UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas
Programa de Pós-Graduação em Ciência da Computação

Alan Deivite Guimarães da Silva

Determinação da Associação de Indivíduos e seu Espaço Social Coletivo para Aplicações em Interação Humano-Robô

Belo Horizonte
2018

Alan Deivite Guimarães da Silva

Determinação da Associação de Indivíduos e seu Espaço Social Coletivo para Aplicações em Interação Humano-Robô

Versão Final

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Ciência da Computação.

Orientador: Douglas Guimarães Macharet

Belo Horizonte
2018

2018, Alan Deivite Guimarães da Silva.
Todos os direitos reservados

Silva, Alan Deivite Guimarães da.

S586d Determinação da associação de indivíduos e seu espaço social coletivo para aplicaçoes em interação humano-robô [recurso eletrônico] / Alan Deivite Guimarães da Silva – 2018.
1 recurso online (109 f. il., color.) : pdf.

Orientador: Douglas Guimarães Macharet.

Dissertação (Mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação.
Referências: f. 93-99.

1. Computação – Teses. 2. Robótica – Teses. 3. Interação homem-máquina – Teses. 4. Robôs móveis - Teses. 5. Espaços públicos – Teses. I. Macharet, Douglas Guimarães. II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação. III. Título.

CDU 519.6*75(043)

Ficha catalográfica elaborada pela bibliotecária Irénquer Vismeg Lucas Cruz
CRB 6/819 - Universidade Federal de Minas Gerais - ICEx

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Determinação da associação de indivíduos e seu espaço social coletivo para aplicações em interação humano-robô

ALAN DEIVITE GUIMARÃES DA SILVA

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

Douglas Macharet

PROF. DOUGLAS GUIMARÃES MACHARET - Orientador
Departamento de Ciência da Computação - UFMG

PROF. MARIO FERNANDO MONTENEGRO CAMPOS
Departamento de Ciência da Computação - UFMG

PROF. LUIZ CHAIMOWICZ
Departamento de Ciência da Computação - UFMG

PROF. ARMANDO ALVES NETO
Departamento de Engenharia Eletronica - UFMG

Belo Horizonte, 29 de maio de 2018.

Dedico este trabalho a minha família e aos amigos com quem convivi ao longo desses anos.

Agradecimentos

Agradeço à minha família, em especial, minha mãe Maria Guimarães que sempre esteve ao meu lado. Ao meu pai Lourival Alves que batalhou por anos para proporcionar a melhor educação para seus filhos.

A todos os meus amigos, agradeço imensamente pelos inúmeros conselhos, incentivo e apoio incondicional. Agradecimentos especiais para Camila Laranjeira, Hugo Oliveira, Felipe Oliveira, Alexsandro Gabriel, Virgínia Mota, Jéssica Sena, Elerson Rubens, Renato Garcia, Mabel Abreu, Wallysson Barbosa e Bruna Frade.

Gostaria de deixar o meu profundo agradecimento ao meu orientador Douglas Guimarães Macharet, pelo empenho aplicado ao meu projeto de pesquisa, confiança e incansável dedicação.

Sou grato ao Laboratory of Computer Vision & Robotics (VeRLab) e ao Programa de Pós Graduação em Ciências Contábeis (PPGCC), desde o pessoal do administrativo até a coordenação do curso, que de alguma forma contribuíram para a realização deste trabalho. Gostaria de agradecer também ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, à Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, por possibilitar a execução deste trabalho científico.

A todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigado.

"A lesson without pain is meaningless. For you cannot gain anything without sacrificing something else in return, but once you have overcome it and made it your own...you will gain an irreplaceable fullmetal heart."

(Hiromu Arakawa)

Resumo

Com o desenvolvimento da robótica, a utilização de robôs móveis autônomos na execução de tarefas cotidianas tende a se tornar cada vez mais factível em diferentes partes da sociedade. Este trabalho possui como objetivo propor um algoritmo baseada em grafo para identificar possíveis associações de indivíduos no ambiente, apresentando uma nova abordagem para computar o nível de interação entre os indivíduos de acordo com sua posição e orientação corporal. Posteriormente, com base nessa detecção dos grupos de interação, propomos uma representação do espaço social coletivo baseada em uma função de densidade, por intermédio da junção de funções Gaussianas assimétricas para delimitação do espaço socialmente aceitável, de forma que um robô móvel consiga realizar tarefas na presença de pedestres sem oferecer nenhum tipo de risco físico ou desconforto social, como colidir com um indivíduo ou invadir o seu espaço pessoal de interação. A metodologia proposta para determinação da associação de indivíduos foi avaliada utilizando *datasets* sintéticos e do mundo real, apresentando performance superior aos trabalhos do estado-da-arte na maioria dos casos. Além disso, avaliamos por meio de simulações a representação do espaço social coletivo, demonstrando que nossa abordagem gera comportamentos mais socialmente favoráveis do que métodos que não consideram tais restrições.

