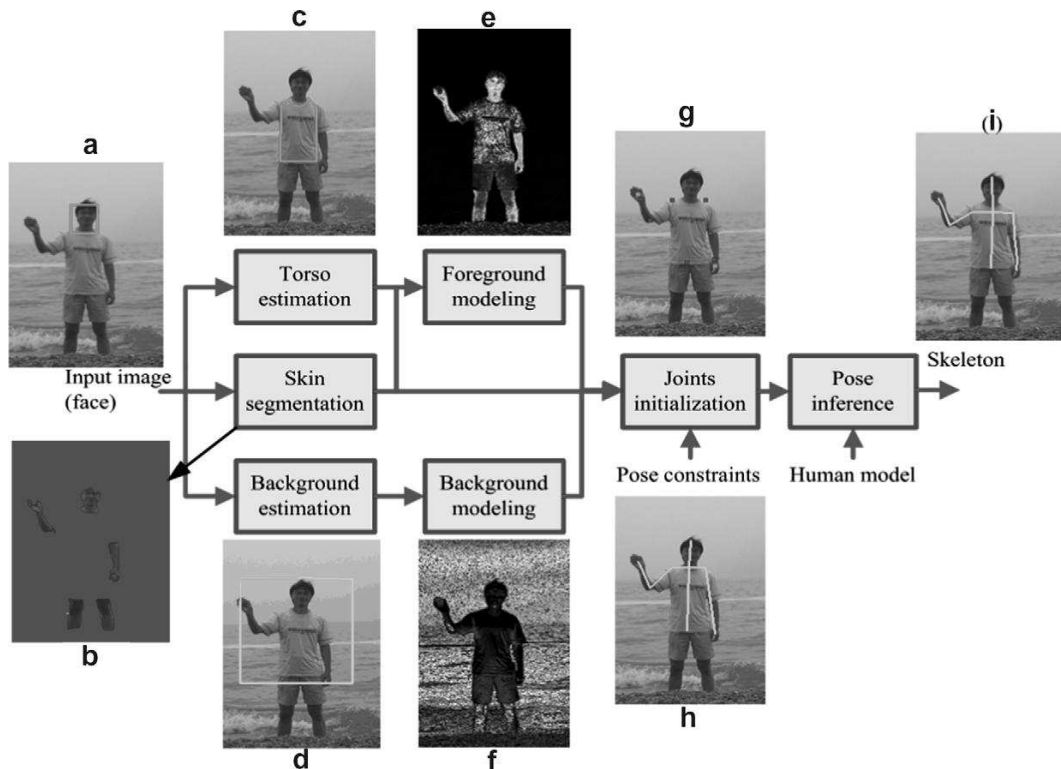


Figura 2.1: Fluxograma do modelo proposto por Hu et al [HWLY09].

de PS tem sido bastante empregadas na literatura recentemente. O trabalho apresentado por Ferrari [FMJZ08] também utiliza *Pictorial Structures* para a detecção de poses humanas em imagens. O modelo apresentado obtém a pose somente na parte superior do corpo, semelhante ao trabalho de Hu et al [HWLY09]. A fase inicial do modelo tem por finalidade determinar a parte superior do corpo e a sua aparência, obtendo somente a localização e a escala aproximada da pessoa. As duas próximas etapas utilizam uma estrutura pictórica que descreve a configuração espacial de todas as partes do corpo e sua aparência. A Figura 2.2 apresenta uma visão global dos passos do modelo durante todo o processo.

Outra abordagem baseada em modelos é apresentada por Jiang [Jia11], que propõe um esquema denominado de *Consistent Max-Covering* consistente para obter a pose humana em imagens. O *Consistent Max-Covering* obtém a pose através de polígonos que representam as partes do corpo humano e tentam cobrir o máximo possível do *foreground* representado pela silhueta humana. Para esta tarefa são utilizadas características locais da imagem e restrições de cor e formas do corpo humano. O modelo pode ser dividido basicamente em 4 passos:

1. Passo 1 - Modelo do corpo e detecção das partes: É utilizado um modelo contendo 10 partes do corpo humano, essas partes são descritas através de retângulos. Este modelo é colocado sobre o *foreground* da imagem que já tinha sido obtido através de uma subtração de fundo ou uma segmentação de cor. Inicialmente, prováveis partes do corpo são cobertas com as partes do modelo e ajustadas pelo método de casamento de Chamfer [Gav07];
2. Passo 2 - *Consistent Max-covering*: Cada provável parte do modelo do corpo cobre alguns

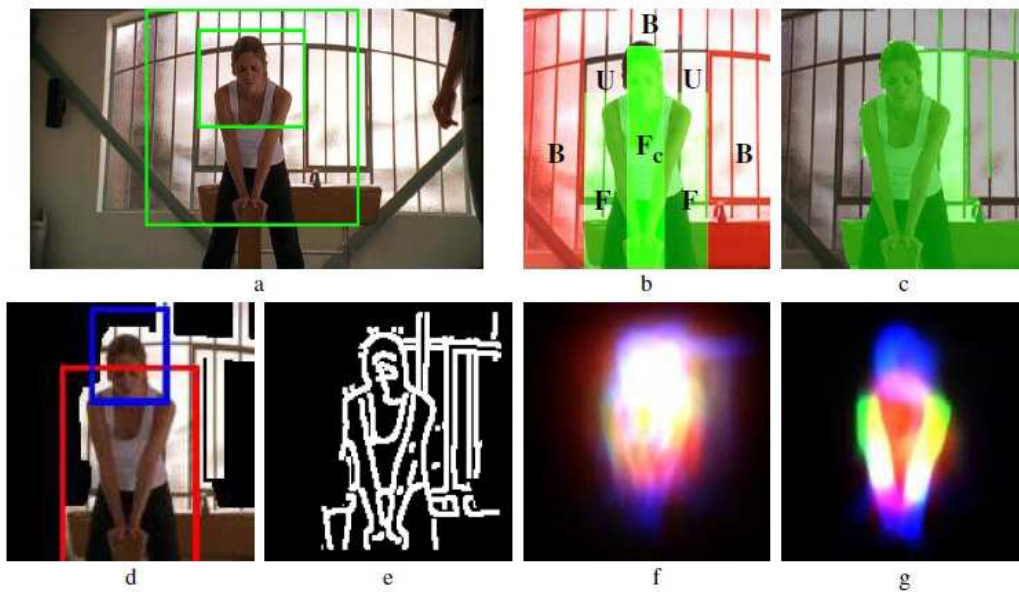


Figura 2.2: Visão geral do modelo proposto por Ferrari, Jiménez e Zisserman [FMJZ08], **1. Detecção da parte superior do corpo:** (a) A pessoa detectada (retângulo interno) e a janela ampliada onde o processamento é aplicado (retângulo externo). **2. Foreground:** (b) sub-regiões para inicialização do Grabcut [RKB04]. (c) saída da região do primeiro plano (*foreground*) pelo Grabcut. **3. Análise:** (d) área  $F$  a ser analisada (dilatada a partir de (c)) e bordas (e) dentro  $F$ . (f) *Pictorial Structures* da posição das partes do corpo após a inferência baseada nas bordas. (g) *Pictorial Structures* após uma segunda inferência baseado nas bordas e na aparência. Figura obtida em [FMJZ08].

pixels do *foreground* da imagem. O *Max-Covering* é formulado como um problema de otimização que procura cobrir o máximo destes potenciais *pixels* que pertence a determinada parte do corpo. Essa otimização é representada através de uma equação composta pelo possível *pixel*, o custo de combinar as partes do corpo que caracteriza a imagem local, o grau de configuração da parte do corpo na sequência do plano do corpo humano e uma penalização pela diferença de cor das partes do corpo simétricas: se as partes simétricas, por exemplo, parte superior do braço, tem diferença de cor grande, a penalização também é grande;

3. Passo 3 - Formulação Linear: Neste passo é realizada a linearização da equação que formula o *max-covering* consistente, obtido através de uma programação linear inteira mista;
4. Passo 4 - Exclusão e Cobertura: Finalizando o modelo, restrições de exclusão são empregadas a fim de penalizar partes do corpo que estiverem sobrepostas. Estas restrições evitam que certas partes do corpo ocupem a mesma localização espacial;

De acordo com os autores o método proposto apresenta correlações entre múltiplas partes do corpo e melhora a robustez da estimativa para representar movimentos complexos. O trabalho também propõe uma formulação linear e um método de relaxamento eficiente. Experimentos em sequências de vídeo mostraram que o método proposto é robusto e eficiente na estimativa de representar a pose humana. Na Figura 2.3, é possível observar alguns resultados obtidos.



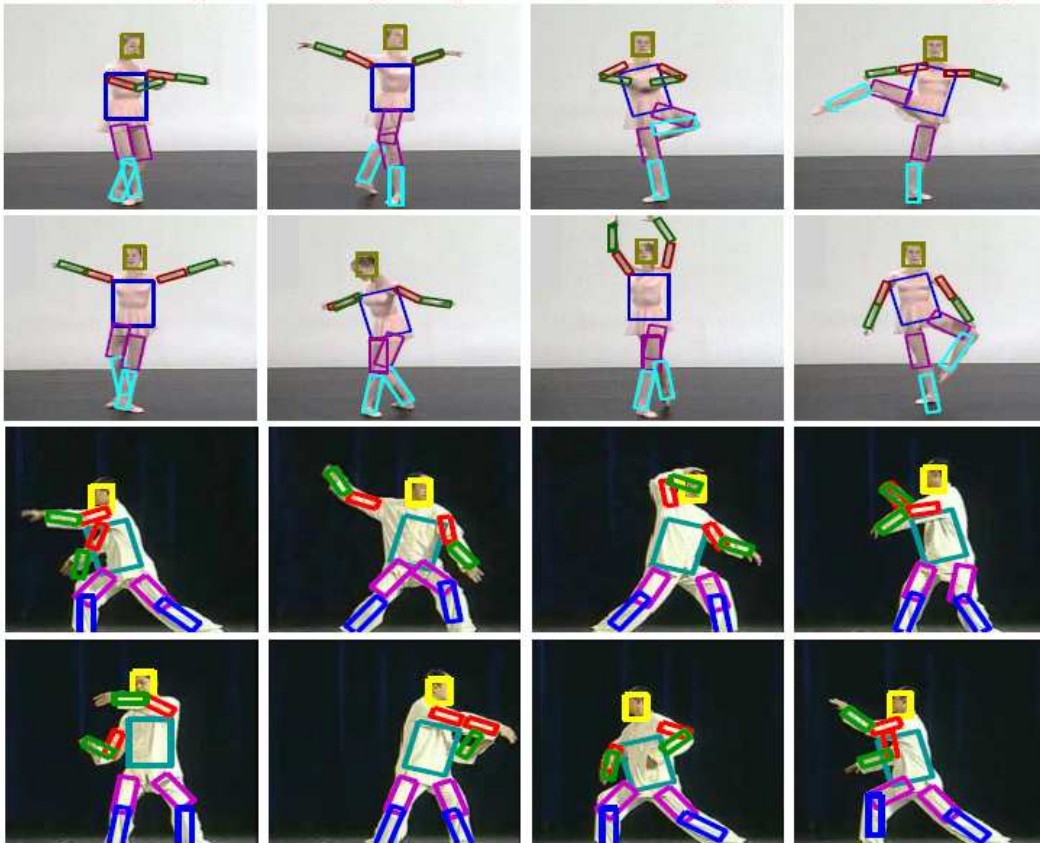


Figura 2.3: Alguns resultados obtidos por Jiang [Jia11]

Os trabalhos baseados em modelos estudados até aqui apresentam resultados muito promissores, eficientes e robustos. Sua principal vantagem é que normalmente são de simples implementação e baixo custo computacional. Mas é possível observar que as soluções obtidas focam especialmente em esqueletos 2D, já que o problema da ambiguidade inerente ao obter um sistema 3D a partir do 2D está ainda em aberto e é de difícil solução. Taylor [Tay00] apresentou uma solução baseada em restrições mas não faz uma definição de quais restrições para o corpo humano, já que seu modelo é mais genérico para corpos articulados, englobando robótica e etc.

A outra escola proposta por Agarwal e Triggs [AT06b] está relacionada com abordagens *baseadas em aprendizagem*, este tipo de abordagem evita a inicialização explícita e tem por objetivo tirar proveito de conjuntos de dados para procura em imagens de treinamento que correspondam a imagem de entrada. Estes conjuntos de dados são basicamente compostos por um grande conjunto de malhas 3D de humanoides obtidos através de *scanners*. Dentre os trabalhos que se enquadram nesta categoria citam-se [MREM04], [AT04], [EL04], [HYW05], [LC06], [AT06b], [GQ06], [GWBB09] e [Jia10]. Alguns destes serão apresentados a seguir.

Malik et al. [MREM04] apresentam um trabalho que aborda o problema de localização de posições das articulações de uma pessoa em uma imagem estática, e usam essa informação para estimar a configuração e postura do corpo em 3D. A abordagem básica é armazenar uma série de exemplos de imagens 2D de corpos humanos em diferentes configurações e pontos de vista da câmera. Em cada uma dessas imagens armazenadas, a localização das articulações (cotovelos, joelhos, etc.) foi

manualmente marcada e rotulada para ser usada posteriormente. A imagem de entrada é então testada com cada uma das amostras armazenadas, e o exemplo mais apropriado é encontrado.

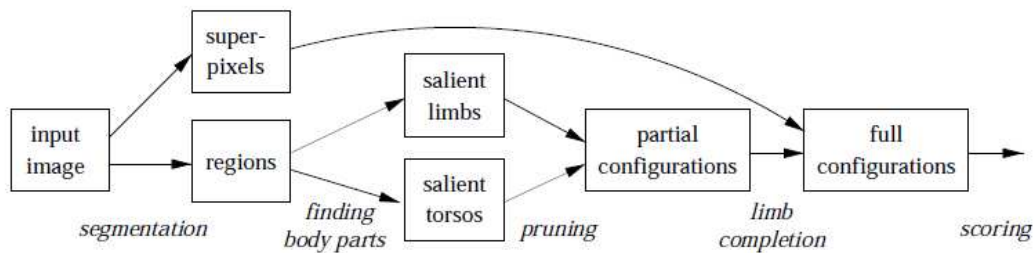


Figura 2.4: Fluxograma do algoritmo apresentado por Malik et al. [MREM04]

A Figura 2.4 apresenta o fluxograma do algoritmo apresentado por Malik et al. [MREM04]. Inicialmente a imagem de entrada é segmentada em “*superpixels*” que são regiões relativamente grandes ou segmentos. Então são detectados os membros superiores e inferiores nestes segmentos. Simultaneamente, são encontradas as potenciais posições da cabeça e do tronco. Ambos os módulos retornam um “*shortlist*” dos melhores candidatos classificados. Em seguida, são combinadas essas partes do corpo em configurações parciais e eliminadas as configurações impossíveis, impondo restrições globais como a escala relativa e simetria nas roupas. Na fase final são completadas as configurações parciais, pela pesquisa de combinações no espaço de *superpixels* para recuperar as configurações do corpo inteiro. A Figura 2.5 apresenta alguns resultados obtidos pelos autores.

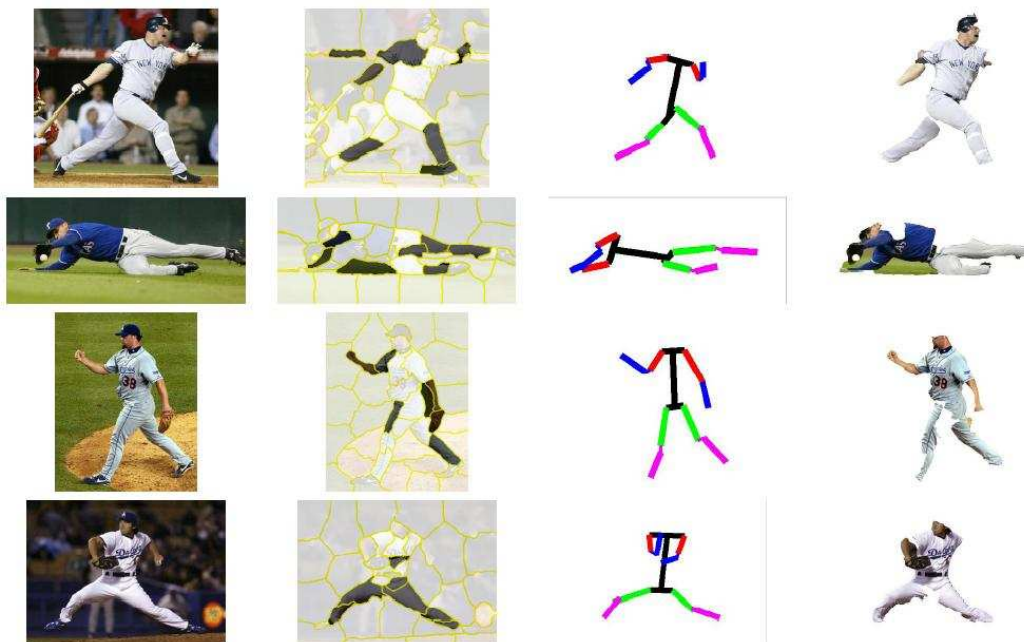


Figura 2.5: Alguns resultados apresentados por Malik et al, Figura obtida em [MREM04]. As colunas da imagem representam respectivamente: as imagens de entrada; os “*superpixels*” obtidos; os esqueletos encontrados; e as máscaras de segmentação.

O trabalho de Lee e Cohen [LC06] propõe a utilização de uma abordagem adaptativa, onde um modelo humano é usado para sintetizar as regiões da imagem correspondente a formas humanas

(dada uma hipótese) e, assim, separar a pessoa do fundo. Para a geração das hipóteses, o método de Monte Carlo via Cadeia de Markov é usado.

Uma visão geral do modelo de Lee e Cohen é apresentada na Figura 2.6. Inicialmente, uma abordagem hierárquica é usada para extrair um conjunto de recursos correspondentes ao contorno do corpo e as suas partes. Estas características foram selecionadas porque são discriminativas e estão diretamente relacionados com as características visíveis na imagem. O modelo descreve uma forma de extrair, pesar e excluir essas características. Primeiro, um *blob* (região predefinida) do primeiro plano (*foreground*) da imagem é extraído pela subtração de fundo (*background subtraction*). Esses *blobs* são rastreados como elipses que representam tanto o corpo inteiro ou somente a parte superior do corpo (Fase A, Figura 2.6). O método de rastreamento incorpora aparência e aprende a lidar com inter-ocluções entre as pessoas. O rastreamento robusto destas elipses fornece estimativas grosseiras da posição do corpo, tamanho e orientação. Características das partes individuais do corpo são detectadas em uma abordagem hierárquica (Fase B, Figura 2.6). Para cada parte, vários candidatos são gerados e um processo de otimização é usado para alinhar os candidatos as características da imagem com precisão. Estes candidatos são avaliados por uma distribuição de probabilidade posterior local aproximada, e os candidatos com baixa confiança são removidos.

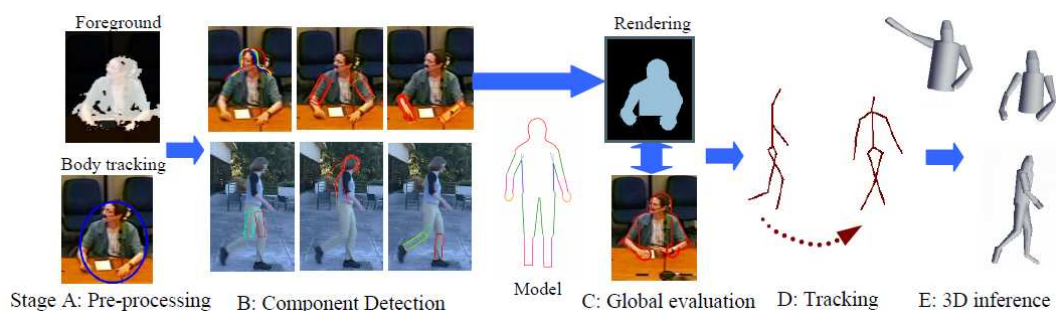


Figura 2.6: Fluxograma do modelo proposto por Lee e Cohen. [LC06]

Os candidatos para as diferentes partes do corpo são combinados para formar um conjunto de hipóteses (Fase C, Figura 2.6). Cada hipótese é avaliada por um cálculo de uma medida global de possibilidade com base em uma distribuição de probabilidade computada a partir de previsões das bordas do corpo. Isso explica o efeito da auto-oclução, bem como o efeito da perspectiva. Com o uso de múltiplas hipóteses de pose para cada quadro da imagem, uma pose ótima da trajetória da sequência é extraída usando programação dinâmica no Estágio D da Figura 2.6. Na Fase E, são estabelecidas restrições físicas de um modelo humano articulado 3D para estimar a pose tridimensional usando uma amostragem baseada no método MCMC.

De acordo com os autores [LC06], os principais pontos fortes deste método incluem as habilidades para lidar com o corpo com formas diferentes, sem inicialização manual e supera significativamente fundos confusos e desordenados e a auto-oclução. Os resultados dos experimentos indicaram boa estimativa da pose, estes resultados podem ser observados na Figura 2.7.

Agarwal e Triggs [AT04] descrevem um método de recuperação da pose do corpo humano em 3D baseado em aprendizagem de imagens únicas e sequências de imagens monoculares. A abordagem

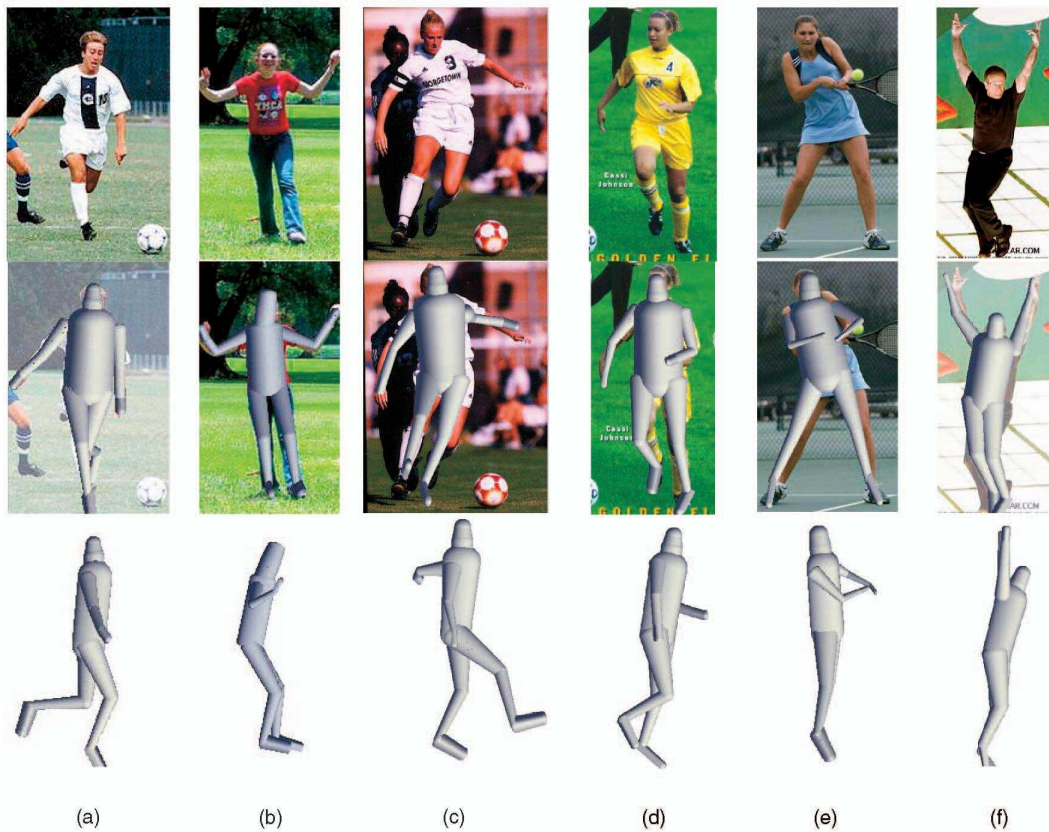


Figura 2.7: Figura obtida em [LC06]. A primeira linha são as imagens de entrada e a segunda e terceira são os resultados obtidos pelo modelo de Lee [LC06]. O problema da ambiguidade devido a profundidade é percebido nas imagens das colunas (b) e (c).

não exige nem um modelo de corpo, nem rotulagem prévia explícita de partes do corpo na imagem. Ela emprega a silhueta e regressores RVM (*Relevance Vector Machine*). RVM é uma técnica de aprendizado de máquina (*machine learning*) que utiliza a inferência bayesiana para obter soluções parcimoniosas para a regressão e classificação. De acordo com os autores, para obter realismo e uma boa generalização no que diz respeito a pontos de visão, os regressores foram treinados utilizando imagens resintetizadas a partir de *motion capture* de humanos reais, e foram testados tanto quantitativamente quanto qualitativamente em sequência de imagens reais. A Figura 2.8 apresenta alguns resultados do modelo proposto.

Outro trabalho que faz parte da escola baseada na aprendizagem é apresentado por [OURHJR04]. O trabalho é um modelo estatístico para a detecção e rastreamento de silhueta humana e da estrutura 3D do esqueleto correspondente em sequências de vídeo. Os resultados são obtidos utilizando técnicas de Análise de Componentes Principais (PCA). O problema da não linearização do PCA é parcialmente resolvido pela aplicação de um PDM (*point distribution model*) diferente, dependendo da pose estimada.

Hua et al [HYW05] propôs uma formulação estatística para estimar a pose humana 2D a partir de uma única imagem. A configuração do corpo humano é modelada por uma rede de Markov e a estimativa do problema é inferir parâmetros de pose da imagem, tais como aparência, forma, cor,



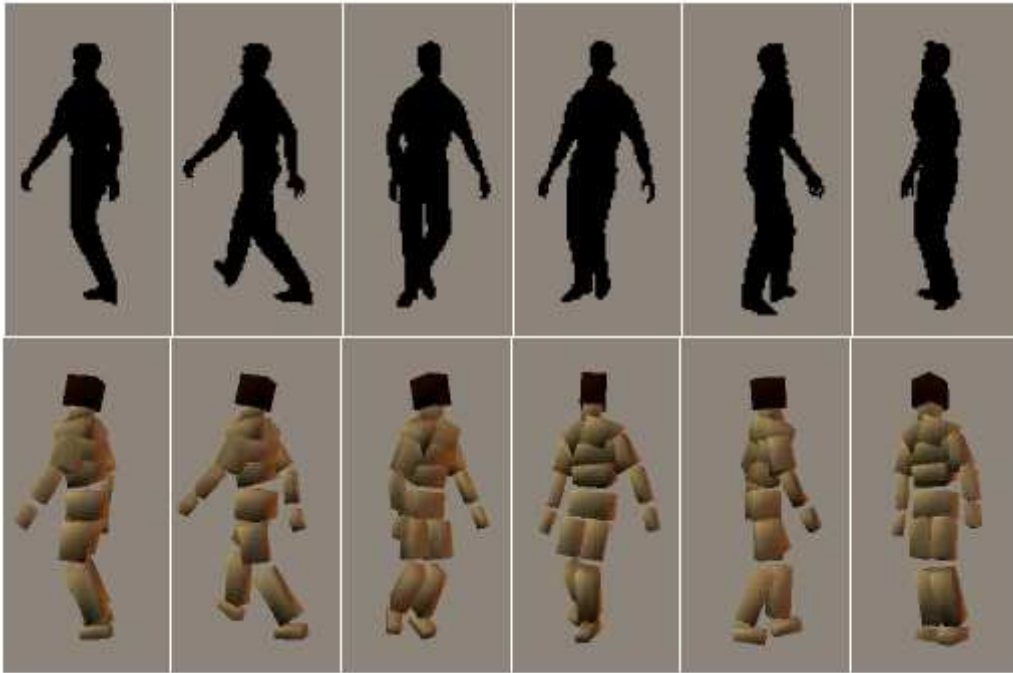


Figura 2.8: Alguns resultados apresentados por [AT04].

borda, etc. A partir de um conjunto de imagens rotuladas a mão, determinando o conhecimento sobre partes do corpo 2D é armazenado previamente pelo aprendizado das suas representações em baixo nível e pela inferência dos parâmetros da pose. Para a inferência é empregado o método de Monte Carlo. O modelo é treinado para imagens de jogadores de futebol conforme pode ser visualizados os resultados na Figura 2.9

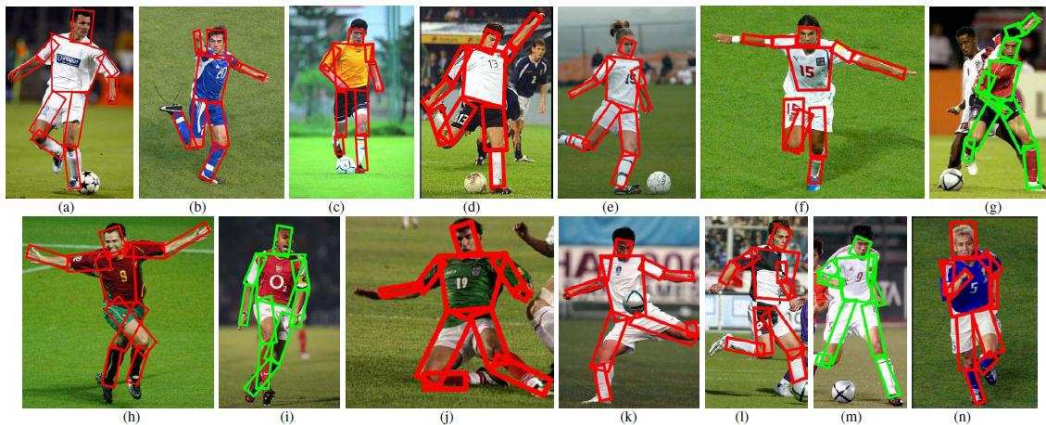


Figura 2.9: Resultados apresentados por Hua et al [HYW05].

Um trabalho mais recente é apresentado por Jiang [Jia10]. Este método estima poses 3D humanas baseadas em imagens únicas usando apenas as articulações correspondentes. Devido à ambiguidade inerente da profundidade, estimar poses 3D de uma visão monocular é um problema desafiador. Este método resolve o problema através de uma pesquisa em uma base de dados de milhões de exemplares de poses. Comparado com os modelos tradicionais paramétricos, de acordo

com os autores, esse método é capaz de lidar com um grande banco de dados de poses, evita ajustes de parâmetros, é fácil de treinar e é eficaz para representar reconstruções 3D complexas. O método proposto estima a parte superior do corpo e a parte inferior do corpo sequencialmente, o que implicitamente divide ao quadrado o tamanho do banco de dados exemplar e permite reconstruir naturalmente poses de forma eficiente. A implementação é baseada no método *kd-tree* (*k-dimensional tree*) possibilitando desempenho em tempo real. Os experimentos em uma variedade de imagens mostram que o método proposto é eficiente e eficaz. A Figura 2.10 apresenta alguns resultados do método proposto por Jiang [Jia10].

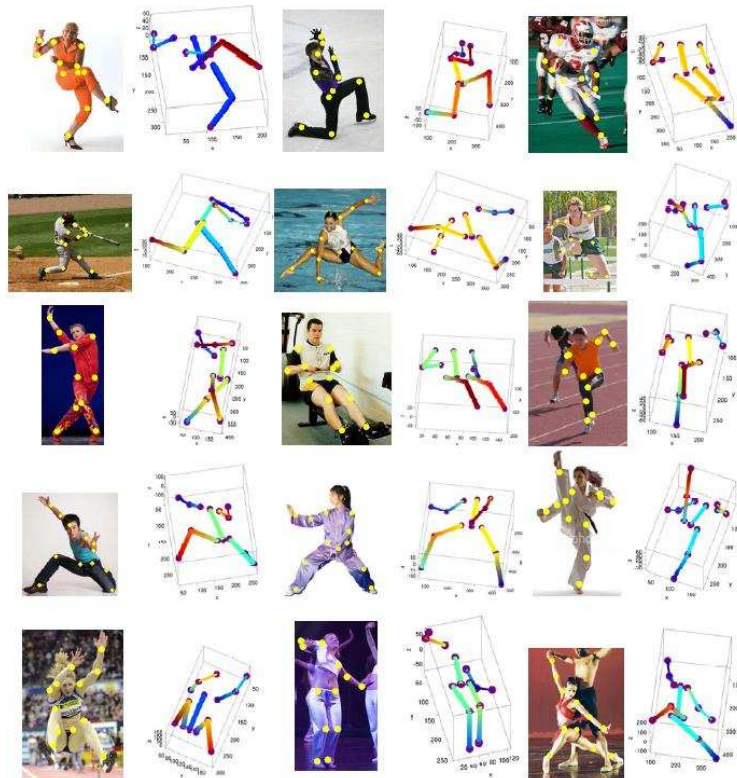


Figura 2.10: Trabalho baseado em uma base de dados de milhares de exemplares [Jia10].

Apesar dos resultados de alta qualidade alcançados por abordagens do tipo de aprendizagem, existem algumas desvantagens que devem ser discutidas. Uma das principais é a necessidade da criação de uma base de dados de poses humanas extensa, já que as possibilidades de poses do ser humano superam os milhares. Outro problema é que normalmente essas bases de dados só podem ser criadas utilizando dispositivos do tipo *scanners* 3D de alta resolução, que muitas vezes estão disponíveis somente em grandes empresas ou em centros de pesquisas avançados, além de serem muito custosos. Outra desvantagem é que essas bases de dados são bem específicas para um determinado tipo de situação, por exemplo, o trabalho apresentado por Hua et al [HYW05] é modelado para jogadores de futebol, que possuem certos padrões de uniformes e poses. Para ser aplicado em outra situação a base de dados e o treinamento deve ser criados novamente. Outra dificuldade é que essas imagens da base de dados normalmente são rotuladas e descritas manualmente, sendo um trabalho demorado, custoso e tedioso além de ser muito subjetivo. A

Figura 2.11 apresenta algumas imagens informadas manualmente dentre um conjunto de 50 imagens usadas no modelo de Hua et al [HYW05].



Figura 2.11: Figuras processadas manualmente para o trabalho de Hua et al [HYW05].

Como trabalhos mais recentes do estado da arte citam-se os modelos propostos por Yi e Ramanan [YR11], o de Simo-Serra et al [SSRA\*12] e o trabalho de Radwan [RDJG12]. Simo-Serra et al [SSRA\*12] propuseram uma nova abordagem para estimar poses humanas 3D mesmo quando as imagens são ruidosas. Sua abordagem é dividida em três partes principais: Detecção das partes em 2D, exploração estocástica das hipóteses ambíguas e a desambiguação. Foi proposta uma estratégia de amostragem estocástica para propagar o ruído do plano de imagem para o espaço de forma. Isto fornece um conjunto de formas 3D ambíguas, que são praticamente indistinguíveis de suas projeções de imagem. A desambiguação é obtida através da imposição de restrições cinemáticas que garantem que a pose resultante é uma possível postura humana 3D. O método foi validado em uma variedade de situações em que os detectores 2D do estado da arte obtiveram tanto estimativas imprecisas ou parcialmente, faltando partes do corpo. A Figura 2.12, mostra de forma geral as 3 principais partes do modelo proposto.

O trabalho de Radwan [RDJG12] é um *framework* para o tratamento de oclusões parciais e uma melhor estimativa da pose 2D. Seu trabalho propõe um *framework* baseado em Estrutura Pictorial - Structure Pictorial (PS), o que resulta em uma melhor estimativa de pose no caso de auto-occlusão em imagens sem restrições. Estimativa de poses de corpos articulados baseado em um *framework* PS atualmente tem atraído muita atenção no desenvolvimento em uma grande variedade de aplicações, tais como segurança automotiva (detecção de pedestres), vigilância, busca por posturas e indexação de vídeo. Modelos PS representam um objeto como um grafo, onde cada nó representa uma parte do corpo e as arestas entre nós codificam as restrições cinemáticas entre cada par ligado das partes do corpo.

Neste capítulo foram apresentados alguns dos trabalhos mais inovadores e recentes no estado da arte com o objetivo de obter posturas humanas a partir de imagens ou vídeos. É possível observar que o foco dos trabalhos é centrado na obtenção de posturas a partir de vídeos ou processando imagens, e que na sua maioria os resultados são somente esqueletos bidimensionais. Quando os resultados focam em esqueletos 3D, são usadas abordagens para situações muito específicas ou

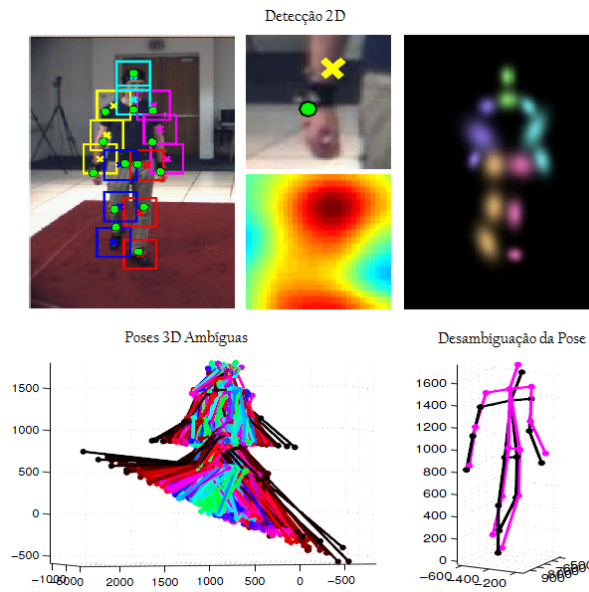


Figura 2.12: Processo em 3 partes proposto por Simo-Serra et al [SSRA\*12].

bastante controladas. Desta forma, uma nova abordagem está sendo proposta baseada em modelos. Essa abordagem será descrita no próximo capítulo.



### 3. MODELO DESENVOLVIDO

O problema de estimar a pose em 3D de uma pessoa a partir de imagens tem recebido uma atenção especial na literatura de visão computacional conforme discutido no Capítulo 2. Isto em parte deve-se ao fato de que as soluções para este problema podem ser empregadas em uma ampla gama de aplicações. Pode-se observar através do estudo realizado no estado da arte que a maioria das pesquisas nesta área têm-se centrado sobre o rastreamento de pessoas através de sequências de imagens, por outro lado, menos atenção tem sido dirigida para o problema de determinar uma postura individual com base em uma única imagem. Na realidade, este problema é um desafio, porque as restrições das imagens 2D muitas vezes não são suficientes para determinar poses 3D de um objeto articulado como o ser humano se enquadra.

De acordo com estes fatos, neste trabalho é proposto um modelo que apresenta uma solução baseado no trabalho de Taylor [Tay00] para recuperar posturas 3D em imagens únicas. O modelo pode ser dividido em 2 partes principais sendo que a segunda parte é subdividida em 4 subpartes :

1. A inicialização manual do esqueleto;
2. A recuperação da pose 3D;
  - (a) A utilização da abordagem de Taylor [Tay00];
    - Obtenção de  $2^{20}$  possíveis poses;
    - Eliminação das poses idênticas;
  - (b) A aplicação de Restrições Biomecânicas nas seguintes partes do corpo:
    - Ombros;
    - Quadril;
    - Coluna; e
    - Joelhos.
  - (c) Medida de Conforto;
    - Obtenção de uma classificação pela ordenação da poses geradas conforme a medida de conforto;
  - (d) Abordagem sobre a Luminosidade;
    - Reordenação da classificação conforme o sombreamento.

A Figura 3.1 apresenta sucintamente as fases do modelo desenvolvido. Nas próximas subseções estas fases serão explicadas detalhadamente.

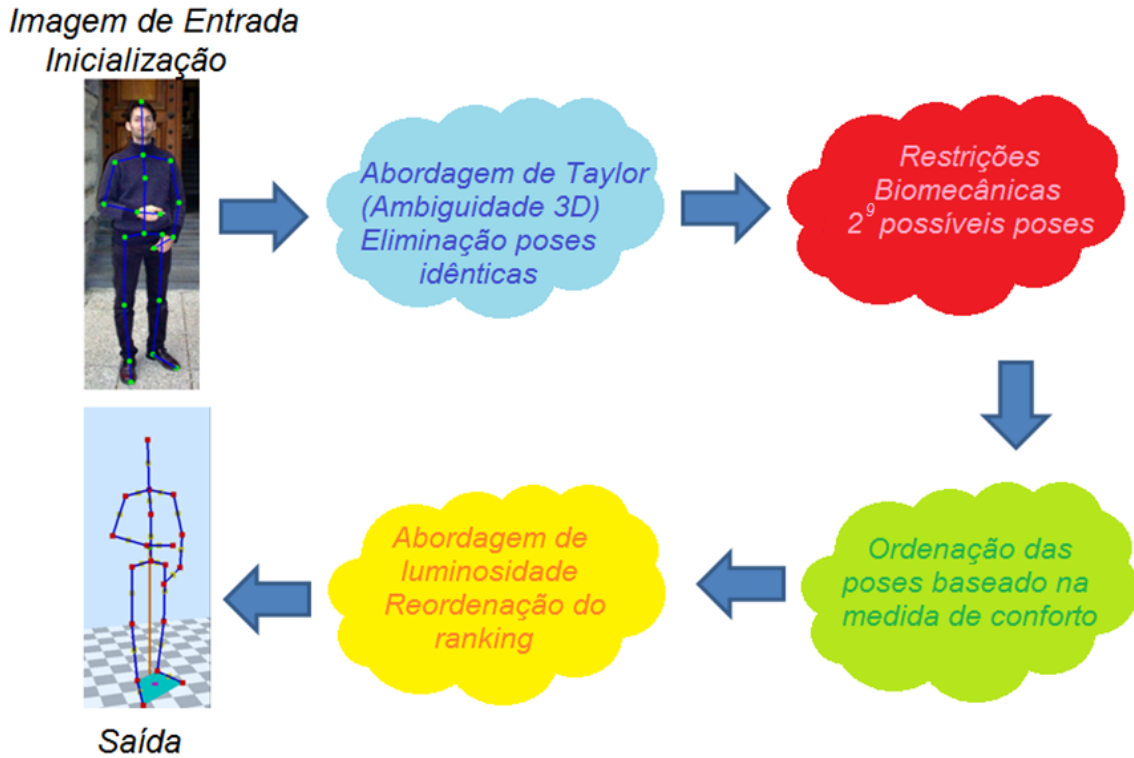


Figura 3.1: Fluxograma do modelo proposto.

### 3.1 Inicialização do Esqueleto

A detecção automática inicial de pessoas em imagens estáticas é uma tarefa desafiadora, ela é o passo inicial para resolver o problema apresentado. De acordo com Hornung et al. [HDK07], a aquisição inicial obtida de forma manual da postura 2D apresenta algumas vantagens quando comparada com procedimentos automáticos, pois a intervenção manual é realizada em pouco tempo e obtém-se resultados mais precisos em poses muito complexas ou parcialmente oclusas. Por esta razão o modelo apresentado também realiza o processo de inicialização do esqueleto de forma manual. Neste trabalho o modelo do esqueleto é composto por 19 ossos  $k$  e 20 articulações  $i$  como pode ser visualizado na Figura 3.2. Todos os ossos possuem inicialmente uma altura e largura parametrizada pela altura  $h$  da pessoa na imagem e pelas dimensões antropométricas de uma pessoa média de acordo com [Til02]. Mais precisamente para determinado osso  $k$  na imagem, é estimado um retângulo cujo comprimento correspondente é  $l_k$  e a largura é  $w_k$ , que são dados por:

$$l_k = h f_{lk}, \quad w_k = \frac{w_i + w_j}{2}, \quad (3.1)$$

onde  $w_i$  e  $w_j$  são dados por  $w_i = h f_{wi}$  e  $w_j = h f_{wj}$  respectivamente e os fatores de proporcionalidade  $f_{lk}$ ,  $f_{wi}$  e  $f_{wj}$  são obtidos de [Til02] e mostrados na Tabela 3.1.  $j$  é o índice da articulação seguinte a articulação  $i$ , de forma hierárquica do esqueleto, que juntos definem um osso, como visto na Figura 3.2.

O processo de inicialização manual é composto de dois estágios: a informação da altura da

pessoa na imagem em pixels e a localização das principais articulações da pessoa em coordenadas de imagem.