Palavras-chave: interação humano-robô; espaço social coletivo; ambiente centrado no homem; robô autônomo

Abstract

With the development of robotics, the use of autonomous mobile robots in the execution of everyday tasks tends to become increasingly feasible in different parts of society. This work aims to propose a graph based algorithm to identify possible associations of individuals in the environment, presenting a new approach to compute the level of interaction between individuals according to their position and body orientation. Subsequently, based on this detection of the interaction groups we propose a representation of the collective social space based on a density function through the junction of asymmetric Gaussian functions to delimit the socially acceptable space, so that a mobile robot can perform tasks in the presence of pedestrians without offering any type of physical risk or social discomfort like colliding with an individual or invading your personal interaction space. The proposed methodology for determining the association of individuals was evaluated using synthetic and real-world datasets, outperforming the state-of-the-art methods on most cases. In addition, we evaluated through simulations the representation of the collective social space demonstrating that our approach generates more socially favorable behaviors than methods that do not consider such restrictions.

Keywords: human-robot interaction; collective social space; human-centered environment; autonomous robot

Lista de Figuras

1.1 Exemplos de robôs que agregaram avanços tecnológicos e revolucionaram a histórica da robótica. 20
1.2 Algumas aplicações existentes que envolvem a tarefas de interação social entre humanos e robôs. 23
1.3 Diferentes tipos de zonas interpessoais definido por Hall [26], onde as distâncias definidas para cada zona podem variar de acordo com o lugar ou a depender da cultura em que se está imergido. 24
1.4 Ilustração das diferentes etapa: (a) Indivíduos espalhados pelo ambiente; (b) Identificação dos grupos de interação; (c) Determinação do espaço Pessoal; (d) exemplo de uso em uma tarefa de navegação. 26

2.1 Zonas de reação pessoal, revelando quanto espaço as pessoas geralmente precisam em várias situações sociais. 29
2.2 Valores médios (em centímetros) para as distâncias social, pessoal e íntima em todas as 42 nações avaliadas. 30
2.3 Tamanho real do espaço pessoal quando as pessoas estão em fila [40]. 31
2.4 Modelo do espaço pessoal com base na orientação do rosto. O espaço pessoal na área frontal de uma pessoa é mais amplo que as áreas traseiras e laterais [3]. 31
2.5 Custo do espaço pessoal para uma pessoa que se desloca a 1.0m/s ao longo do eixo Y positivo (para cima) [34]. 32
2.6 Ilustração do mapeamento das interações sociais de três participantes ($\mathbf{p}_1$, $\mathbf{p}_2$ e $\mathbf{p}_3$) usando a abordagem proposta por Papadakis et al. [46] em vários intervalos de tempo. 32
2.7 O método de estimativa do estado proxemic objetivo usa uma abordagem baseada em amostragem para estimar como o ser humano produz sinais sociais, e como o robô os percebe usando modelos baseados em dados. 33
2.8 Exemplo da multiplicidade de pistas comportamentais não-verbais expressada na atitude de alguém em relação a uma situação social, tais como a postura corporal, gesto, olhar, expressão facial, análise conversacional e proxemics [71]. 34
2.9 Arranjos espaciais de F-formation humana: (a) representação sócio-espacial do padrão F-Formation [32]. (b) representação dos F-formations mais frequentes para grupos de 2 pessoas [11]. 35