- A informação da altura da pessoa em pixels: Há duas formas diferentes de se obter a altura da pessoa através da intervenção manual. Quando a pessoa está em uma posição ereta na imagem e todo o seu corpo é visível, o usuário simplesmente clica no topo da cabeça da pessoa e na parte inferior dos pés, obtendo  $h$  diretamente. Em outra situação, por exemplo, quando a pessoa está sentada ou com grande parte do corpo oclusa na imagem, a altura pode ser estimada a partir de qualquer osso, incluindo a face. A parte escolhida (osso) deve estar paralela com o plano da imagem a fim de evitar problemas de perspectiva. Por exemplo, se o usuário escolher o comprimento da face como referência, ele deve clicar no topo da cabeça e na parte inferior do queixo, para obter o comprimento da face  $h_f$ . Então a altura da pessoa é calculada pela Equação  $h = h_f/f_{l0}$ , onde  $f_{l0}$  foi obtido da Tabela 3.1. O processo pode ser realizado para qualquer outro osso, desde que ele esteja paralelo ao plano da imagem. A Seção 3.2 explica como é utilizado o valor de  $l_k$  para obter a pose 3D.
- A posição das principais articulações da pessoa na imagem: Este processo é realizado também através dos cliques do usuário. São 20 pontos clicados sobre o corpo na imagem compostos de 15 articulações e 5 extremidades do corpo (Figura 3.2). Estes cliques devem ser dados de forma hierárquica e ordenada:
  1. A extremidade inicial é a posição acima da cabeça - ( $J_1$ );
  2. Pescoço - ( $J_2$ );
  3. Peito - ( $J_3$ );
  4. Abdômen - ( $J_4$ );
  5. Ombro direito - ( $J_5$ );
  6. Cotovelo direito - ( $J_6$ );
  7. Pulso direito - ( $J_7$ );
  8. Extremidade da mão direita - ( $J_8$ );
  9. Ombro esquerdo - ( $J_9$ );
  10. Cotovelo esquerdo - ( $J_{10}$ );
  11. Pulso esquerdo - ( $J_{11}$ );
  12. Extremidade da mão esquerda - ( $J_{12}$ );
  13. Quadril direito - ( $J_{13}$ );
  14. Joelho direito - ( $J_{14}$ );
  15. Tornozelo direito - ( $J_{15}$ );
  16. Extremidade do pé direito - ( $J_{16}$ );

17. Quadril esquerdo - ( $J_{17}$ );
18. Joelho esquerdo - ( $J_{18}$ );
19. Tornozelo esquerdo - ( $J_{19}$ ); e
20. Extremidade do pé esquerdo ( $J_{20}$ ).

É importante salientar que o modelo atual não trata o problema de oclusões parciais, devendo o usuário clicar em uma posição aproximada caso um dos pontos  $J_i$  esteja em uma região não visível ao observador da imagem. Obtidas as informações do processo inicial, o próximo passo é a geração da pose 3D, que é descrita na próxima seção.

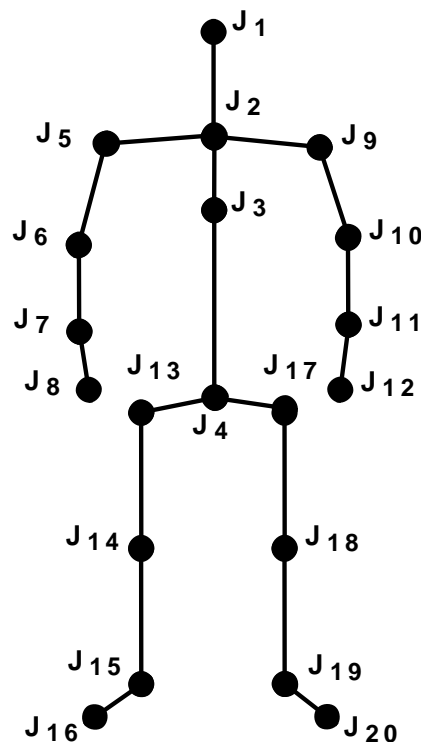


Figura 3.2: O modelo do esqueleto utilizado na proposta.

### 3.2 Identificação da pose 3D

O trabalho de Taylor [Tay00] apresenta um método para a recuperação de informações sobre a configuração de objetos articulados a partir de uma única imagem. As semelhanças com este trabalho é que ambos os métodos assumem uma projeção ortográfica e ambos empregam informações geométricas como restrições. De acordo com Taylor, tendo-se um segmento de reta de comprimento conhecido  $l$  na imagem em projeção ortográfica em escala (que corresponde a um osso do esqueleto), os dois pontos finais em 3D (articulações do referido osso)  $(x_1, y_1, z_1)$  e  $(x_2, y_2, z_2)$  são projetadas para  $(u_1, v_1)$  e  $(u_2, v_2)$ , respectivamente. Se o fator de escala  $s$  do modelo de projeção é conhecido, é uma questão simples calcular a profundidade relativa dos dois pontos finais, denotado por  $\Delta Z = z_1 - z_2$ , usando a seguinte equação [Tay00]:

Tabela 3.1: Os ossos, articulações e os fatores de proporcionalidade  $f_{lk}$ ,  $f_{wi}$  and  $f_{wj}$ .

$k$	Ossos	Articulações $i - j$	$f_{lk}$	$f_{wi}$	$f_{wj}$
0	Cabeça	$(J_1 - J_2)$	0.20	0.0883	0.0883
1	Peito	$(J_2 - J_3)$	0.098	0.2040	0.1765
2	Abdômen	$(J_3 - J_4)$	0.172	0.1765	0.1649
3	Ombro Direito	$(J_2 - J_5)$	0.102	0.0607	0.0607
4	Braço Direito	$(J_5 - J_6)$	0.159	0.0607	0.0376
5	Antebraço Direito	$(J_6 - J_7)$	0.146	0.0376	0.0289
6	Mão Direita	$(J_7 - J_8)$	0.108	0.0289	0.0289
7	Ombro Esquerdo	$(J_2 - J_9)$	0.102	0.0607	0.0607
8	Braço Esquerdo	$(J_9 - J_{10})$	0.159	0.0607	0.0376
9	Antebraço Esquerdo	$(J_{10} - J_{11})$	0.146	0.0376	0.0289
10	Mão Esquerda	$(J_{11} - J_{12})$	0.108	0.0289	0.0289
11	Quadril Direito	$(J_4 - J_{13})$	0.050	0.1013	0.1013
12	Coxa Direita	$(J_{13} - J_{14})$	0.241	0.1013	0.0607
13	Panturrilha Direita	$(J_{14} - J_{15})$	0.240	0.0607	0.0549
14	Pé Direito	$(J_{15} - J_{16})$	0.123	0.0549	0.0549
15	Quadril Esquerdo	$(J_4 - J_{17})$	0.050	0.1013	0.1013
16	Coxa Esquerda	$(J_{17} - J_{18})$	0.241	0.1013	0.0607
17	Panturrilha Esquerda	$(J_{18} - J_{19})$	0.240	0.0607	0.0549
18	Pé Esquerdo	$(J_{19} - J_{20})$	0.123	0.0549	0.0549

$$\Delta Z^2 = l^2 - \frac{(u_1 - u_2)^2 + (v_1 - v_2)^2}{s^2}. \quad (3.2)$$

Tal formulação gera ambiguidades para cada segmento, uma vez que o sinal de  $\Delta Z$  não pode ser determinado (ou seja, têm-se a possibilidade de  $z_1 > z_2$  ou  $z_2 > z_1$ ). De fato,  $\Delta Z$  é calculado para cada osso e aplicado na segunda junta do osso calculado e na primeira junta do próximo osso na hierarquia. Se o esqueleto tem vinte articulações, há no máximo  $2^{20}$  posturas possíveis, no pior cenário ( $\Delta Z_i \neq 0$  para todas  $i$  articulações). Por outro lado, quando  $\Delta Z_i = 0$ , não há ambiguidade na articulação  $i$ . A fim de reduzir o número de posturas possíveis, em primeiro lugar definiu-se a equivalência de posturas que são encontradas quando todas as articulações têm as mesmas coordenadas  $z$ . Isso acontece porque quando as pessoas estão em pé, muitas articulações estão geralmente alinhadas no plano da imagem (por exemplo,  $z_i$  é equivalente nas posturas analisadas). Um aspecto importante neste modelo é que considera o esqueleto hierárquico, ou seja, o  $\Delta Z$  calculado é somado aos ossos que seguem-se na hierarquia, a fim de encontrar a postura final. Em uma próxima etapa, apresenta-se um conjunto de restrições que visam eliminar posturas impossíveis para os seres humanos.

### 3.2.1 Restrições Biomecânicas

Após a eliminação das poses equivalentes, um conjunto de restrições biomecânicas do corpo de acordo com Tilley [Til02] é aplicado para minimizar o problema de ambiguidade nas poses 3D articuladas. As limitações biomecânicas do corpo consideram a relação de ossos ligados através das articulações, e também os seus respectivos ângulos de rotação. As restrições preservam as distâncias entre quaisquer duas articulações, independentemente do movimento do corpo humano, e evitam ângulos impossíveis na postura. Inicialmente, através da Equação 3.2 pode-se definir o valor de  $\Delta Z$  para cada junta relacionada com a junta anterior no esqueleto. Contudo, visto que  $\Delta Z$  pode ter tanto um valor positivo como negativo, é definido um conjunto de suposições e restrições que são aplicadas em algumas articulações<sup>1</sup> ( $J_i$  da Figura 3.2), a fim de reduzir a ambiguidade, que serão apresentadas a seguir.

i)  $\Delta Z_{J_1} = 0$

Suposição 1: O topo da cabeça está sempre no plano da imagem. A Figura 3.3 ilustra essa Suposição.

ii)  $sign(\Delta Z_{J_4}) = sign(\Delta Z_{J_3}) = sign(\Delta Z_{J_2})$

Restrição 1: O objetivo é evitar que as articulações da coluna fique alternadas com valores positivos e negativos, o que seria irreal do ponto de vista de postura humana. Desta forma, uma vez que  $\Delta Z_{J_2}$  é determinado,  $\Delta Z_{J_3}$  e  $\Delta Z_{J_4}$  são configurados de acordo, adquirindo o mesmo sinal. Ver na Figura 3.4

iii)  $\Delta Z_{J_5} \geq 0$  e  $\Delta Z_{J_9} \geq 0$

Suposição 2: Inicialmente assume-se que  $\Delta Z_{J_5}$  e  $\Delta Z_{J_9}$  são positivos, mas eles podem ser manualmente definidos pelo usuários.

Restrição 2: A fim de evitar inconsistências antropométricas, é incluída uma restrição biomecânica que define o ângulo entre os dois ombros, definidos pelos vetores:  $\vec{v} = J_5 - J_2$  e  $\vec{u} = J_9 - J_2$ . Se o ângulo interno entre  $\vec{v}$  e  $\vec{u}$  é menor que  $165^\circ$  (de acordo com [NH94]), então o sinal de  $\Delta Z_{J_9}$  é oposto a  $\Delta Z_{J_5}$ . A Figura 3.5 ilustra essa Restrição.

iv)  $sign(\Delta Z_{J_8}) = sign(\Delta Z_{J_7})$  and  $sign(\Delta Z_{J_{12}}) = sign(\Delta Z_{J_{11}})$

Suposição 3: Assume-se que o deslocamento das mãos está sempre com o mesmo sinal do deslocamento do antebraço.

v)  $\Delta Z_{J_{13}} \geq 0$  e  $\Delta Z_{J_{17}} \geq 0$

Suposição 4: Similar a Suposição 2, assume-se inicialmente que  $\Delta Z_{J_{13}}$  e  $\Delta Z_{J_{17}}$  são positivos.

Restrição 3: Análoga a Restrição 2, mas utilizando o vetores entre os pontos  $J_{13}$ ,  $J_4$  e  $J_{17}$ , configurado por  $\vec{v} = J_{13} - J_4$  e  $\vec{u} = J_{17} - J_4$ , testando se o ângulo interno entre os vetores é menor que  $180^\circ$  [NH94]. Se isto é verdade, o sinal de  $\Delta Z_{J_{17}}$  é oposto ao  $\Delta Z_{J_{13}}$ .

vi)  $\Delta Z_{J_{16}} > 0$  e  $\Delta Z_{J_{20}} > 0$

Suposição 5: Assume-se que as extremidades dos pés são sempre deslocadas positivamente, visto que as fotos devem ser de poses frontais, conforme definido nos objetivos deste trabalho.

<sup>1</sup>As suposições e restrições são aplicadas quando  $\Delta Z_i \neq 0$ .

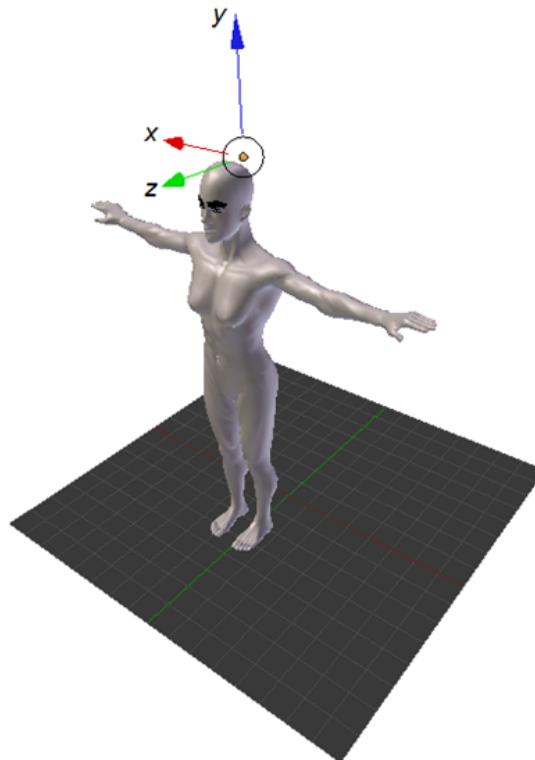


Figura 3.3: Suposição 1: O topo da cabeça está sempre no plano da imagem.

As restrições biomecânicas propostas, de fato reduzem o número de ambiguidades a um máximo de  $2^9$  possíveis poses. Assim, após a aplicação destas restrições, há ainda um conjunto de poses que podem gerar ambiguidades se  $\Delta Z_i \neq 0$  são:  $J_2, J_6, J_7, J_{10}, J_{11}, J_{14}, J_{15}, J_{18}$  e  $J_{19}$ .

Uma restrição adicional foi inserida a fim de reduzir as poses impossíveis de joelhos. Esta restrição é aplicada somente nos casos em que o ângulo entre a coxa e a parte inferior da perna é maior do que  $180^\circ$  [NH94], sua finalidade é impedir poses com “joelhos quebrados”. A Figura 3.6 apresenta dois resultados do modelo baseado na imagem de entrada - Figura 3.6(a). É importante lembrar que as 2 posturas apresentam projeção 2D equivalentes. Os resultados são visualizados de um ponto de vista lateral da pessoa na imagem para melhor observar os resultados. A Figura 3.6(b) é o resultado sem a restrição do joelho e a Figura 3.6(c) com a restrição.

De fato, na maioria dos casos, o número de posturas geradas é menor que 512, devido à natureza intrínseca das imagens usuais, ou seja, apresentando muitas articulações com  $\Delta Z = 0$ , não sendo computado um valor de deslocamento para  $\Delta Z$ . Isso faz com que muitas poses sejam iguais e possam ser descartadas conforme descrito anteriormente.

O conjunto de restrições biomecânicas apresentadas nesta seção foram publicadas no trabalho de Jacques et al [JDJM13]. A próxima seção descreve uma abordagem para ordenação das poses restantes.

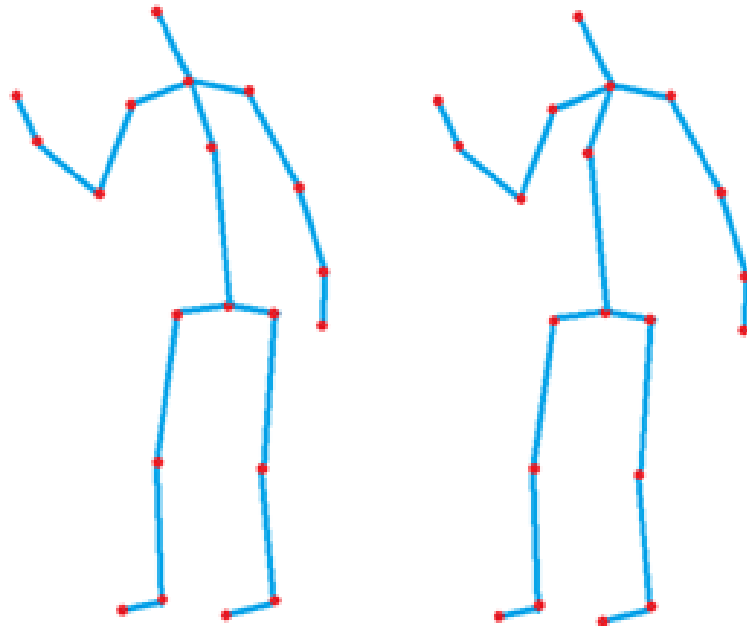


Figura 3.4: Suposição 2: Os ossos da coluna devem estar com o sinal de delta Z em acordo.

### 3.2.2 Ordenação das posturas geradas baseado na medida de conforto

Como o modelo apresentado possui 19 ossos e 20 articulações, e as suposições e restrições biomecânicas são aplicadas somente a 11 articulações, existe ainda um conjunto de  $2^9$  possibilidades de poses no pior caso. Desenvolveu-se então um modelo baseado no conforto das posturas a fim de prover ordenação das poses geradas e facilitar a escolha de uma pose que melhor se enquadre na postura da imagem.

De acordo com Mochizuki [LM03] o controle postural é extremamente complexo. Duas grandezas podem ser obtidas por meio da biomecânica para o estudo da postura, que é o centro de massa da pessoa e o centro de apoio. O autor afirma que para manter o equilíbrio, a pessoa deve minimizar a diferença entre essas duas grandezas. Baseado nestas observações, verificou-se que algumas das posturas geradas são mais improváveis que outras devido ao esforço físico que seria necessário para a pessoa manter determinada pose. A fim de estabelecer um critério para a classificação desse conjunto de posturas geradas de acordo com sua probabilidade de ocorrência, foi proposta uma forma de estimar o “conforto” de cada postura. Este conforto da postura foi modelado como a diferença entre o centro de massa da pessoa e a localização do centro de apoio (localizado em torno do pé) da pessoa. Se essa diferença for maior que um determinado limiar, obtido empiricamente, considera-se a postura desconfortável (assumindo que a pessoa está em pé). Observou-se também que, quando os valores dessa diferença são grandes, a pessoa normalmente está sentada ou em fotos tiradas durante algum movimento, por exemplo dançando. Na verdade, quanto maior a distância entre a projeção do centro de massa e a projeção do centro de apoio, será necessário um esforço



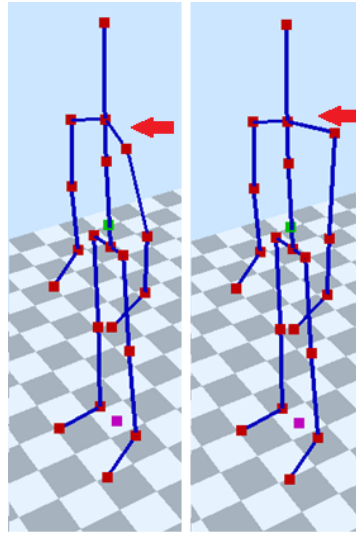


Figura 3.5: Restrição 3: Define o ângulo entre os dois ombros, de acordo com os vetores:  $\vec{v} = J_5 - J_2$  e  $\vec{u} = J_9 - J_2$ . Se o ângulo interno entre  $\vec{v}$  e  $\vec{u}$  é menor que  $165^\circ$  (de acordo com [NH94]), então o sinal de  $\Delta Z_{J_9}$  é oposto a  $\Delta Z_{J_5}$ .

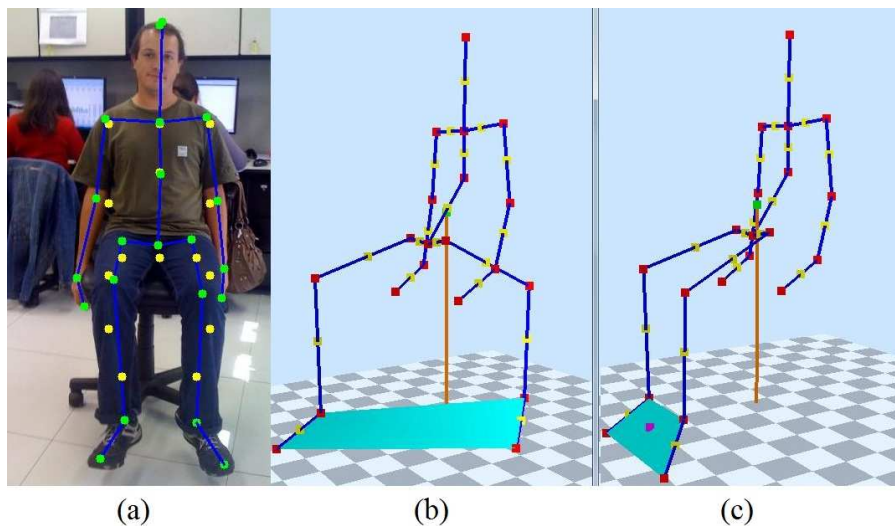


Figura 3.6: Resultado após a aplicação da restrição dos joelhos, (a) Imagem de entrada, (b) Resultado sem restrição do ângulo entre coxa e a parte inferior da perna. (c) Resultado com restrição: evita a pose que a pessoa estaria com a coxa para trás e a parte inferior da perna para frente.

maior para a pessoa manter seu equilíbrio.

Para calcular o centro de massa de uma determinada postura, cada parte do corpo (relacionado com determinado osso  $k$ ), foi modelado através de um tronco de cone (Figura 3.7), no qual sua altura coincide com o comprimento  $l_i$  do osso correspondente.

Os raios do tronco de cone são determinados pela largura do osso  $k$  e são calculados usando  $w_i$  e  $w_j$  (valores especificados pelas articulações  $i$  e  $j$  conforme a Tabela 3.1). Assumindo-se que a base do tronco de cone está na origem de um sistema de coordenadas local em 3D, e sua altura é alinhada com o eixo  $z$ , seu centro geométrico da massa será de  $(0, 0, cm_k)$ , onde

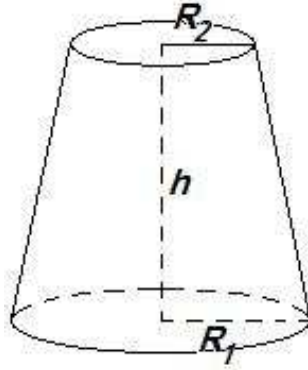


Figura 3.7: Cada osso do corpo foi modelado como um tronco de cone, a fim de calcular seu centro de massa.

$$cm_k = \frac{l_k (r_k^2 + 2r_k R_k + 3R_k^2)}{4 (r_k^2 + r_k R_k + R_k^2)}, \quad (3.3)$$

e  $l_k$  é a altura do tronco de cone,  $r_k = \frac{w_i}{2}$  e  $R_k = \frac{w_j}{2}$ . Deve-se notar que o centro de massa de uma determinada parte do corpo é sempre sobre o seu eixo central (isto é, seu osso correspondente), e  $cm_k$  representa sua distância da base do osso. Por isso, é trivial obter as coordenadas 3D  $cm_k = (cm_k^x, cm_k^y, cm_k^z)$  do centro de massa no sistema de coordenadas global, como ilustrado pelos pontos azuis na Figura 3.8.

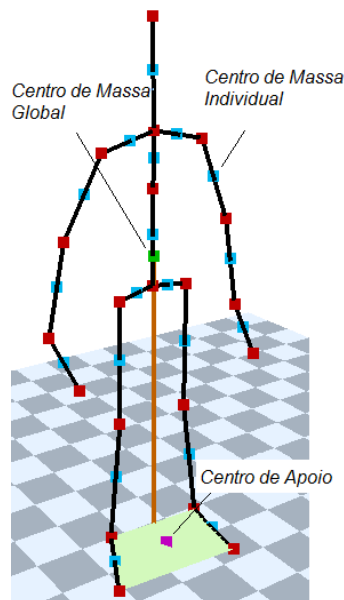


Figura 3.8: Os pontos azuis representam o centro de massa individual de cada osso, o ponto magenta indica o ponto de apoio do corpo e o ponto verde mostra o centro de massa global do corpo.

Dado o centro de massa de cada osso, então é possível calcular o centro de massa global do esqueleto. O centro de massa global  $gcm$  do esqueleto (considerando a densidade de massa de cada parte do corpo constante) é calculado por:

$$gcm = \frac{\sum_{k=1}^n fv_k cm_k}{\sum_{k=1}^n fv_k}, \quad (3.4)$$

onde  $n$  é total de numero de ossos, e  $fv_k$  é o volume do frustrum de cada osso  $k$ , dado por

$$fv_k = \frac{1}{3}\pi (r_k^2 + r_k R_k + R_k^2). \quad (3.5)$$

Após o cálculo centro de massa global, é calculado o centro de apoio do esqueleto (*sup*) que é dado por um ponto central de um quadrilátero formado pelos pés da pessoa em um plano  $xz$ , como pode ser visto na Figura 3.8. Após, o *gcm* é projetado até o plano do  $xz$  do *sup*, conforme a linha laranja da Figura 3.8. Então, para cada postura gerada (máximo de  $2^9$  poses), a distância Euclidiana ( $dC$ ) entre *gcm* projetado e *sup* é calculada e normalizada de acordo com a altura  $h$  da pessoa na imagem

$$dC = \frac{d(gcm, sup)}{h}. \quad (3.6)$$

Supõem-se que quanto mais perto a projeção do centro de massa está do centro de apoio, mais provável é o equilíbrio estático do corpo. Com base nesta suposição, a hipótese é de que valores menores da distância normalizada  $dC$  indicam que a postura gerada é a mais equilibrada e, portanto, mais confortável para a pessoa. Assim, se a média dos valores  $dC$  das  $2^9$  possíveis poses é relativamente baixo, assume-se que a pessoa está de pé em uma postura equilibrada, ou seja, confortável. Se a média de  $dC$  para todas as posturas geradas é relativamente alta, isto indica que a pessoa não está sendo suportada pelos seus pés em uma condição equilibrada, ou que a pessoa provavelmente não está em pé.

Pode-se observar, através de experimentos, a aceitação desta hipótese. Dadas estas definições, foi realizada uma classificação inicial, de posturas em pé, comparando a média de  $dC$  de todas as posturas geradas ( $\bar{dC}$ ), com um limiar de  $T_{\bar{dC}} = \alpha.h$  (foi configurado inicialmente  $\alpha = 0.1$ ). Assim sendo, considera-se que a pessoa está em pé sempre que  $dC < T_{\bar{dC}}$ . Se uma pose em pé é detectada, todas as poses possíveis são classificados em ordem crescente de acordo com  $dC$ , e mostradas para o usuário. Por outro lado, para as posturas não em pé, a classificação é realizada em ordem decrescente. Dois exemplos da classificação são apresentados na Figura 3.9. Nesta figura podem ser vistas as poses classificadas nas 4 primeiras colocações e para cada pose é mostrada a distância de conforto obtida pelo modelo proposto. É importante notar que nesta fase não são excluídas nenhuma das poses geradas, o modelo somente fornece um método de sugerir aos usuários as posturas mais prováveis baseado na classificação.

Nesta fase observa-se que em certas posturas ainda há ambiguidade no critério de conforto, uma vez que pode existir a mesma distância de conforto para diferentes posturas, conforme ilustrado pela Figura 3.10. A fim de melhorar a classificação das posturas, é feita uma análise do sombreamento da imagem. Esta abordagem é descrita na próxima seção.

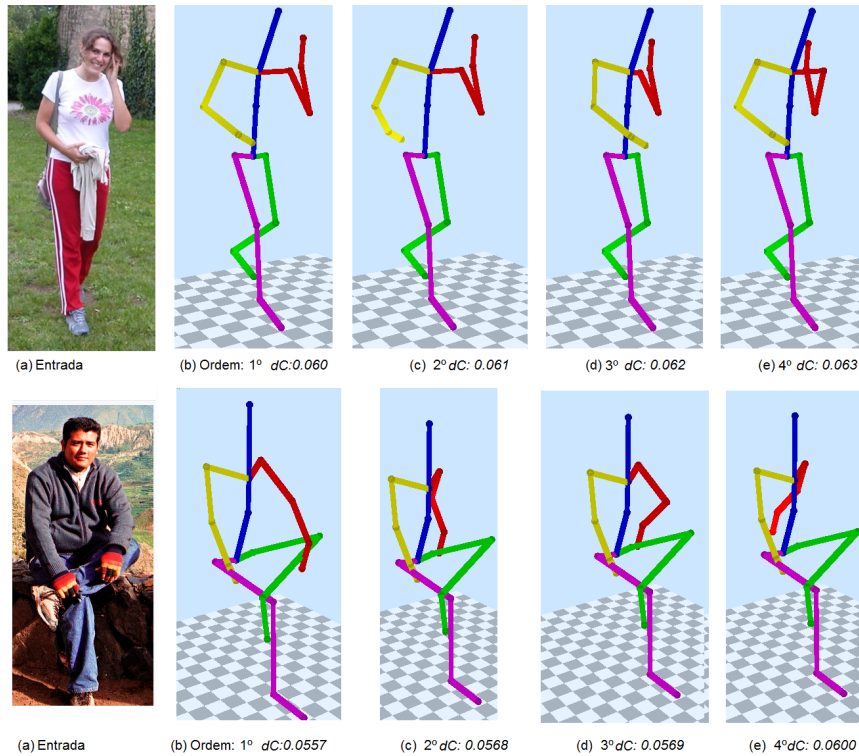


Figura 3.9: Exemplo das imagens de entrada e as poses ordenadas. As poses são mostradas em pontos diferentes de visão a fim de ter uma noção melhor da postura 3D. Abaixo de cada pose, aparece a sua posição na classificação de conforto e a sua distância  $dC$

### 3.2.3 Abordagem através da Luminosidade

Como mencionado anteriormente, após calcular a classificação de conforto, observa-se que diferentes poses podem ter mesma distância  $dC$  do conforto. Isto acontece com poses simétricas, como pode ser visto na Figura 3.10. Nesta Figura é possível ver duas situações em que a única diferença na postura está relacionada com  $\Delta Z$  dos braços. À esquerda, o braço direito do humano virtual está na parte da frente do corpo e do braço esquerdo está por trás do corpo, enquanto que o oposto acontece na imagem do lado direito. Isso faz com que a medida de conforto  $dC$  para as duas posturas seja a mesma. A fim de melhorar a classificação das poses, foi criada uma abordagem para minimizar essa ambiguidade das poses simétricas através de uma análise de luminosidade dos membros.

A fim de que a abordagem tenha um resultado satisfatório, duas informações iniciais são obtidas automaticamente pelo modelo. A primeira é verificar se a posição da fonte de luz no momento da criação da imagem é frontal à pessoa na fotografia e a segunda é verificar se a diferença de intensidade de luminosidade entre membros simétricos é satisfatória para prover um resultado eficiente.

Para a verificação da posição da fonte de luz foi proposta a seguinte abordagem: Durante o processo de inicialização, conforme a Seção 3.1, uma área retangular  $A$  da região dos pés da pessoa na imagem é recuperada conforme a Figura 3.11 (b). Esta área retangular é obtida através das posições das articulações  $J_{15}$ ,  $J_{16}$ ,  $J_{19}$  e  $J_{20}$  e aumentada em 5% do seu tamanho inicial a fim de proporcionar as informações sobre a iluminação desta região. Nesta região é aplicada a abordagem

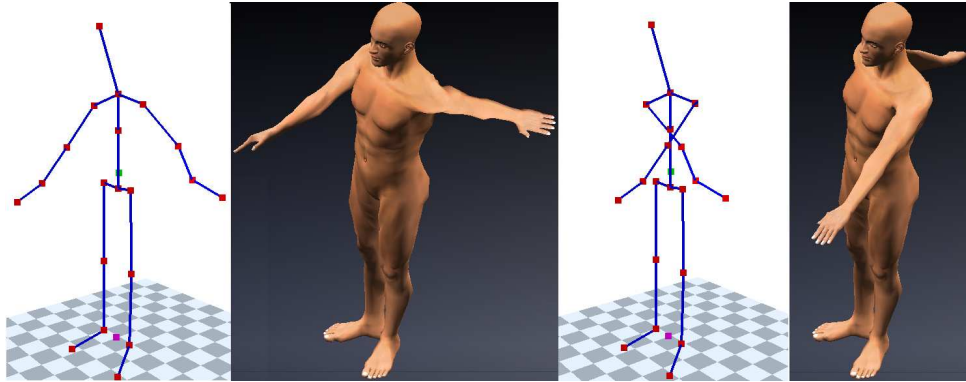


Figura 3.10: Devido à semelhança das posições dos braços, a distância  $dC$  é a mesma nas duas posturas. No primeiro caso, o braço direito do humano virtual está a frente do corpo e o braço esquerdo está para trás do corpo. E o oposto acontece na Figura da direita. Como consequência, as posturas de ambos têm a mesma medida de conforto.

de detecção de sombra baseada na proposta por Guo, Daim e Hoiem [GDH11]. Esta abordagem, diferentemente dos métodos tradicionais que exploram *pixel* ou informações de borda, é baseada em regiões. A fim de determinar qual região é considerada sombra, este método utiliza um classificador. Este classificador é treinado através de regiões rotuladas e classificadas como sombra. Este modelo de detecção de sombra irá determinar quais *pixels* da região  $A$  são considerados sombras e quais não são.

Para determinar a localização da fonte de luz no momento da geração da imagem, a região  $A$  é então dividida em 2 quadrantes, quadrante superior  $Q_{sup}$  e quadrante inferior  $Q_{inf}$ . Estes quadrantes podem ser visualizados pela linha em vermelho na Figura 3.11(c). Baseado nos resultados obtidos pelo processo de detecção de sombra, para cada quadrante é feito o somatório dos *pixels* considerados sombra pelo modelo de Guo, Daim e Hoiem [GDH11]. Para determinar se a localização da fonte de luz está em uma posição frontal o resultado do somatório dos pixels considerados sombra no quadrante  $Q_{sup}$  deve ser maior que a do quadrante  $Q_{inf}$ , caso contrário define-se que a fonte de luz está atrás da pessoa na imagem. Quando o sistema determina que a fonte de luz é frontal à pessoa, a abordagem de luminosidade é aplicada ao processo de detecção de poses, caso contrário a abordagem de luminosidade não é aplicada e a ordenação não é modificada.

Para o prosseguimento da abordagem, as regiões de interesse (*ROI*) dos ossos do esqueleto que já foram automaticamente determinadas durante o processo de inicialização conforme a Seção 3.1, são utilizadas. O exemplo de uma destas regiões pode ser visualizada na Figura 3.12 pelo retângulo em preto. A posição das ROIs são calculadas utilizando as posições das articulações  $J_i$  e  $J_j$  clicadas pelo usuário e os seus tamanhos são obtidos através dos valores de  $l_k$  e  $w_k$  de cada osso (Tabela 3.1). Para a *ROI* de cada osso calcula-se  $med_k$  que é dada pela mediana do valor da luminância  $V$  do espaço de cor *HSV* de cada pixel da *ROI*. O modelo de cor *HSV* é um tipo de método para definir a cor de acordo com três características básicas: Matiz (*Hue*), Saturação (*Saturation*) e Luminância (*Luminance*). Este modelo foi apresentado por Alvy Ray Smith, em 1978. Ele é uma transformação não linear do modelo de espaço de cor *RGB*. As definições matemáticas do modelo

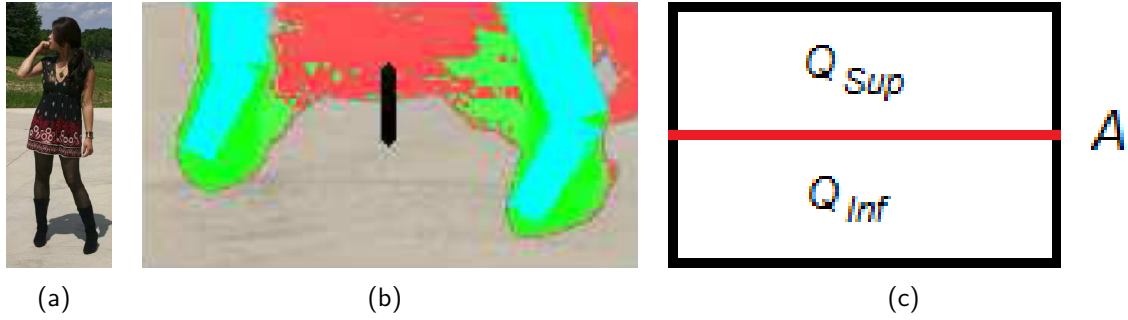


Figura 3.11: A região  $A$ , mostrada na imagem (b), é calculada pelas articulações  $J_{15}$ ,  $J_{16}$ ,  $J_{19}$  e  $J_{20}$  clicadas pelo usuário no processo de inicialização com a entrada da imagem (a). A imagem (c) mostra como a região  $A$  é então dividida em dois quadrantes ( $Q_{sup}$  e  $Q_{inf}$ ). Para cada quadrante é realizado o somatório dos pixels considerados sombras pela abordagem proposta por Guo, Daim e Hoiem [GDH11] (*pixels* em vermelho e verde) da imagem (b). Caso  $Q_{sup}$  tenha um resultado superior a  $Q_{inf}$  é definido que a fonte de luz está na parte frontal da pessoa na imagem.

$HSV$  são dadas por

$$H = \begin{cases} 0^0 & \text{if } \max = \min \\ 60^0 \times \frac{G-B}{\max-\min} + 0^0, & \text{if } \max = R \text{ and } G \geq B \\ 60^0 \times \frac{G-B}{\max-\min} + 360^0, & \text{if } \max = R \text{ and } G < B \\ 60^0 \times \frac{B-R}{\max-\min} + 120^0, & \text{if } \max = G \\ 60^0 \times \frac{R-G}{\max-\min} + 240^0, & \text{if } \max = B \end{cases} \quad (3.7)$$

$$S = \begin{cases} 0, & \text{if } \max = 0 \\ \frac{\max-\min}{\max} = 1 - \frac{\min}{\max}, & \text{otherwise} \end{cases} \quad (3.8)$$

$$V = \max. \quad (3.9)$$

onde  $\max = \text{máximo}(R, G, B)$  do espaço de cores  $R$ ,  $G$  e  $B$  e  $\min = \text{mínimo}(R, G, B)$ .



Figura 3.12: A região de interesse (retângulo em preto) é calculada usando a posição dos pontos clicados pelo usuário e as medidas antropométricas.

Depois de determinado que a posição da fonte de luz no momento em que foi criada a imagem estava em uma posição frontal à pessoa e o valor de  $med_k$  de cada  $ROI$ , o próximo passo do processo é verificar a diferença das  $med_k$  entre os membros simétricos, por exemplo, braço esquerdo com braço direito e coxa esquerda com coxa direita. Inicialmente, antes de realizar a comparação de luminância entre membros, é verificado se a área da  $ROI$  de cada um dos membros possui uma área maior que 30% (valor foi obtido empiricamente) da área do membro (osso) correspondente utilizando os valores de  $f_{li}$  e  $f_{wi}$ , de acordo com a Tabela 3.1. Esta verificação é realizada inicialmente para os dois braços, se um deles não possuir o tamanho mínimo (provavelmente devido a projeção da perspectiva) o processo não é realizado. Isto significa que a  $ROI$  não possui o tamanho necessário para ser empregada no processo de decisão.

Caso as  $ROI$ s passem no teste da área, o processo segue as avaliando. No passo seguinte a luminosidade é usada para decidir se os braços estão na mesma direção ou estão em direções opostas (um para frente e o outro para trás). Para realizar este passo é calculada a diferença de luminância entre os dois braços. Empiricamente, se a diferença entre as  $med_k$  for menor que 10% do canal  $V$ , supõe-se que os braços devem estar na mesma direção. Quando os braços estão na mesma direção, é verificado se eles estão projetados para trás ou para frente do corpo. Esta decisão é tomada comparando a luminância dos braços com a luminância do tronco. Se  $med_k$  de algum dos braços for maior que a luminância do tronco, é determinado que os braços estão a frente do corpo sendo positivo os sinais de  $\Delta Z$  das articulações  $J_5$  e  $J_9$ . Caso contrário, os membros estão atrás do tronco sendo os sinais de  $\Delta Z$  negativos.

Quando é verificado que os braços estão em direções opostas, é feita uma comparação da luminância entre eles. O braço que possuir a  $med_k$  maior é considerado à frente do corpo e o outro, atrás do corpo. Desta forma, é determinada a direção dos braços da pessoa na imagem. A Figura 3.13 mostra o fluxograma do algoritmo utilizado para a definição da projeção dos braços baseado na luminosidade da imagem.

Similarmente, o processo é aplicado para os antebraços e pernas, especificamente no conjunto  $A$  definido pelas articulações  $\{J_6, J_7, J_{10}, J_{11}, J_{14}, J_{15}, J_{18}, J_{19}\}$ . Como  $\Delta Z$  para as articulações analisadas com base na luminosidade pode apresentar resultados diferentes se comparados com as melhores posições na classificação baseada no conforto, foi proposta uma forma de integrar ambos os critérios a fim de proporcionar uma nova classificação. Definiu-se um coeficiente de penalização  $PC_p$  para cada possível pose  $p$ , definido por

$$PC_p = \begin{cases} (d_p f_p) & \text{if } \bar{d} \leq T_{\bar{d}C}, \\ \left(\frac{d_p}{f_p}\right) & \text{otherwise,} \end{cases} \quad (3.10)$$

onde  $d_p$  é a medida de conforto da pose possível,  $T_{\bar{d}C}$  é o liminar para a caracterização da pose em pé e  $p$  e  $f_p$  são definidos como segue: Se  $S_L(A_x)$  denota o sinal da junta  $x$  (1 ou -1) baseado na análise de luminância, e  $S_P(A_x, p)$  denota o sinal da junta  $x$  relativa a pose  $p$  calculada conforme

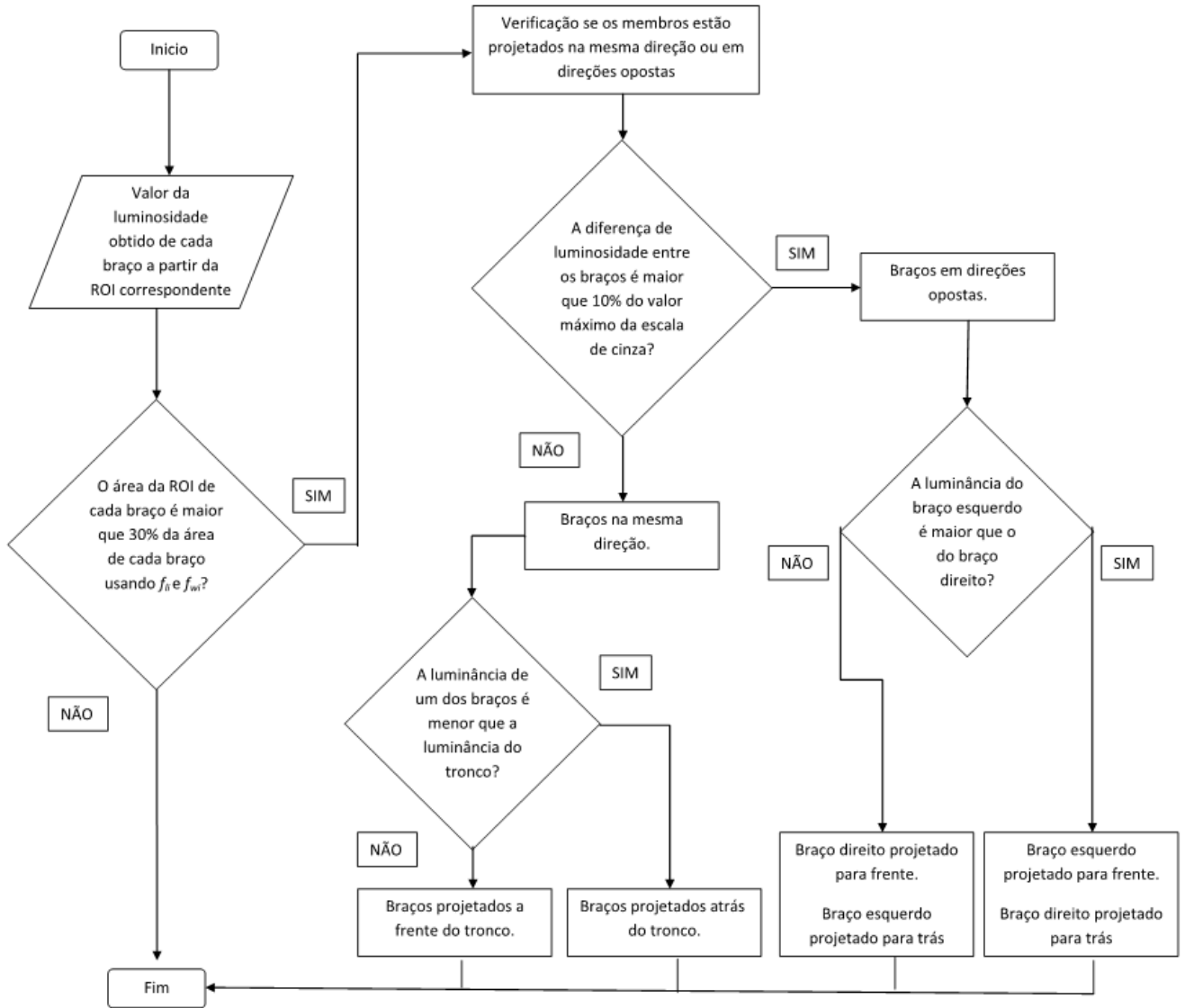


Figura 3.13: Fluxograma do algoritmo para a obtenção da posição dos braços baseado na luminosidade da imagem.

o conforto da postura, a função proposta  $f_p$  é dada por

$$f_p = \sum_{x=1}^m \tilde{\delta}(S_L(A_x) - S_P(A_x, p)), \quad (3.11)$$

onde  $\tilde{\delta}(n)$  é o complemento binário da função discreta, i.e.  $\tilde{\delta}(n) = 1$  if  $n \neq 0$ , e  $\tilde{\delta}(n) = 0$  caso contrário.  $A_x$  é o  $x$ -ésimo elemento de  $A$  e  $m$  é o número de elementos de  $A$ . Desta forma,  $f_p$  conta o número de articulações que  $S_L$  e  $S_P$  estão em desacordo.