2.10 Exemplificação da abordagem proposta. (a-c) Dois indivíduos posicionados frontalmente a um voto de distância fixa para um mesmo centro da circunferência do *o-space*. (d) Os indivíduos não estão de forma exatamente frontal em casos reais. (e-f) Várias posições e orientações da cabeça são desenhadas a partir de distribuições Gaussianas de modo a lidar com a incerteza de cenários reais [14]. 35
2.11 Fluxograma resumindo a abordagem. Linhas verdes: F-formations anotados, Caixa amarela: associado, Linhas vermelhas: detecções [31]. 36
2.12 Exemplo de disposição poligonal de 5 pessoas em um F-formation (Esquerda) e um real F-formation com o relacionado *o-space* destacado em vermelho (Direita) [55]. 36
2.13 Ilustração da descoberta do grupo interagindo: Dois grupos são descobertos e duas pessoas que não pertencem a qualquer grupo. As pessoas não pertencentes ao grupo são eliminadas na análise da atividade grupal. [62]. 37
2.14 Na figura é mostrado o processo de geração do tronco. 38
2.15 (Esquerda) *O-space* do F-formation envolvendo três alvos k, r e q com a sua pose do corpo. (Centro) vetores de direção $d_{(.)}$ obtidos através do regressor de pose do corpo são mostrados usando setas azuis, enquanto $c_{(.)}$ (pontos amarelos) denotam os locais do centro eleitos. 38
2.16 Resultados qualitativos para os *datasets eth* (primeira linha), *hotel* (segunda linha) e *zara01* (último linha). 39
2.17 Exemplo do rastreador. A linha indica o deslocamento do quadro anterior. Menos de 0,625 fps, é difícil encontrar uma correspondência entre quadros sem prévia [72]. 40
2.18 (a) e (b) mostram um grupo uniforme e não uniforme, respectivamente. 40
2.19 Grupos pequenos são prevalentes em cenas de pedestres. O algoritmo proposto detecta grupos de pessoas que viajam juntas através de agrupamento hierárquico em trajetórias extraídas automaticamente de vídeo de multidões em várias condições [22]. 41
2.20 Destaques das propriedades dos grupos sociais: (a) coerência hierárquica, (b) invariância de densidade e (c) transitividade. [60]. 41
2.21 Uma visão do caminho do robô (preto), um caminho humano (laranja / cinza escuro) e o caminho mais curto (verde / cinza claro). 43
2.22 Navegação segura de um robô móvel através de multidões de agentes dinâmicos com trajetórias incertas. 43
2.23 Os custos em frente do humano em movimento aumentou em comparação com o humano parado. 44

2.24 Uma pessoa bloqueia um caminho significativamente mais curto causando um desvio para o robô. Alternativamente, o robô pode tentar espantar esta pessoa e seguir o caminho mais curto [39]. 44
2.25 O plano estático resultante e os planos dinâmicos são mostrados em verde e laranja, respectivamente [13]. 45
2.26 Simulação do robô navegando em um ambiente povoado por pessoas em três momentos diferentes. 45
2.27 Algumas das *features* empregadas na função de custo social aprendida. 46
2.28 Resultados para a solução de observação, abordando um grupo de humanos [24]. 47
2.29 Ilustração do resultado do *framework* de conforto e segurança humano em diferentes cenários. 47
2.30 Pose inicial e final do robô em relação a um grupo detectado para um caso simulado [42]. 48
2.31 Conversação de grupo simulada entre um robô e quatro pessoas. Os círculos vermelho e verde em cima dos agentes identificam o falante e os destinatários, respectivamente. O grande círculo cinza representa o *o-space* do grupo F-formation [69]. 48
2.32 Robô aproximou-se do humano H1, para iniciar o processo de orientação. O trajeto verde mostra o caminho planejado para o objetivo após o compromisso conjunto de orientação ter sido estabelecido com H1. [45]. 49
2.33 (a) Representação de um ambiente com pessoas e robôs. (b) Computação do contorno convexo. (c) Interpolação do contorno convexo com fórmula *Newton Backward Divided Difference* (d) Cálculo da trajetória para resgatar o indivíduo 49
2.34 Visão da câmera aérea com uma sobreposição de pose do robô e eixos mostrando a posição detectada do parceiro social [20]. 50
2.35 Caminhos da pessoa e do robô em torno de cantos, para cada uma das duas abordagens. 51
2.36 Usuário remoto e o acompanhante caminhando juntos enquanto conversam [12]. 51

3.1 Representação da mudança no formato do espaço social segundo a velocidade de deslocamento do indivíduo [34]. 56
3.2 Definição dos limiares que modificarão o formato do espaço interpessoal segundo a zona social desejada (e.g. Íntima, Pessoal, Social ou Pública). 57
3.3 Determinação do grupo. (a) O círculo tracejado vermelho representa a região de busca de um indivíduo v_i dada uma distância social δ. (b) Grafo de interação completo para todos os indivíduos na cena. 58
3.4 Ilustração da distância de disparidade no foco de atenção entre um par de indivíduos v_i e v_j. 59