Computados o  $PC_p$  de cada pose  $p$ , a classificação de conforto é reordenado, obtendo-se uma nova classificação. Essa reordenação é realizada de acordo com as penalidades impostas pelo modelo, fazendo com que as poses que estejam menos penalizadas tenham uma classificação superior as mais penalizadas, a partir da classificação do conforto. No Capítulo 4 é apresentada uma comparação entre as duas formas de ordenação das imagens e discute-se o impacto da análise de luminosidade



para encontrar as melhores poses. A Figura 3.14 apresenta o resultado da aplicação do modelo de luminosidade em uma imagem e a sua influência na classificação das poses geradas pelo modelo.

No próximo capítulo serão descritas as análises realizadas a fim de avaliar o modelo proposto.

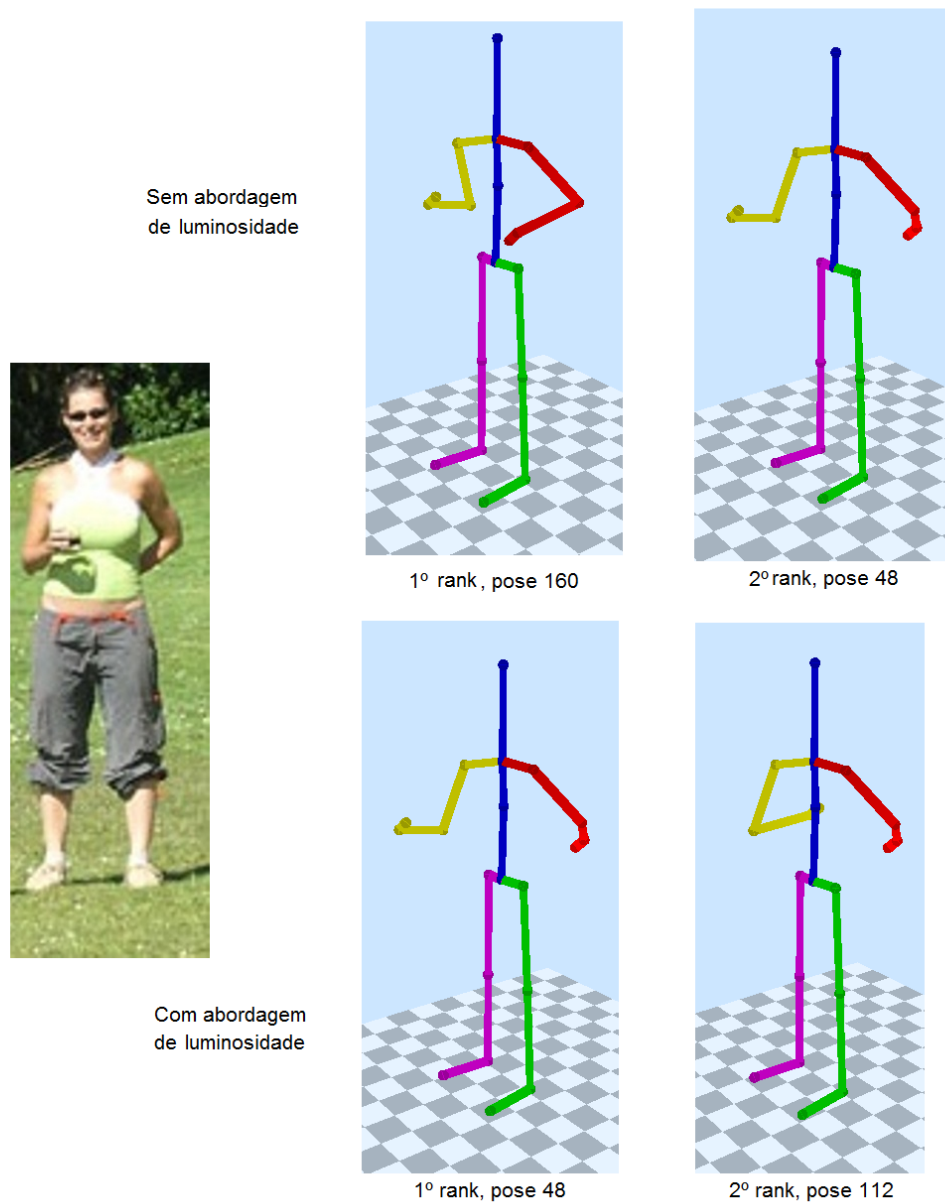


Figura 3.14: Resultado após a aplicação do modelo de luminosidade sobre a classificação das posturas baseadas no conforto. As imagens superiores apresentam a classificação das poses obtidas somente com o critério do conforto, onde a pose correta está na segunda posição. Nas imagens inferiores é apresentada a classificação com a aplicação do modelo de luminosidade onde a pose gerada correta passou para a primeira posição devido a sua menor penalização.

## 4. RESULTADOS

Neste capítulo são descritos os experimentos realizados nesta tese a fim de realizar uma avaliação do modelo apresentado. Inicialmente, foram selecionadas 443 imagens contendo pessoas em diversas posturas. Estas imagens foram obtidas em bases de dados disponíveis na internet de acordo com os trabalhos [FMJZ08, BM09, DT05] e também algumas foram criadas no Laboratório de Humanos Virtuais - VHLab do PPGCC da PUC RS. As imagens obtidas são fotografias de pessoas com aparência total do corpo e de frente para a câmera. Para cada imagem 2D foi gerado o seu *ground truth* em 3D de forma manual. O *ground truth* para cada imagem é definido pelo valor do sinal de  $\Delta Z$  para cada junta  $i$  do esqueleto 3D.

Para a realização das análises foi desenvolvido um protótipo de acordo com o modelo proposto. Este protótipo foi desenvolvido utilizando a linguagem c++. Para a manipulação das imagens, onde é feita a recuperação das articulações da pessoa e a análise de luminosidade, foi empregada a biblioteca OpenCV <sup>1</sup>, e para a criação da poses em tridimensionais foi utilizada a biblioteca OpenGL <sup>2</sup>. O protótipo para analisar e gerar as posturas 3D a partir das informações de entrada do modelo teve um custo médio de tempo de 15s para cada imagem.

Foram realizadas 5 tipos de análise a fim de validar o modelo de recuperação de posturas humanas 3D a partir de imagens bidimensionais. A primeira análise teve por objetivo verificar se a pose correta da pessoa no mundo real permanece no conjunto final das possíveis poses geradas pelo modelo completo. Juntamente com esta análise foi realizado um experimento com a finalidade de verificar qual o impacto da abordagem de luminosidade sobre o modelo como um todo, já que em alguns casos esta abordagem é muito sensível aos fatores de ruídos, luminosidade e sombreamentos da imagem. A fim de atingir estes objetivos, foram gerados resultados sobre imagens seguindo o modelo completo e gerados resultados sem a fase da abordagem de luminosidade sobre as mesmas imagens. Esta análise está descrita na Seção 4.2. A terceira análise teve como objetivo verificar a sensibilidade do modelo quando diferentes usuários realizam o passo de inicialização (clique com o *mouse* sobre as articulações na imagem) e avaliar se este método influencia de forma a impactar sobre o resultado final obtido. A quarta análise foi baseada em uma pesquisa (*survey*) com o intuito de avaliar de forma visual se a pose correta gerada pelo modelo estava presente no conjunto de poses finais apresentadas ao usuário. Finalizando, foi realizada uma análise quantitativa através das poses humanas 3D recuperadas pelo modelo proposto quando comparadas com as poses 3D obtidas através do *hardware Microsoft Kinect* ©, o qual possui informação de profundidade e pode eliminar as possíveis ambiguidades.

---

<sup>1</sup><http://opencv.org/>

<sup>2</sup><http://www.opengl.org/>

#### 4.1 Análise sobre a recuperação da pose 3D

Esta primeira análise teve por finalidade verificar se a pose real em 3D da pessoa a partir de uma fotografia é encontrada no conjunto de poses 3D geradas pelo modelo proposto nesta tese. Essa análise foi realizada de forma qualitativa e quantitativa através de inspeção visual e verificação das poses corretas encontradas no conjunto de possíveis poses geradas pelo modelo. O processo do modelo foi realizado em um conjunto de 443 imagens, sendo aplicado de duas formas: Primeiramente, usando a Biomecânica juntamente com a abordagem de ordenação baseada no conforto e, também empregando o mesmo processo incluindo a análise da luminosidade. O experimento foi realizado desta forma a fim de verificar além da validação do modelo, também avaliar a abordagem da luminosidade no método, já que este tipo de abordagem é bastante suscetível a inúmeras variações, como por exemplo, posição da fonte de luz, sombras de outras pessoas ou objetos, etc.

No primeiro caso, o gráfico da Figura 4.1 apresenta como resultado o tamanho dos conjuntos de possíveis poses obtidas para cada imagem processada pelo modelo. Esses conjuntos variaram de uma única pose até 512 que é o conjunto máximo possível de poses para uma imagem após a aplicação das restrições biomecânicas. Para melhor visualização, os conjuntos com mais de 64 poses foram agrupados em uma única classe do histograma representada pela classe > 64. Pelo gráfico da Figura 4.1 é possível verificar que para 29 imagens do conjunto de 443 de entrada foram geradas apenas uma possível pose 3D, para 40 imagens foram geradas 2 possíveis poses e assim sucessivamente.

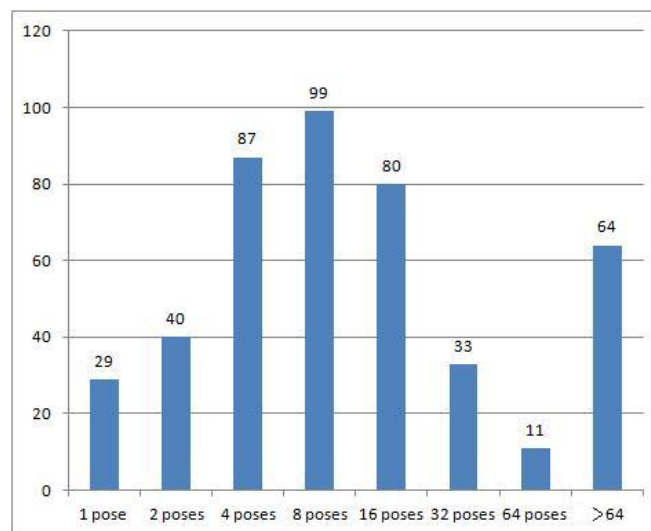


Figura 4.1: Resultado da quantidade de conjuntos de possíveis poses obtidos pelo modelo.

É importante salientar que em 13 imagens deste experimento a pose correta não foi obtida pelo modelo, portanto não estão no gráfico da Figura 4.1. Isto ocorreu nas imagens em que as posturas não eram frontais ou era poses em perspectiva. A Figura 4.2 mostra um exemplo destas imagens e o resultado da primeira pose gerada pelo modelo. O problema da perspectiva será discutido na Seção 6.1.

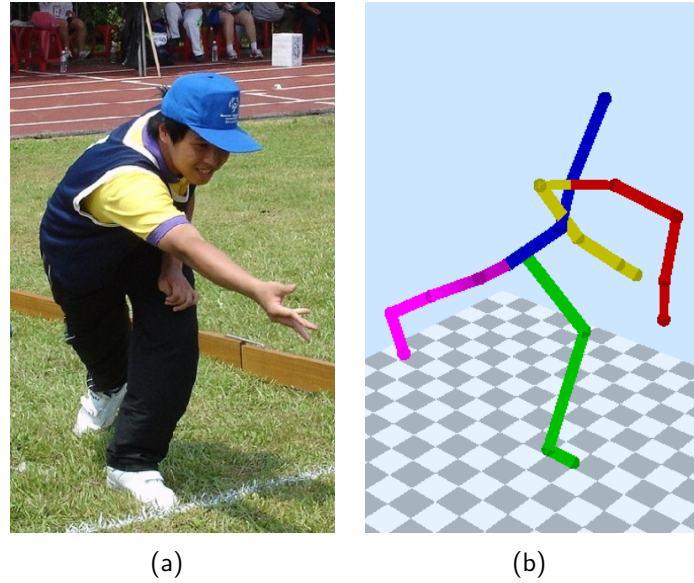


Figura 4.2: Resultado onde a pose real não foi definida corretamente devido a influência da perspectiva sobre o modelo de Taylor [Tay00]. A imagem (a) é a imagem de entrada, vista de um ângulo de cima e do lado direito da pessoa na foto. A imagem (b) é um dos resultados gerados pelo modelo do Taylor que é suscetível à perspectiva.

No gráfico da Figura 4.3 é mostrada em qual posição da classificação, a pose correta foi definida. É possível verificar que em 341 imagens a pose correta está entre as 10 primeiras posições classificadas pelo modelo e conferidas através do *ground truth*, representando 79% das imagens. Em 62 imagens a posição da melhor imagem ficou a partir da vigésima sétima colocação. Estas colocações maiores ocorrem normalmente em poses onde a pessoa está em uma atividade física, por exemplo correndo ou saltando, e Taylor [Tay00] gera um grande conjunto poses possíveis. Na curva cumulativa apresenta as posições em que foram obtidas as poses de acordo com o *ground truth*.

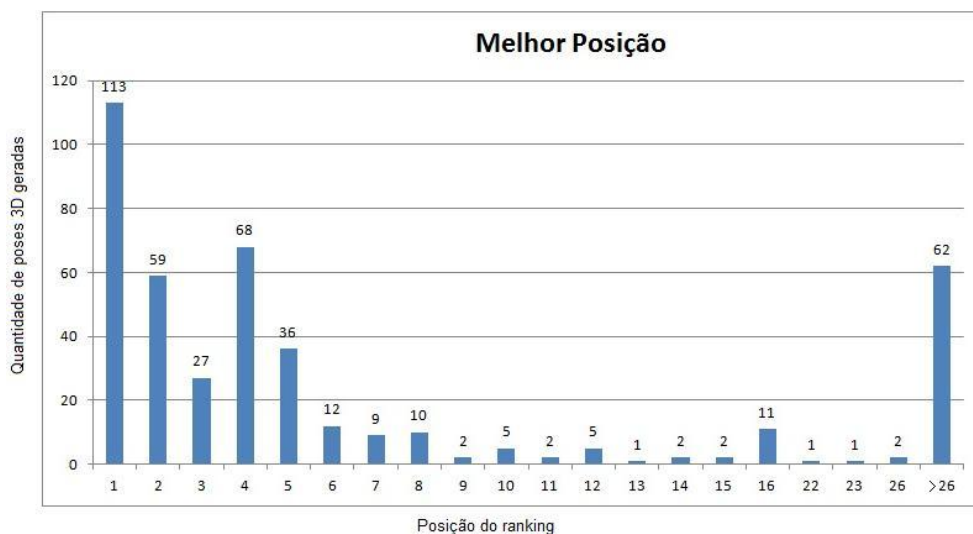


Figura 4.3: O Histograma mostra para cada posição da classificação quantas poses 3D geradas pelo modelo estão de acordo com a pose real na fotografia.

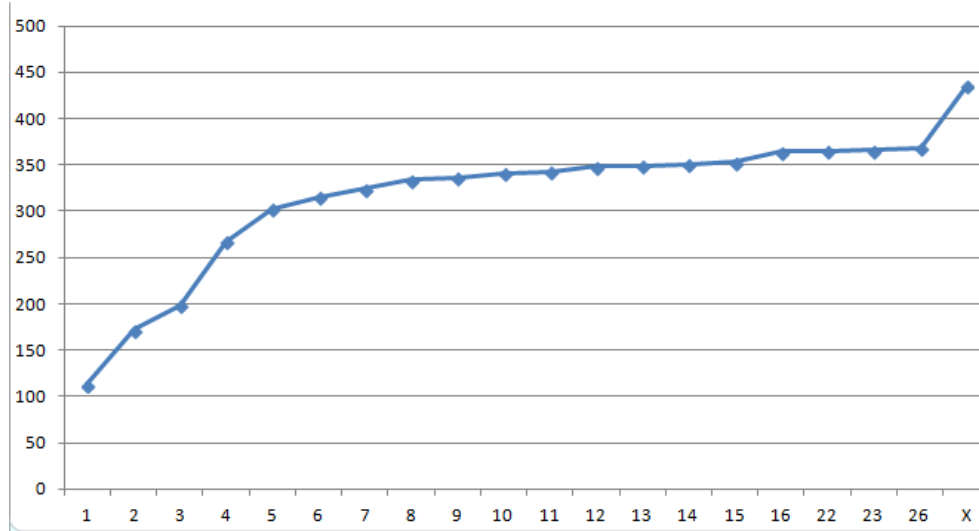


Figura 4.4: A curva cumulativa apresenta as posições onde foram encontradas as posturas de acordo com o *ground truth*.

Finalizando, as Figuras 4.5 e 4.6 apresentam alguns resultados obtidos pelo modelo proposto completo, mas somente o resultado da pose recuperada na primeira posição da classificação. A primeira coluna destas figuras é a imagem de entrada e a segunda e terceira coluna apresentam visões frontal e lateral do esqueleto recuperado, respectivamente.

## 4.2 Análise da abordagem de Luminosidade

A fim de verificar os efeitos obtidos pela abordagem da luminosidade, foi realizado o seguinte teste. De um conjunto de 28 imagens, aplicou-se o modelo proposto usando apenas a abordagem de conforto (RANKING 1) e depois, usou-se a abordagem do conforto juntamente com a de luminosidade (RANKING 2). Em seguida, verificou-se a posição da postura obtida pelo *ground truth* em ambas as classificações: quanto mais inicial a posição, melhor o resultado. Como mostrado na Tabela 4.1, O RANKING 2 proporcionou resultados melhores ou iguais ao RANKING 1 em 85% dos casos. Com esta análise verificou-se que apesar da abordagem de luminosidade ser uma técnica simples e poder ser suscetível a ruídos, sombras e a posição da fonte de luz no momento da obtenção da imagem, ela apresentou resultados satisfatórios.

## 4.3 Análise do método de inicialização do modelo

Esta análise teve por finalidade avaliar o impacto do método de inicialização descrito na Seção 3.1 nos resultados finais do modelo. Apesar do método de aquisição manual da postura 2D apresentar algumas vantagens de acordo com Hornung, Deckers e Kobbelt [HDK07], realizou-se este experimento a fim de verificar quanto a sensibilidade do clique nas imagens pode impactar no resultado do modelo, já que este procedimento pode variar de pessoa para pessoa. A seleção da posição das articulações  $J_i$  de uma pessoa na imagem pode variar de um usuário para outro. O

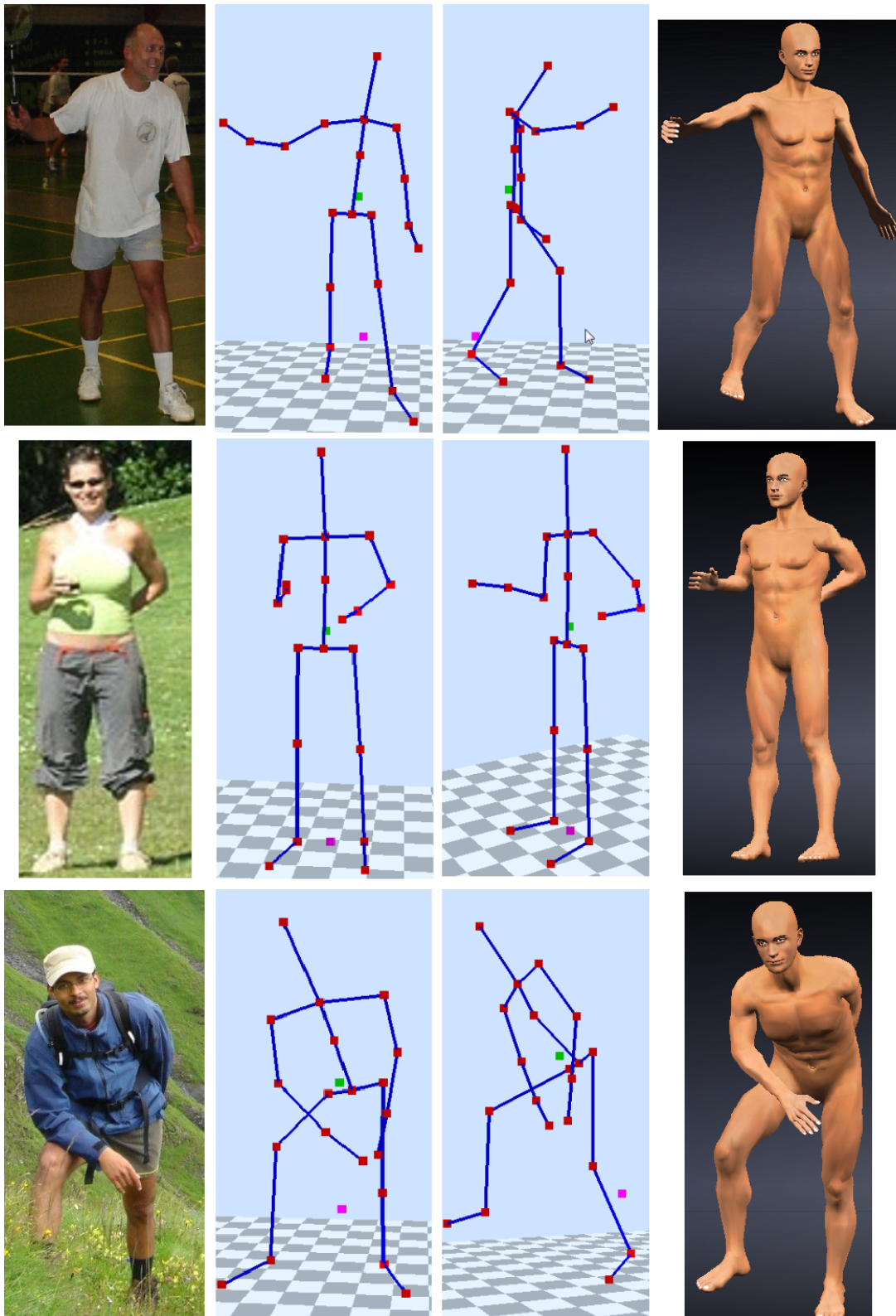


Figura 4.5: Resultados do esqueleto 3D obtido pelo modelo proposto classificado em 1º lugar. A primeira coluna representa a imagem de entrada, a segunda coluna é uma visão frontal do esqueleto gerado e a terceira coluna é a imagem de um ponto de vista lateral do esqueleto. A quarta coluna é um humano virtual gerado na mesma postura obtida pelo modelo.



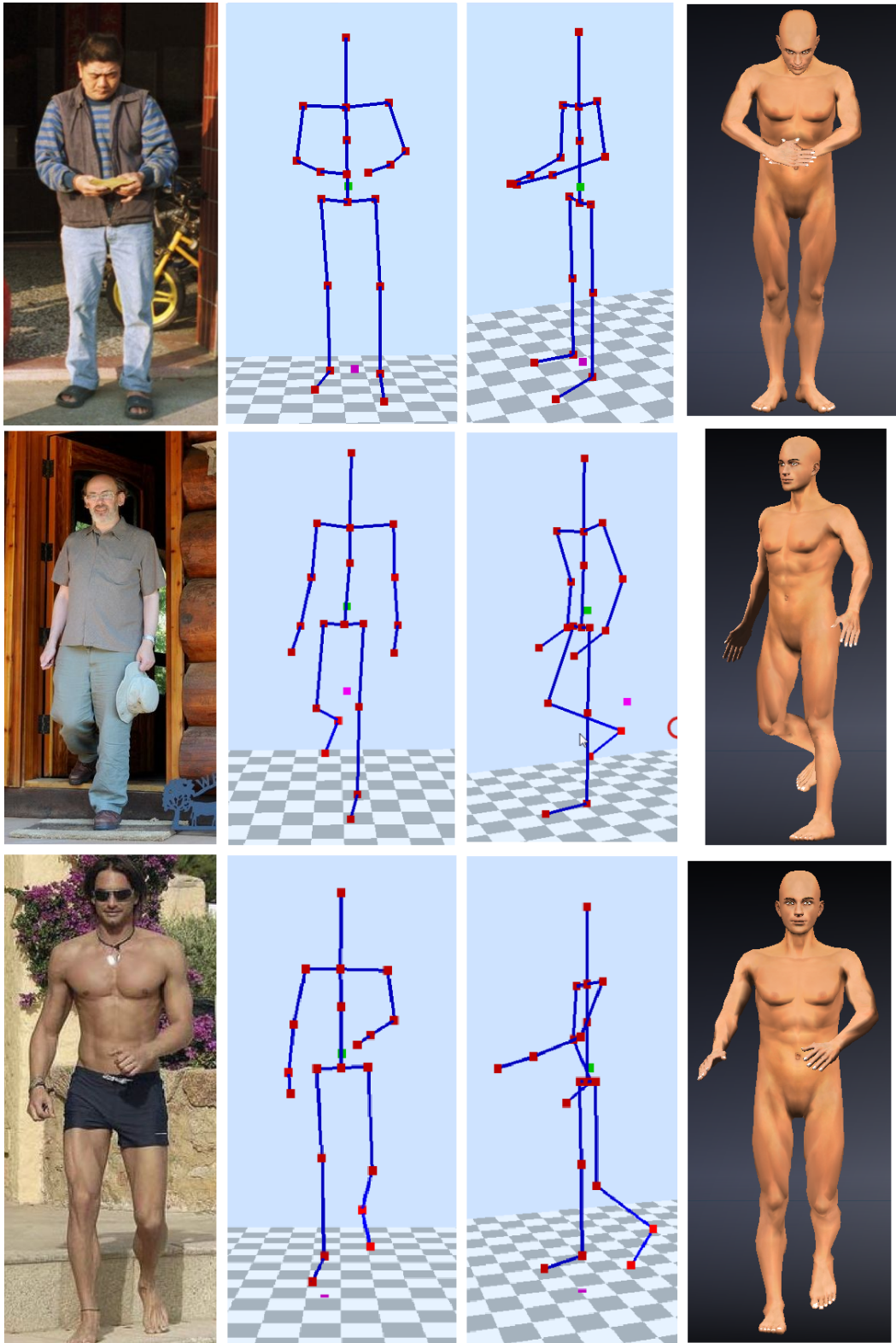


Figura 4.6: Outros resultados obtidos pelo modelo proposto. As colunas estão na mesma ordem da Figura 4.5.



Tabela 4.1: O impacto obtido pela abordagem da luminosidade na classificação. A primeira coluna indica a imagem de entrada, a segunda coluna mostra a posição da postura de acordo com o *ground truth* na classificação gerada pelo modelo usando apenas abordagem de conforto. Na terceira coluna são as posições na classificação gerados pelo modelo com a abordagem de conforto junto com a abordagem de luminosidade. A quarta coluna é a situação comparativa de melhora ou não no posicionamento.

Imagem de entrada	RANKING 1	RANKING 2	Situação
0	4	4	✓
1	6	3	✓
2	3	3	✓
3	1	1	✓
4	1	1	✓
5	2	2	✓
6	5	8	✗
7	4	1	✓
8	16	8	✓
9	4	1	✓
10	8	3	✓
11	4	1	✓
12	2	1	✓
13	1	1	✓
14	4	3	✓
15	3	1	✓
16	16	4	✓
17	4	3	✓
18	1	1	✓
19	1	1	✓
20	3	2	✓
21	1	1	✓
22	1	2	✗
23	4	3	✓
24	5	1	✓
25	1	2	✗
26	1	2	✗
27	3	2	✓

experimento seguiu os seguintes passos: foi solicitado a 5 pessoas a realização do procedimento de obtenção da pose bidimensional através de cliques na imagem. Para cada pessoa foi solicitada a realização deste processo para cinco fotografias com pessoas em poses diferentes. Essas imagens

podem ser visualizadas na Figura 4.7.

A partir da pose 2D obtida de cada fotografia pelo usuário foram analisados os resultados obtidos pelo modelo e uma comparação dos conjuntos de possíveis poses geradas pelo modelo. Foi observado que apesar de haver uma variação na quantidade de poses finais geradas pelo modelo, o mesmo ainda gerou a pose real correta de acordo com o *ground truth* em todos os conjuntos de possíveis poses obtidas. Houve também uma variação da posição da postura correta na classificação proposta pelo modelo. Estes resultados estão descritos na tabela 4.2. As colunas "QT\_G" de cada usuário é a quantidade de poses possíveis geradas pelo modelo através do seu modo de inicialização. As colunas "Pos" indicam a posição da pose correta na classificação de acordo com o *ground truth*. Salienta-se que esta variação apesar de afetar a posição da pose correta na classificação, foi causada pela diferença da altura das pessoas em pixels obtidas de forma manual pelo usuário. Isto afeta o modelo de Taylor, nos tamanhos esperados de cada osso.

Imagens	Us 1		Us 2		Us 3		Us 4		Us 5	
	QT_G	Pos	QT_G	Pos	QT_G	Pos	QT_G	Pos	QT_G	Pos
Imagem a	2	1 <sup>a</sup>	2	1 <sup>a</sup>	4	1 <sup>a</sup>	8	1 <sup>a</sup>	2	1 <sup>a</sup>
Imagem b	1	1 <sup>a</sup>	1	1 <sup>a</sup>	4	4 <sup>a</sup>	8	7 <sup>a</sup>	2	1 <sup>a</sup>
Imagem c	4	1 <sup>a</sup>	8	1 <sup>a</sup>	32	1 <sup>a</sup>	4	1 <sup>a</sup>	4	2 <sup>a</sup>
Imagem d	8	3 <sup>a</sup>	4	1 <sup>a</sup>	4	2 <sup>a</sup>	4	1 <sup>a</sup>	4	2 <sup>a</sup>
Imagem e	8	6 <sup>a</sup>	8	4 <sup>a</sup>	8	2 <sup>a</sup>	4	2 <sup>a</sup>	4	1 <sup>a</sup>

Tabela 4.2: Resultado da análise dos cliques no processo de inicialização do modelo por cinco usuários voluntários. As colunas QT\_G de cada usuário é a quantidade de poses possíveis geradas pelo modelo através do seu modo de inicialização. As colunas Pos indicam a posição da pose correta na classificação de acordo com o *ground truth*.

A Figura 4.7 apresenta os resultados dos cliques no processo inicial dos cinco usuários voluntários nas cinco imagens selecionadas para o experimento.

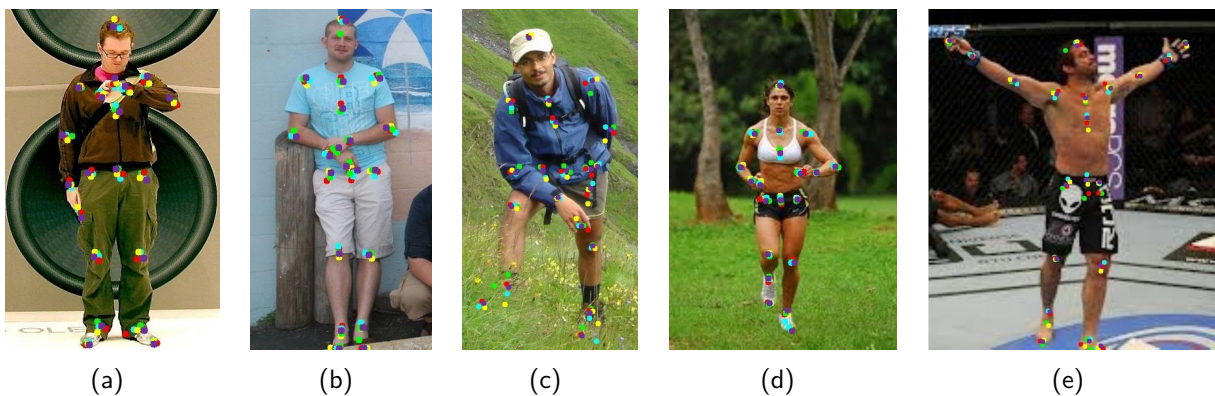


Figura 4.7: Resultado dos cliques realizados nas articulações de quatro pessoas a fim de verificar a sensibilidade do processo de inicialização do modelo proposto.

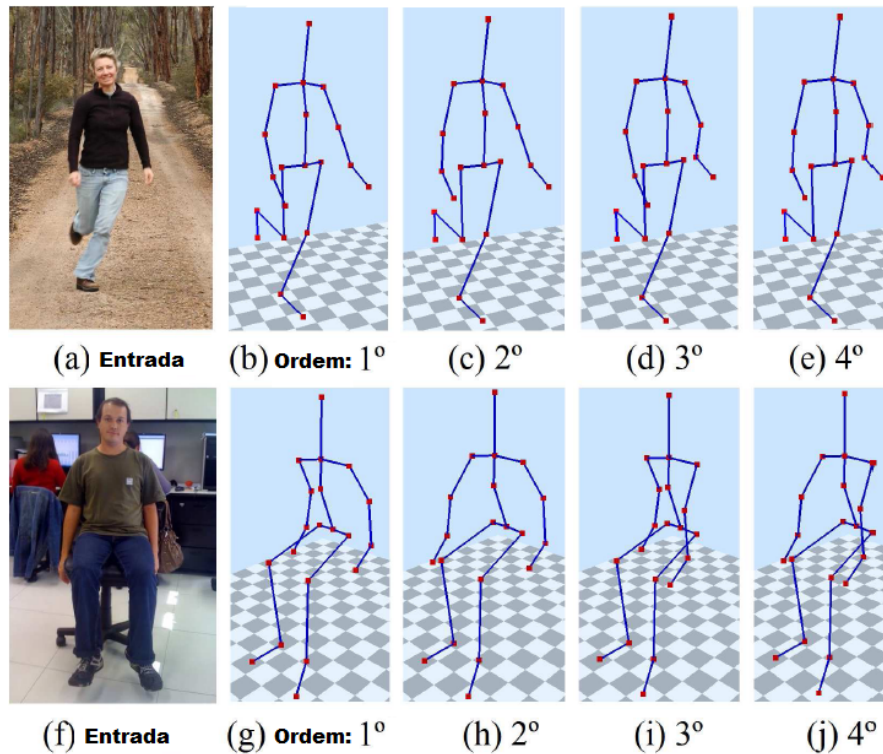


Figura 4.8: Exemplo das imagens de entrada e as poses ordenadas. As poses são mostradas em pontos diferentes de visão a fim de ter uma noção melhor da postura 3D.

#### 4.4 Análise através de pesquisa

A fim de avaliar as poses 3D geradas pelo modelo, foi realizada uma pesquisa composta por 10 perguntas com voluntários. Em cada pergunta, uma imagem com uma pessoa foi mostrada, junto com as primeiras quatro poses na classificação gerado pelo modelo, como mostrado na Figura 4.8. Os voluntários foram solicitados a classificar as posturas 3D que, em sua opinião, correspondiam a pose da pessoa na foto. As pessoas também poderiam rejeitar uma pose ou mesmo todas geradas pelo sistema. Na análise dos resultados, considerou-se o resultado como correto (acerto), se a postura classificada pelo sistema como a melhor também era classificada pelo usuário como a melhor ou a segunda melhor para a pose da pessoa na foto. Decidiu-se considerar a segunda posição na classificação, porque em muitos casos, as poses geradas são muito semelhantes. De acordo com este critério, dada as 100 respostas (10 pessoas responderam 10 questões cada), obteve-se uma média de 77% de acertos entre a escolha dos voluntários com a escolha do modelo. A Figura 4.8 apresenta dois resultados presentes no questionário. A Figura 4.8 (a:e) mostra os piores resultados obtidos em comparação com as opiniões dos voluntários, devido ao fato de que as poses recuperadas são bastante similares, pois as diferenças entre os quatro esqueletos ordenados estão localizados somente nos braços. A Figura 4.8 (f:j) mostra uma imagem com um percentual elevado (100%) de respostas corretas, mesmo que a diferença esteja também somente nos braços, a percepção dessa diferença é mais notória nestas poses. O teste teve por finalidade gerar resultados e obter uma avaliação visual desses resultados.

#### 4.5 Análise utilizando o sensor *Kinect* © para obter o *Ground Truth*

Outra análise realizada foi a comparação das posições das articulações do esqueleto gerado pelo modelo proposto nesta tese com o software de rastreamento do esqueleto fornecido pelo *Microsoft Kinect SDK* <sup>3</sup>. O sensor *Kinect* foi lançado pela *Microsoft* no ano de 2010 para criar uma maior interação do usuário com os jogos do videogame *Xbox 360 Microsoft* ©. Este sensor é composto por uma câmera *RGB*, que capta a imagem do usuário e uma câmera infravermelha que com o auxílio de um projetor de raios infravermelhos captura a profundidade da imagem e consequentemente o usuário em *3D*. O Kit de Desenvolvimento de Software - *Software Development Kit* -(*SDK*) fornece um software denominado *skeletal tracking* que através das câmeras, rastreia até dois usuários em *3D* conforme a Figura 4.9.

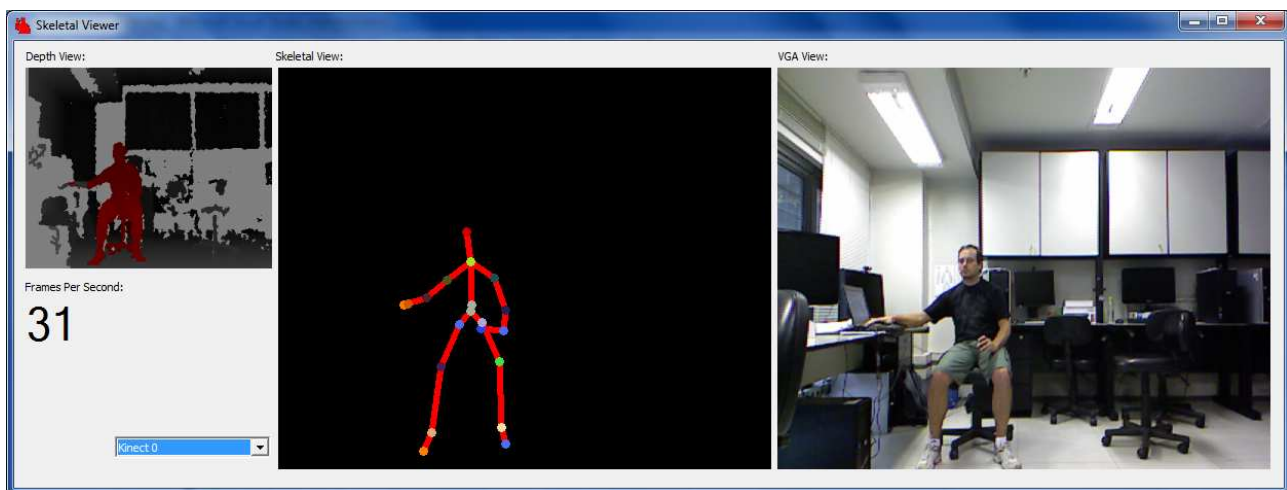


Figura 4.9: Software fornecido pelo *Microsoft Kinect SDK* fazendo o rastreamento de usuários pelo seu sensor.

Desta forma foi realizado o seguinte experimento: utilizando o *hardware Kinect* e o software *skeletal tracking* identificou-se, em 25 quadros de vídeos da câmera do *Kinect* onde foram capturadas pessoas em diversas poses, o esqueleto 3D e o valor de  $\Delta Z$  para cada junta  $i$ , gerando o *ground truth*. Depois, para esses mesmos 25 quadros, aplicou-se o modelo desta tese e calculou-se a classificação de poses de acordo com o conforto. Por fim, verificou-se qual das poses na classificação gerada está de acordo com o esqueleto obtido pelo *Microsoft Kinect*®. A Figura 4.10 apresenta os resultados quantitativos. Os dados horizontais são posições da classificação e os dados verticais são as quantidades de posturas que está de acordo com o *ground truth*. Pode-se observar no gráfico que em 92% das imagens, o modelo classificou a pose correta de acordo com o *ground truth* nas cinco primeiras posições da classificação gerada. Se computadas apenas as três primeiras posições, o modelo classificou as posturas corretamente em 64% dos casos. É importante ressaltar que neste processo foi usado somente as restrições biomecânicas e a abordagem de conforto do modelo em razão das imagens que foram obtidas pelo *Kinect* serem de baixa resolução ( $480 \times 390$  pixels) e

<sup>3</sup><http://www.microsoft.com/en-us/kinectforwindows>

muito ruidosas o que faz com que as informações de luminosidade não possam ser consideradas. É importante salientar que uma pose gerada pelo modelo proposto foi considerada "de acordo" com o *ground truth*, se todos os sinais de  $\Delta Z$  da articulação  $i$  da pose estavam de acordo com os sinais de  $\Delta Z$  da articulação  $i$  do esqueleto fornecido pelo software *skeletal tracking*.

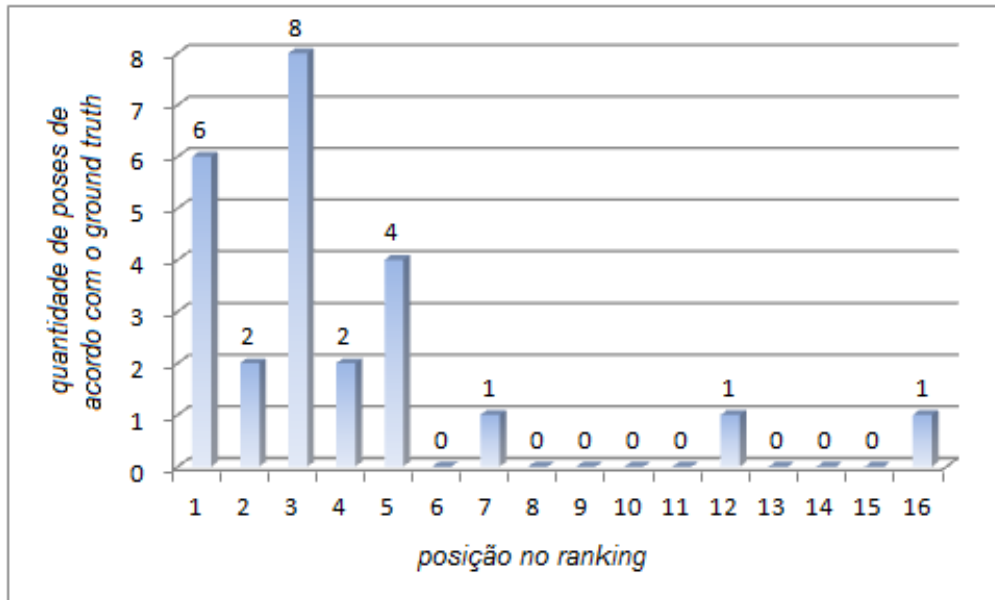


Figura 4.10: Resultados da localização das posturas na classificação de acordo com o *ground truth* obtido pela Microsoft Kinect<sup>®</sup>.

No próximo Capítulo serão apresentadas duas aplicações que empregaram o modelo proposto nesta tese. Estas aplicações tiveram por objetivo além da validação do modelo a verificação da sua aplicabilidade.