3.5 Exemplificação das circunstâncias onde se faz necessário a adição de restrições ao modelo matemático proposto. 60
3.6 Exemplo de valores de peso para o nível de interação de acordo com a Equação 3.9 considerando diferentes configurações de F-formation. 61
3.7 Exemplo das etapas de definição do espaço social coletivo para $\delta = 1.2$m. . . . 62

4.1 Gráfico de radar contendo as variáveis quantitativas referentes ao F_1-score para os *datasets* organizados em sentido horário segundo sua ordem decrescente de qualidade de detecção. 70
4.2 Distribuição temporal dos valores de F_1-score obtidos por *frame* para $T = 1$. As circunstâncias mais desfavorecidas são destacadas no sombreado roxo. . . . 71
4.3 Distribuição temporal dos valores de F_1-score obtidos por *frame* para $T = 2/3$. As circunstâncias mais desfavorecidas são destacadas no sombreado roxo. . . . 72
4.4 *Frame* de exemplo que ilustra casos de classificação errônea de grupos interativos considerando o *dataset* sintético. Os círculos tracejados vermelhos representam os espaços pessoais dos indivíduos. A cor do círculo externo representa a nossa classificação, e a cor interna o *ground truth*. 73
4.5 Exemplo de classificação incorreta de grupos interagindo considerando o *dataset* Cocktail Party. Os círculos azuis representam a classificação segundo o *ground truth*, e os círculos tracejados vermelhos a distância social. 74
4.6 Exemplo com *frames* ilustrando casos de classificação errônea em grupos interativos considerando o *dataset* SALSA. A cor do círculo externo representa a nossa classificação, e a cor interna o *ground truth*. 75
4.7 Avaliação do nível de acurácia para diferentes distâncias sociais. 77
4.8 Gráfico de radar contendo as variáveis quantitativas referentes ao F_1-score incluindo os resultados obtidos após a otimização dos parâmetros para cada *dataset*. 79
4.9 Nível de acurácia aplicando filtro da mediana para diferentes janelas. 80
4.10 Representação do espaço social atômico/coletivo para diferentes zonas de proxemics. 81
4.12 Exemplo de formação do espaço social coletivo para $\delta = 1.25$m, destacando a presença de um indivíduo *singleton* e a delimitação do seu espaço pessoal no ambiente. 82
4.11 Sequência de dados temporal demonstrando a inserção de um novo indivíduo em um grupo de interação e a readaptação do espaço social coletivo para incorporar o espaço interpessoal de todos os indivíduos do grupo. 83
4.13 Exemplo de formação do espaço social coletivo para $\delta = 1.25$m, destacando a interação entre dois indivíduos e a delimitação do espaço pessoal abrangendo o grupo. 84

4.14 Configuração utilizada para avaliar a utilização do espaço social coletivo de forma socialmente aceitável (Figura (a) extraída de Pérez-Higueras et al. [47]). 85
4.15 Configurações aplicadas ao modelo de navegação diferencial do Pioneer p3dx. . 86
4.16 Trajetória executadas pelo robô nos experimentos realizados por Pérez-Higueras et al. [47]. 87
4.17 Trajetórias executadas pelo robô para indivíduos singulares considerando as diferentes etapas dos experimentos. 87
4.18 Trajetórias executadas pelo robô para grupo de indivíduos considerando as diferentes etapas dos experimentos. 88

A.1 Resultados relativos ao tempo necessário para atingir o objetivo (T_p), para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação. 101
A.2 Resultados relativos ao comprimento do percurso (L_p), para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação. 102
A.3 Resultados relativos a alterações cumulativas de orientação (CHC), para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação. 103
A.4 Resultados relativos a distância média para a pessoa mais próxima (CP_{avg}), para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação. 104
A.5 Resultados relativos a distância mínima para a pessoa mais próxima (CP_{min}), para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação. 105
A.6 Resultados relativos a distância máxima para a pessoa mais próxima (CP_{max}), para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação. 106
A.7 Resultados relativos a intrusões de espaço pessoal (CP_{prox}) na distância íntima, para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação. 107

A.8 Resultados relativos a intrusões de espaço pessoal (CP_{prox}) na distância pessoal, para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação. 108
A.9 Resultados relativos a intrusões de espaço pessoal (CP_{prox}) na distância social+pública, para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação. 109

Lista de Tabelas

2.1 Alguns dos principais trabalhos encontrados na literatura, destacando as caracteristicas mais relevantes abordadas por cada um deles. 53