## 5. APLICAÇÃO DO MODELO PROPOSTO

A fim de analisar a aplicabilidade do modelo proposto nesta tese, foi proposta uma abordagem para minimizar o problema da ambiguidade em posturas usando o modelo de detecção de auto-occlusão obtida pelo processo de segmentação desenvolvido por Jacques Junior et al. [JJM12]. Verificou-se que se o modelo de detecção de auto-occlusão é capaz de definir automaticamente a sequência dessa auto-occlusão dos membros, ou seja, a ordem que estes membros estão a frente da câmera durante o processo de obtenção da imagem, então é possível determinar quais as poses são válidas a partir de uma determinada imagem eliminando poses ambíguas. Esta aplicação é descrita na Seção 5.1.

### 5.1 Estimativa da pose 3D utilizando informação de auto-occlusão

Durante o desenvolvimento desta tese verificou-se a possibilidade de minimizar o problema da geração de poses 3D ambíguas a partir de uma imagem, utilizando um modelo de estimativa de detecção de auto-occlusão (*self-occlusion*). Inicialmente, o modelo de segmentação proposto por Jacques Junior et al. [JJM12] é usado para capturar partes rotuladas do corpo humano em uma imagem 2D. Em seguida, as partes do corpo oclusas são detectadas quando outras partes diferentes do corpo se sobrepõem. Elimina-se a ambiguidade das posturas analisando a energia dos contornos correspondentes em torno dos pontos de intersecção entre estes membros. Os resultados estimados das oclusões alimentam o algoritmo de estimação de pose que reconstrói um conjunto de posturas plausíveis em 3D. Os resultados experimentais indicam que a técnica proposta funciona bem em imagens não triviais, pois estimam corretamente as partes do corpo oclusas reduzindo o número de possíveis posturas 3D. As próximas seções apresentam detalhes desta pesquisa.

#### 5.1.1 Auto-occlusão

A auto-occlusão (*self-occlusion*) em poses humanas é um problema clássico em visão computacional, atualmente existem diversas abordagens que buscam solucionar deste problema. Sigal e Black [SB06] apresentaram um modelo para estimar poses 2D que emprega probabilidades locais da imagem que aproximam da probabilidade global através da contagem das oclusões e explicações concorrentes de provas na imagem em várias partes. A abordagem proposta por Huang e Yang [HY10] usa um modelo de regressão para aprender o mapeamento do espaço de características da imagem para o espaço da pose, mas difere de [SB06], em que as representações esparsas são aprendidas a partir de exemplos com a capacidade demonstrada para lidar com oclusões. Radwan et al. [RDJG12] usou um modelo de regressão gaussiana para aprender os parâmetros de partes do corpo oclusos. No modelo de Kim e Kim [KK10] é detectado se uma dada parte do corpo sofre oclusão ou não analisando os valores próprios dos dados 3D da imagem recolhidos a partir do ponto de articulação de cada parte do corpo. Em [CYL12], os autores propuseram um método de estimação do estado

de auto-occlusão. Em sua abordagem, um campo aleatório de Markov é utilizado para modelar o estado de oclusão que representa a ordem de profundidade entre duas partes simétricas do corpo humano, como braço esquerdo com braço direito ou perna esquerda com perna direita.

Neste aplicação é empregada uma abordagem de detecção de auto-occlusão proposta por Jacques Junior et al. [JDJM13] e a informação obtida neste processo irá refletir sobre o problema da ambiguidade de poses 3D que possuem a mesma projeção 2D, descrita no Capítulo 3. A partir dos pontos obtidos na imagem 2D, que representam as articulações do corpo humano conforme o processo descrito na seção 3.1, o objetivo é encontrar as partes do corpo que possuem auto-occlusão e que fornecem a informação necessária para remover a ambiguidade do modelo em determinadas situações. Por exemplo, se o braço esquerdo da pessoa está sendo ocluído pelo braço direito, esta informação indica que o braço direito está mais próximo à câmera do que o braço esquerdo, logo é possível descartar posturas em que isto não seja verdade.

O primeiro passo da abordagem é associar um modelo de esqueleto 2D, conforme a (Figure 5.1(a)), na pessoa na fotografia (Figure 5.1(b), em ciano). Este estágio pode ser feito manualmente ou automaticamente, dependendo da aplicação. O segundo estágio é a segmentação da pessoa na imagem. Para este propósito foi utilizado o algoritmo proposto por Jacques Junior et al. [JJM12], o qual segmenta a pessoa com informação semântica baseada no valor do contorno da energia para cada parte do corpo de acordo com a informação do gradiente, coerência dos ossos para cada parte do corpo e distâncias antropométricas. Tal modelo produz um contorno fechado, onde cada ponto do contorno é associado à uma parte específica do corpo conforme ilustrado pela Figura 5.1(b) usando-se diferentes cores. O terceiro passo do modelo é identificar as intersecções das partes do corpo em um alto nível, por exemplo, o braço e o torso estão se interceptando, conforme ilustra a Figura 5.1(c), caracterizando a oclusão. O quarto estágio é analisar em baixo nível cada intersecção candidata (os pontos em vermelho mostrados na Figura 5.1(c)) para identificar qual parte do corpo está sob oclusão. A desambiguação das partes ocluídas do corpo é feita por um procedimento de avaliação da função de energia do contorno na vizinhança dos pontos de intersecção. A saída do modelo é uma lista de pares de intersecção e informação de auto-occlusão, por exemplo, o braço direito está ocluído pela mão esquerda (Figura 5.1(d)). Esta informação irá alimentar o modelo de recuperação da pose 3D, que é usado para reconstruir a pose 3D (Figura 5.1(e)). O modelo para a obtenção e tratamento das auto-occlusões foi desenvolvido por Jacques Junior et al. A descrição deste modelo está especificada no artigo [JDJM13]. Na próxima seção é descrito como é minimizada a ambiguidade do modelo de Taylor utilizando esta abordagem.

### 5.1.2 Minimização da ambiguidade do modelo de Taylor usando a detecção de auto-occlusão

Com a informação da ordem de oclusão entre os membros pode-se aplicar uma abordagem para minimizar o problema de ambiguidade gerado pelo modelo de Taylor [Tay00]. Para cada pose calcula-se a posição 2D do cruzamento entre os dois ossos envolvidos na detecção de auto-occlusão. Com base nos níveis de profundidade (coordenadas  $z$ ) de cada um desses ossos neste ponto de intersecção, pode-se verificar se a postura 3D está de acordo com o resultado de auto-occlusão. Se



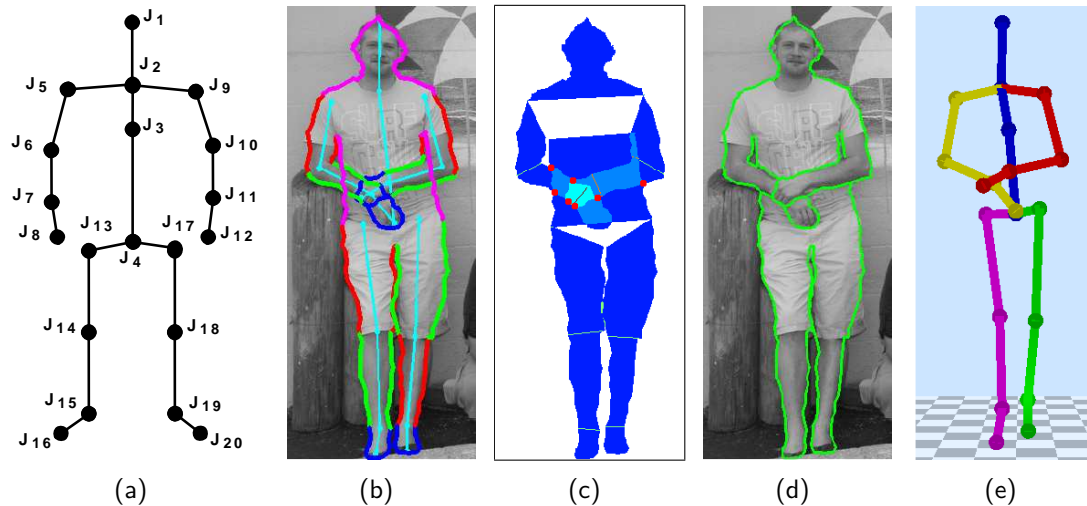


Figura 5.1: Visão geral do modelo proposto. (a) Modelo de esqueleto proposto. (b) O resultado da segmentação. (c) As intersecções entre as partes do corpo e os pontos de intersecção (em vermelho). (d) Ilustração do resultado da estimação de auto-occlusão. (e) A pose 3D estimada.

não, a postura é descartada. Um exemplo é mostrado na Figura 5.2. A Figura 5.2 (a) ilustra a imagem de entrada com as partes detectadas do corpo e de detecção de oclusão. Figura 5.2 (b) ilustra uma das três poses 3D incorretas obtidas usando apenas a abordagem de Taylor e as restrições biomecânicas (o antebraço esquerdo está por trás do tronco). A adição da informação de auto-occlusão descarta as poses 3D incorretas, e apenas uma pose 3D (correta) permanece, ilustrado na Figura 5.2 (c).

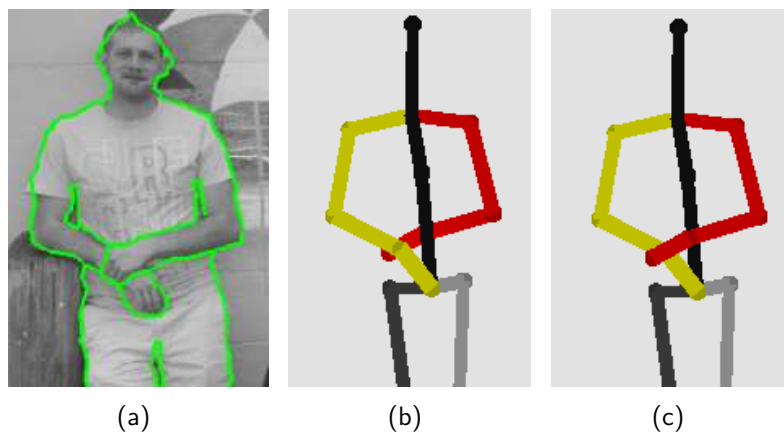
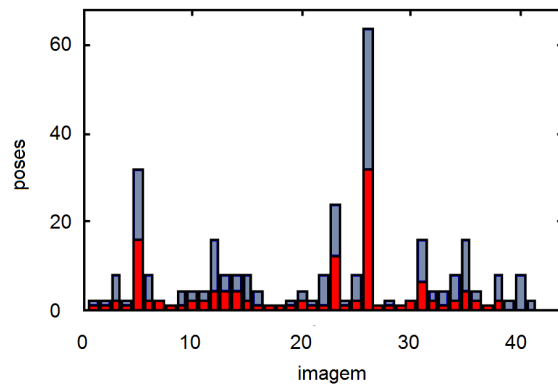


Figura 5.2: (a) Auto-occlusão detectada. (b) Uma das possíveis poses incorretas gerada apenas com as restrições biomecânicas, pois o braço esquerdo da pessoa está por trás do torso. (c) A única pose obtida explorando a abordagem de auto-occlusão.

A Figura 5.3 mostra um gráfico com dados sobre minimização da quantidade de poses geradas a partir do modelo de auto-occlusão.

Foram realizados testes em imagens em um grupo de 41 imagens contendo auto-occlusão. Para a análise quantitativa do modelo de obtenção de auto-occlusão, foi gerado o *ground truth* manualmente, no qual a informação de auto-occlusão é anotada em cada imagem em alto nível, o qual pares de



(a)

Figura 5.3: Número de poses 3D geradas por imagem, usando somente restrições biomecânicas (azul) e incluindo o modelo de auto-oclusão (vermelho).

partes do corpo estão em conflito e também em baixo nível no qual determina qual parte está em frente de outra parte. No Anexo 1 é possível verificar o resultado final das 41 imagens. O número total de oclusões na base de dados, observado pelo usuário, foi de 106, e a abordagem de detecção de auto-oclusão proposta identificou corretamente 89.6% das auto-oclusões, das quais 83.16% representam ambiguidades eliminadas corretamente. O procedimento de detecção também obteve em torno de 10% de falsos positivos, ou seja, o resultado das detecções não estavam de acordo com os dados do *ground truth*. É importante salientar que todo o cálculo foi feito utilizando o valor do pixels em escala de cinza das imagens de acordo com o trabalho apresentado por Jacques et al [JJM12]), e o uso das imagens coloridas poderia melhorar os resultados da segmentação bem como à estimativa da auto-oclusão.

Quanto ao problema de recuperação de poses 3D, a validação foi realizada por inspeção visual, uma vez que não se tem a posição exata da postura 3D que melhor se refere a uma determinada imagem 2D. A Figura 5.3 ilustra o número de poses possíveis detectadas utilizando apenas restrições biomecânicas (barras em azul) e utilizando ambas as técnicas de restrições biomecânicas e os resultados da auto-oclusão (barras vermelhas). O número de posturas restantes depois da aplicação dos métodos nestes dois cenários são 7.52 e 3.04, respectivamente. Considerando a abordagem integral (biomecânica + oclusão), a pose correta fazia parte do conjunto de posturas selecionadas pelo modelo em 70.7% dos casos. A exclusão da postura correta acontece devido a erro na parte de segmentação corporal inicial e na detecção/desambiguação da oclusão. Além disso, em alguns casos (por exemplo, a imagem da última linha da Tabela 5.2), as restrições biomecânicas geraram um conflito com a detecção de auto-oclusão, resultando em um conjunto vazio de possíveis posturas.

Finalizando, as Figuras 5.4 apresentam alguns resultados finais de poses humanas 3D obtidas a partir do modelo de segmentação e da detecção de auto-oclusão. As Figuras 5.4(a),(c) e (e) mostram os resultados da segmentação com a detecção de auto-oclusão e as Figuras 5.4 (b), (d) e (f) são as posturas humanas em 3D recuperadas. Este modelo foi desenvolvido por Jacques et al [JDJM13].

Na próxima subseção foi feita uma análise a fim de verificar o melhoramento dos resultados

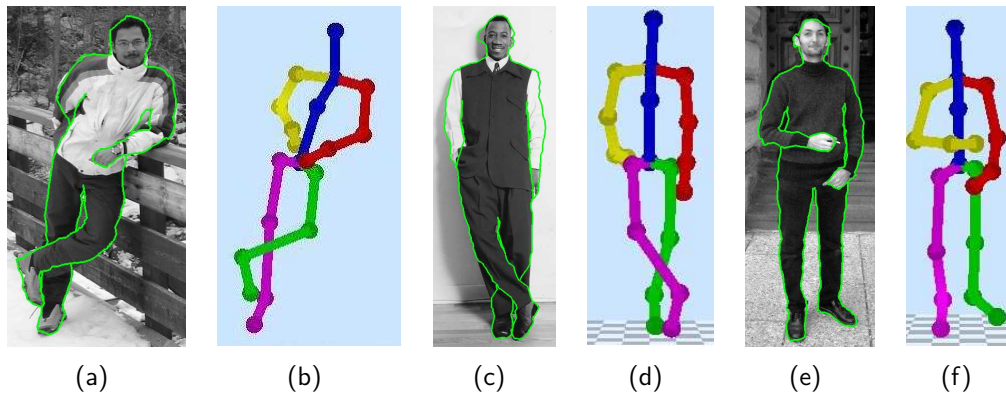


Figura 5.4: A detecção de auto-occlusão e os resultados da recuperação da pose 3D.

obtidos através do modelo proposto com o auxílio da abordagem de obtenção de auto-occlusões em relação aos membros das pessoas.

### 5.1.3 Análise dos resultados utilizando a técnica de auto-occlusão

Nesta seção são apresentados os resultados experimentais do modelo proposto baseado na auto-occlusão. Nesta aplicação não foram utilizadas as abordagens do conforto da postura e a análise da luminosidade por que estas abordagens não influenciam na quantidade de poses finais obtidas, somente fazem uma classificação das mesmas. As Tabelas 5.1 e 5.2 apresentam os resultados obtidos. Para a compreensão desses resultados, deve ser considerada as seguintes informações a respeito das imagens mostradas nas tabelas:

- Cada linha mostra uma imagem analisada;
- Na primeira coluna são visualizados os esqueletos 2D de entrada (ciano) e o resultado da segmentação (linhas verde e vermelha - as vermelhas indicam as conexões das partes adjacentes do corpo);
- Os resultados do procedimento de estimativa de auto-occlusão são mostrados na segunda coluna (a visualização da intersecção entre as partes do corpo, em alto nível, bem como o exemplo da estimativa em baixo nível);
- As três colunas seguintes da tabela mostram, respectivamente, (i) o número de posturas 3D estimadas usando somente as restrições biomecânicas, (ii), o número de posturas 3D estimadas incluindo as informações de auto-occlusão, e (iii) se a pose correta, avaliada por inspeção visual, está ou não no conjunto gerado. É importante salientar que as restrições biomecânicas não removem a pose correta, mas uma análise errada da auto-occlusão pode descartar a pose correta;
- As duas últimas colunas mostram, respectivamente, a postura correta e uma outra pose gerada (incorreta), caso esta exista depois de todo processo realizado. As posturas são mostradas em três diferentes pontos de vista;

- É possível observar que algumas posturas geradas estão incoerentes devido a problemas de perspectiva, por exemplo, as pernas dobradas nos resultados das colunas  $C6$  e  $C7$  da Tabela 5.1 da primeira imagem. Isto ocorre devido ao modelo do Taylor [Tay00] e as medidas antropométricas que foram empregadas na qual utilizou-se medidas de uma pessoa média de acordo com Tilley [Til02], o que pode impactar sobre o modelo quando se aplica a imagens de crianças ou pessoas muito altas.

As imagens foram selecionadas aleatoriamente e mostram uma grande variedade de pessoas, aparências, poses, qualidade de imagem e portanto complexidade. A base de dados inclui algumas imagens públicas obtidas de outros trabalhos [FMJZ08, BM09, DT05], bem como algumas imagens adquiridas durante o desenvolvimento desta tese. Os rótulos das colunas são, respectivamente, como se segue:

- C1: Resultado da inicialização do esqueleto e da segmentação;
- C2: Resultado do processo de detecção de auto-occlusão;
- C3: Número de poses após as restrições biomecânicas;
- C4: Número de poses quando se utiliza a informação de auto-occlusão;
- C5: A pose correta é gerada para esta entrada?
- C6: A pose correta verificada por inspeção visual (frontal, vista diagonal e lateral);
- C7: Outra pose gerada (frontal, vista diagonal e lateral).

C1	C2	C3	C4	C5	C6			C7		
		2	1	S						
		2	1	N						
		8	2	S						
		2	1	S						
		8	2	S						
		2	2	S						
		1	1	S						
		4	1	S						
		4	2	N						

Tabela 5.1: Resultados obtidos utilizando a abordagem de auto-occlusão.



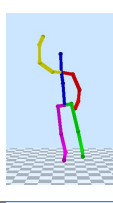
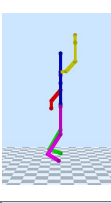
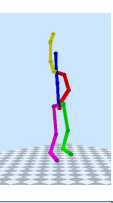
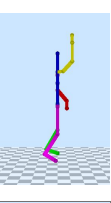


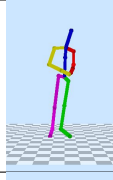
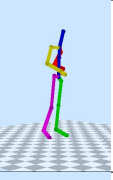
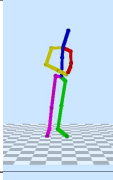
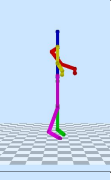


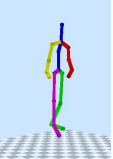
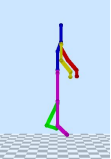
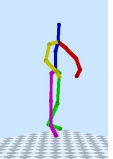
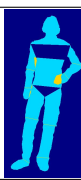
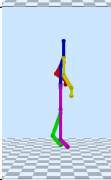
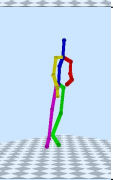
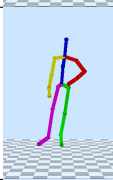
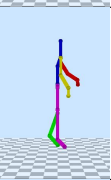
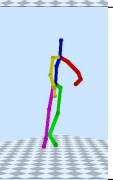



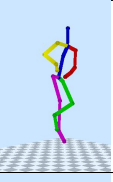

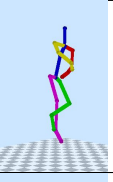

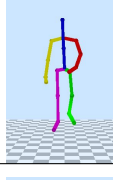
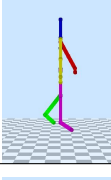
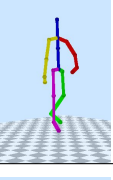
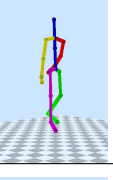

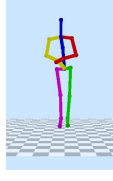
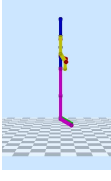
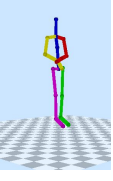
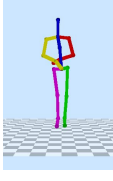
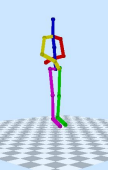
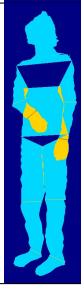
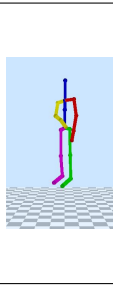
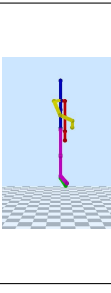
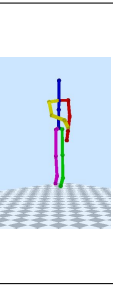
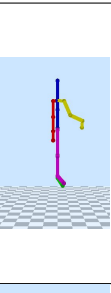
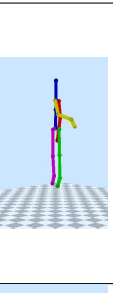



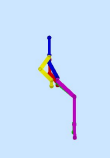
C1	C2	C3	C4	C5	C6			C7		
		4	2	S						
		16	4	N						
		8	4	S						
		8	4	S						
		8	2	S						
		4	2	S						
		4	1	S						
		8	2	S						
		8	0	N						

Tabela 5.2: Mais resultados obtidos utilizando a abordagem de auto-oclução. A última imagem não obteve a postura correta por motivo da projeção errada.

## 6. CONSIDERAÇÕES FINAIS E PERSPECTIVAS

Esta tese descreveu um modelo para a recuperação de poses em 3D a partir de uma única imagem 2D. Verificou-se no estado da arte que a maioria dos problemas impostos nesta área estão em aberto, não existindo uma solução definitiva para o tema. As dificuldades como por exemplo, perspectiva, ambiguidade nas poses, luminosidade, ruídos, oclusões parciais, vestuários complexos dentre outras, tornam o problema extremamente complexo. O modelo apresentado nesta tese é baseado em projeção ortográfica que gera um conjunto de poses ambíguas mas é minimizado por técnicas de restrições biomecânicas. O modelo também obtém uma classificação do conjunto final das poses geradas através de uma abordagem baseada no conforto das posturas das pessoas na imagem. Esta classificação é ainda melhorada através de um processo que analisa a luminosidade das partes do corpo da pessoa na fotografia penalizando as poses que não são coerentes com essa análise. Este processo para o melhoramento da classificação das posturas obtidas, apesar de ser bastante sensível a determinadas situações do ambiente, como ruídos, posição da fonte de luz e posição da pessoa na imagem, ainda assim mostrou-se eficaz conforme pode ser visualizado nos resultados obtidos pela análise descrita na seção 4.2.

Com o objetivo de avaliar o modelo proposto nesta tese foram desenvolvidas 5 análises diferentes. Inicialmente, verificou-se a quantidade de poses corretas e as suas posições na classificação determinada pelo modelo completo. Também foi feita uma análise do impacto da abordagem de luminosidade sobre o modelo. A terceira análise foi a verificação da influência da fase inicial de obtenção da pose humana através dos cliques manuais na imagem. Uma quarta análise foi feita em forma de questionário a fim de determinar de forma visual a coerência das poses geradas pelo modelo e o que os usuários definiam como pose correta. E finalizando foram analisados os dados obtidos através de um comparativo entre o resultado do modelo proposto nesta tese e os dados obtidos através do *hardware Microsoft Kinect*.

Durante o desenvolvimento da tese também foi proposta uma aplicação utilizando os conceitos do modelo proposto com a finalidade de obter uma avaliação da sua aplicabilidade. Esta aplicação é baseada em uma abordagem de segmentação de pessoas em imagens e na detecção de auto-occlusão dos membros destas pessoas. Com base nestas informações e usando o modelo de geração de posturas 3D juntamente com as restrições biomecânicas definiu-se uma abordagem com a finalidade de minimizar o problema da ambiguidade gerada pela projeção da imagem 2D para a obtenção da postura humana 3D.

### 6.1 Deficiências do modelo

Verificou-se através dos experimentos realizados no período das análises que determinadas posturas geradas pelo modelo possuem divergência com a realidade. Este problema ocorre nas seguintes situações:



- Em imagens de pessoas em movimento, normalmente durante a prática de esportes, onde o centro de apoio determinado pelo sistema não coincide com a realidade. Este problema não influencia na obtenção final da postura correta, mas afeta sua posição na ordenação relativa a abordagem do conforto. A Figura 6.1 mostra uma imagem de entrada (a) de uma pose onde o centro de apoio não está de acordo com a realidade mas o modelo gera a pose correta, no caso em 5ª posição da classificação (b);
- Imagens com poses em perspectiva, conforme a Figura 6.2 (a) e (b), afetam o resultado do modelo proposto nesta tese, que utiliza projeção ortográfica, ver Figura 6.3. Apesar de existir em modelos no estado da arte que minimizam esta deficiência [BK00, LC85], este problema não foi abordado devido ao escopo do trabalho proposto, ficando esta tarefa como um objetivo para trabalhos futuros.
- Outro problema detectado é que a abordagem da luminosidade sofre muito com ruídos da imagem, e a posição da fonte de luz não estando na frente da pessoa. Novamente, este problema não afeta a recuperação final da pose correta, mas somente a sua localização na classificação. Um exemplo desta limitação pode ser vista na Figura 6.4;
- Posturas em que a pessoa não está de frente para o dispositivo de obtenção da imagem. Neste caso ocorrem distorções na criação das poses, visto que o modelo das restrições biomecânicas impostas é baseado na posição da pessoa, como ilustra também a Figura 6.4;

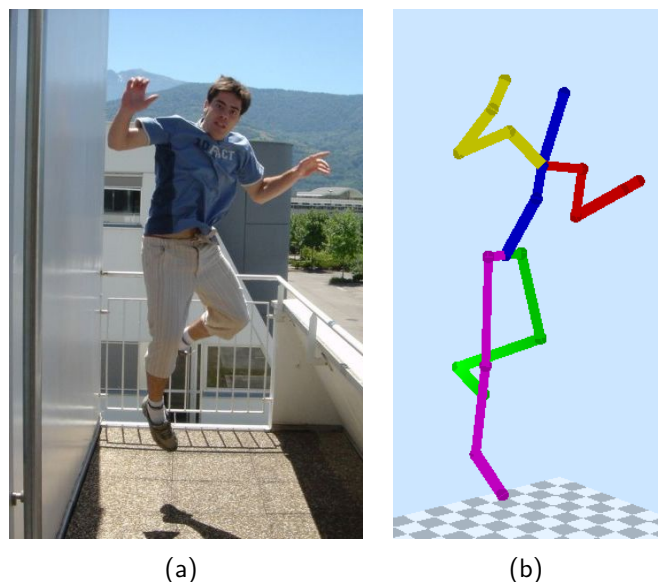


Figura 6.1: Limitação do modelo: Mesmo que o centro de apoio da pessoa não esteja de acordo com a realidade (a), o sistema gera a pose correta (b).

Durante os testes verificou-se que o modelo proposto não é invariante a rotação. Por exemplo, a Figura 6.5 mostra duas imagens que apresentam a mesma distância de conforto. De fato, pode-se concluir que em relação as mãos, braços, pernas e posição das articulações, estas duas poses são



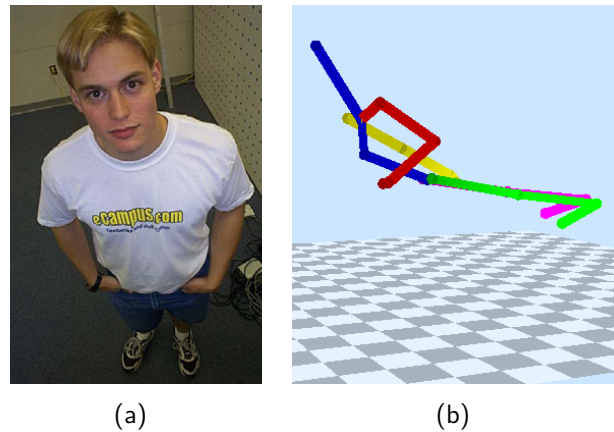


Figura 6.2: Pose obtida através de uma perspectiva sobre a pessoa (a). A Figura (b) é o resultado gerado pelo modelo.

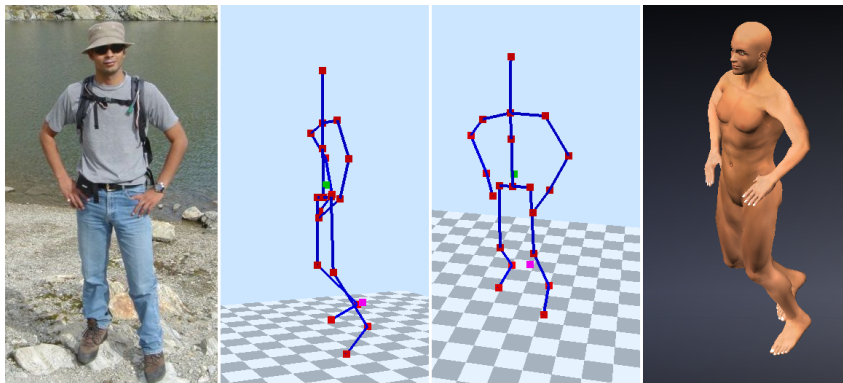


Figura 6.3: Limitações do modelo apresentado. A imagem obtida em perspectiva afeta o tamanho das pernas.

equivalentes, mas elas representam duas posturas diferentes no sentido do seu conforto, por exemplo, são poses diferentes se uma pessoa realiza essas posturas na vida real. O tratamento deste tipo de problema ficará aberto para ser tratado em trabalhos futuros.

## 6.2 Trabalhos Futuros

Em trabalhos futuros, pretende-se lidar com projeção em perspectiva na imagem. Além disso, o método de sombreamento pode ser melhorado incluindo técnicas avançadas de processamento de imagens para definir a luminosidade na imagem com melhor precisão. Abaixo uma lista dos trabalhos futuros que possibilitarão a continuidade dos estudos desta tese.

1. Aplicações de novas regras de restrições biomecânicas (incluir restrições para braços e antebraços);
2. Comparativo da abordagem de luminosidade com outras abordagens semelhantes do estado da arte a fim de verificar a sua eficiência e seus problemas;

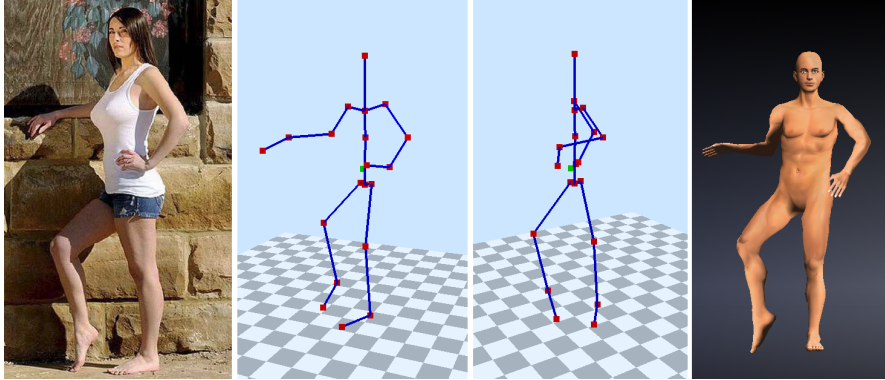


Figura 6.4: Outra limitação do modelo é quando a fonte de luz não está em frente a pessoa, o que gera informação incorreta para o modelo de luminosidade.

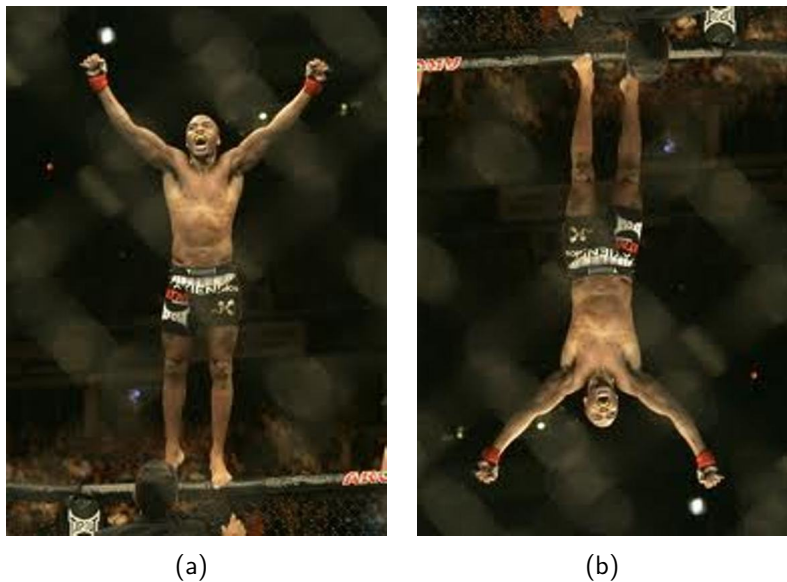


Figura 6.5: Limitação do modelo: A imagem foi invertida verticalmente e possui o mesma distância do conforto, o qual não é consistente com a realidade. Esquerda: Imagem original. Direita: Imagem invertida verticalmente.

3. Verificação da posição da fonte de luz de forma automática;
4. Correção do problema causado pela perspectiva;

Concluindo, o modelo proposto reduz a possibilidade de posturas em relação a abordagem de Taylor [Tay00] e apresenta uma classificação onde a pose correta encontra-se normalmente nas primeiras posições. Por fim, acredita-se que esta tese obteve sucesso em gerar um novo modelo para a obtenção de posturas 3D a partir de uma imagem bidimensional, além proporcionar as contribuições da aplicação de restrições biomecânicas, conforto da postura e análise da luminosidade minimizando a ambiguidade de posturas geradas e reduzindo o espaço de busca para a solução da postura correta. Apresentou um conjunto de novas características baseadas no conforto da postura humana, sendo que estas características podem ser aplicadas a fim de especificar poses e mostrou uma aplicação para determinar poses humanas baseada em um detector de auto-occlusão.

## Bibliografia

- [AC99] AGGARWAL J. K., CAI Q.: Human motion analysis: A review. *Computer Vision and Image Understanding* 73 (1999), 428–440.
- [ARS09] ANDRILUKA M., ROTH S., SCHIELE B.: Pictorial structures revisited: People detection and articulated pose estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2009).
- [AT04] AGARWAL A., TRIGGS B.: 3d human pose from silhouettes by relevance vector regression. In *Proceedings of the 2004 IEEE computer society conference on Computer vision and pattern recognition* (Washington, DC, USA, 2004), CVPR'04, IEEE Computer Society, pp. 882–888.
- [AT06a] AGARWAL A., TRIGGS B.: A local basis representation for estimating human pose from cluttered images. In *Asian Conference on Computer Vision - ACCV* (2006), pp. 50–59.
- [AT06b] AGARWAL A., TRIGGS B.: Recovering 3d human pose from monocular images. *Pattern Analysis and Machine Intelligence* 28 (2006), 44 –58.
- [BK00] BARRON C., KAKADIARIS I.: On the improvement of anthropometry and pose estimation from a single uncalibrated image. In *Human Motion, 2000. Proceedings. Workshop on* (2000), pp. 53–60.
- [BM09] BOURDEV L., MALIK J.: Poselets: Body part detectors trained using 3d human pose annotations. In *Computer Vision, 2009 IEEE 12th International Conference on* (2009), pp. 1365–1372.
- [CL92] CHEN Z., LEE H.-J.: Knowledge-guided visual perception of 3-d human gait from a single image sequence. *Systems, Man and Cybernetics, IEEE Transactions on* 22, 2 (1992), 336–342.
- [CYL12] CHO N.-G., YUILLE A., LEE S.-W.: Self-occlusion robust 3d human pose tracking from monocular image sequence. In *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on* (2012), pp. 254–257.
- [DT05] DALAL N., TRIGGS B.: Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005* (2005), vol. 1, pp. 886 – 893.
- [EF09] EICHNER M., FERRARI V.: Better appearance models for pictorial structures. In *British Machine Vision Conference 2009* (2009), pp. 110–119.

- [EL04] ELGAMMAL A., LEE C.-S.: Inferring 3d body pose from silhouettes using activity manifold learning. In *Proceedings of the 2004 IEEE computer society conference on Computer vision and pattern recognition* (Washington, DC, USA, 2004), CVPR'04, IEEE Computer Society, pp. 681–688.
- [FMJZ08] FERRARI V., MARIN-JIMENEZ M., ZISSERMAN A.: Progressive search space reduction for human pose estimation. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (june 2008), pp. 1–8.
- [Gav99] GAVRILA D. M.: The visual analysis of human movement: A survey. *Computer Vision and Image Understanding* 73 (1999), 82–98.
- [Gav07] GAVRILA D.: A bayesian, exemplar-based approach to hierarchical shape matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29, 8 (2007), 1408–1421.
- [GDH11] GUO R., DAI Q., HOIEM D.: Single-image shadow detection and removal using paired regions. In *CVPR* (2011), IEEE, pp. 2033–2040.
- [GQ06] GUO F., QIAN G.: Learning and inference of 3d human poses from gaussian mixture modeled silhouettes. In *Proceedings of the 18th International Conference on Pattern Recognition - Volume 02* (Washington, DC, USA, 2006), ICPR '06, IEEE Computer Society, pp. 43–47.
- [GWBB09] GUAN P., WEISS A., BALAN A. O., BLACK M. J.: Estimating human shape and pose from a single image. In *Computer Vision, 2009 IEEE 12th International Conference on* (2009), pp. 110–119.
- [HDK07] HORNING A., DECKERS E., KOBELT L.: Character animation from 2d pictures and 3d motion data. In *ACM Transactions on Graphics* (2007), vol. 26, pp. 1–9.
- [HWLY09] HU Z., WANG G., LIN X., YAN H.: Recovery of upper body poses in static images based on joints detection. *Pattern Recogn. Lett.* 30 (April 2009), 503–512.
- [HY10] HUANG J.-B., YANG M.-H.: Estimating human pose from occluded images. In *Proceedings of the 9th Asian conference on Computer Vision - Volume Part I* (Berlin, Heidelberg, 2010), ACCV'09, Springer-Verlag, pp. 48–60.
- [HYW05] HUA G., YANG M.-H., WU Y.: Learning to estimate human pose with data driven belief propagation. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02* (Washington, DC, USA, 2005), CVPR '05, IEEE Computer Society, pp. 747–754.

- [JDJ\*10] JACQUES J. C. S., DIHL L. L., JUNG C. R., THIELO M. R., KESHET R., MUSSE S. R.: Human upper body identification from images. In *ICIP (2010)*, IEEE, pp. 1717–1720.
- [JDJM13] JACQUES J., DIHL L., JUNG C., MUSSE S.: Self-occlusion and 3d pose estimation in still images. In *Image Processing (ICIP), 2013 20th IEEE International Conference on - Accepted for publication (2013)*.
- [Jia10] JIANG H.: 3d human pose reconstruction using millions of exemplars. In *Pattern Recognition (ICPR), 2010 20th International Conference on (aug. 2010)*, pp. 1674–1677.
- [Jia11] JIANG H.: Human pose estimation using consistent max covering. *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (September 2011), 1911–1918.
- [JJM12] JACQUES J., JUNG C., MUSSE S.: Skeleton-based human segmentation in still images. In *Image Processing (ICIP), 2012 19th IEEE International Conference on (2012)*, pp. 141–144.
- [KK10] KIM D., KIM D.: Self-occlusion handling for human body motion tracking from 3d tof image sequence. In *Proceedings of the 1st international workshop on 3D video processing (New York, NY, USA, 2010)*, 3DVP '10, ACM, pp. 57–62.
- [LC85] LEE H.-J., CHEN Z.: Determination of 3d human body postures from a single view. *Computer Vision, Graphics, and Image Processing* 30, 2 (1985), 148 – 168.
- [LC06] LEE M. W., COHEN I.: A model-based approach for estimating human 3d poses in static images. *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (June 2006), 905–916.
- [LM03] L. MOCHIZUKI A. C. A.: Aspectos biomecãnicos da postura ereta: a relaãõ entre o centro de massa e o centro de pressãõ. *Revista portuguesa de ciãncias do desporto* 3 (2003), 77–83.
- [LN07] LEE M. W., NEVATIA R.: Body part detection for human pose estimation and tracking. In *Proceedings of the IEEE Workshop on Motion and Video Computing (Washington, DC, USA, 2007)*, IEEE Computer Society, pp. 23–.
- [MBR06] MENIER C., BOYER E., RAFFIN B.: 3d skeleton-based body pose recovery. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on (june 2006)*, pp. 389 –396.
- [MG01] MOESLUND T. B., GRANUM E.: A survey of computer vision-based human motion capture. *Comput. Vis. Image Underst.* 81 (March 2001), 231–268.

- [MHK06] MOESLUND T. B., HILTON A., KRÜGER V.: A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.* 104, 2 (Nov. 2006), 90–126.
- [MM06] MORI G., MALIK J.: Recovering 3d human body configurations using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 28, 7 (July 2006), 1052–1062.
- [MREM04] MORI G., REN X., EFROS A., MALIK J.: Recovering human body configurations: combining segmentation and recognition. In *Computer Vision and Pattern Recognition, 2004.* (2004), vol. 2, pp. 326–333.
- [NH94] NIGG B., HERZOG W.: *Biomechanics of the Musculo-skeletal System.* John Wiley and Sons, inc, 1994.
- [OURHJR04] ORRITE-URUNUELA C., RINCON J. M. D., HERRERO-JARABA J. E., ROGEZ G.: 2d silhouette and 3d skeletal models for human detection and tracking. In *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 4 - Volume 04* (Washington, DC, USA, 2004), ICPR '04, IEEE Computer Society, pp. 244–247.
- [PC04] PARAMESWARAN V., CHELLAPPA R.: View independent human body pose estimation from a single perspective image. In *Proceedings of the 2004 IEEE computer society conference on Computer vision and pattern recognition* (Washington, DC, USA, 2004), CVPR'04, IEEE Computer Society, pp. 16–22.
- [PJA\*12] PISHCHULIN L., JAIN A., ANDRILUKA M., THORMAHLEN T., SCHIELE B.: Articulated people detection and pose estimation: Reshaping the future. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (2012), pp. 3178–3185.
- [RDJG12] RADWAN I., DHALL A., JOSHI J., GOECKE R.: Regression based pose estimation with automatic occlusion detection and rectification. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on* (2012), pp. 121–127.
- [RKB04] ROTHER C., KOLMOGOROV V., BLAKE A.: "grabcut": interactive foreground extraction using iterated graph cuts. In *ACM SIGGRAPH 2004 Papers* (New York, NY, USA, 2004), SIGGRAPH '04, ACM, pp. 309–314.
- [SB06] SIGAL L., BLACK M.: Measure locally, reason globally: Occlusion-sensitive articulated pose estimation. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* (2006), vol. 2, pp. 2041–2048.

- [SSRA\*12] SIMO-SERRA E., RAMISA A., ALENYA G., TORRAS C., MORENO-NOGUER F.: Single image 3d human pose estimation from noisy observations. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (2012), pp. 2673–2680.
- [Tay00] TAYLOR C. J.: Reconstruction of articulated objects from point correspondences in a single uncalibrated image. *Comput. Vis. Image Underst.* 80 (2000), 349–363.
- [Til02] TILLEY A. R.: *The measure of man and woman - Human factors in design*. John Wiley & Sons, inc, 2002.
- [VJ04] VIOLA P., JONES M. J.: Robust real-time face detection. *Int. J. Comput. Vision* 57 (May 2004), 137–154.
- [WC09] WEI X. K., CHAI J.: Modeling 3d human poses from uncalibrated monocular images. In *Computer Vision, 2009 IEEE 12th International Conference on* (2009), pp. 1873–1880.
- [WL06] WANG R., LEOW W. K.: Human posture analysis under partial self-occlusion. In *ICIAR (1)'06* (2006), pp. 874–885.
- [YR11] YANG Y., RAMANAN D.: Articulated pose estimation with flexible mixtures-of-parts. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition* (Washington, DC, USA, 2011), CVPR '11, IEEE Computer Society, pp. 1385–1392.