4.1 Principais características dos *datasets* [56]. 65
4.2 Imagens correspondentes aos *datasets* utilizados. 65
4.3 Avaliação da determinação de grupo considerando $T = 1$ e valores de comparação retirados de Setti et al. [56]. Nestes experimentos consideramos uma distância social fixa de $\delta = 1.20$m, que se encontra na zona de proxemics Pessoal. 69
4.4 Avaliação da determinação de grupo considerando $T = 2/3$ e valores de comparação retirados de Setti et al. [56], Alameda-Pineda et al. [2]. Nestes experimentos consideramos uma distância social fixa de $\delta = 1.20$m, que se encontra na zona de proxemics Pessoal. 69
4.5 Resultados considerando diferentes distâncias sociais para cada *dataset* avaliado, onde a maximização da métrica F_1-score foi utilizada como critério para determinação da melhor distância δ. 74
4.6 Resultados incluindo os melhores parâmetros para cada *dataset* avaliado, onde a maximização da métrica F-score foi utilizada como critério para determinação dos melhores parâmetros de distância social e janela temporal. 78
4.7 Resultados para a avaliação da navegação social em um cenário estático com duas pessoas. 90
4.8 Resultados para a avaliação da navegação social em um cenário estático com um grupo formado por duas pessoas. 90

Lista de Abreviações e Siglas

AGF	Asymmetric Gaussian Function
CAO	Cognitive-based Adaptive Optimization
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CBA	Commonwealth Bank of Australia
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
COMPANION	Constraint-Optimizing Method for Person-Acceptable NavigatION
CSS	Collective Social Space
DSZ	Dynamic Social Zone
FAPEMIG	Fundação de Amparo à Pesquisa do Estado de Minas Gerais
FN	False Negative
FP	False Positive
HANP	Human-Aware Navigation Planner
HRI	Human-Robot Interaction
HSI	Human-Swarm Interaction
IGP	Interacting Gaussian Processes
IRL	Inverse Reinforcement Learning
KLM	Koninklijke Luchtvaart Maatschappij
LoI	Level of Interaction
MIT	Massachusetts Institute of Technology
Non-SNS	Non-Social Navigation System
RA-L	IEEE Robotics and Automation Letters
ROS	Robot Operating System
RRT*	Optimal Rapidly-exploring Random Trees
SNS	Social Navigation System
SSVM	Structural Support Vector Machine
TP	True Positive
V-REP	Virtual Robot Experimentation Platform

Sumário

1 Introdução **19**

1.1 Contextualização . . . 20
1.2 Motivação . . . 22
1.3 Problema . . . 23
1.4 Objetivo . . . 25
1.5 Contribuições . . . 27
1.6 Organização do Texto . . . 27

2 Trabalhos Relacionados **28**

2.1 Espaço Social . . . 28
2.2 Percepção de Grupos de Pessoas . . . 33
2.2.1 Grupos de Interação . . . 34
2.2.2 Rastreamento de Grupos Sociais . . . 39
2.3 Aplicações de Robôs em Contextos Sociais . . . 41
2.3.1 Navegação Autônoma . . . 42
2.3.2 Abordando Pessoas . . . 46
2.3.3 Escoltando Pessoas . . . 48
2.3.4 Seguindo Pessoas . . . 50
2.4 Contextualização . . . 52

3 Metodologia **54**

3.1 Espaço Social Atômico . . . 54
3.1.1 Determinação de Interação entre Indivíduos . . . 57
3.2 Espaço Social Coletivo . . . 60

4 Experimentos **64**

4.1 Datasets . . . 65
4.2 Determinação do Grupo . . . 67
4.3 Espaço Social Atômico/Coletivo . . . 78

5 Conclusão e Trabalhos Futuros **91**

5.1 Conclusão . . . 91
5.2 Trabalhos Futuros . . . 92

Referências **93**

Capítulo 1

Introdução

A utilização de robôs autônomos está sofrendo uma mudança de cenário, migrando de ambientes controlados para ambientes cada vez mais não estruturados, em que as pessoas são parte integrante. Neste contexto, avanços recentes na pesquisa com *Human-Robot Interaction* (*Interação Humano-Robô*, HRI) descobriram uma ampla gama de novos problemas e provocaram novos desafios para várias classes de problemas já conhecidos. Dentre estes, um problema bem conhecido é posto pela questão básica da robótica móvel relacionada à navegação: "Como faço para alcançar um determinado objetivo?". Isto está diretamente ligado à estratégia usada pelo robô para atingir com segurança uma posição objetivo. Em um contexto social, não é apenas importante determinar caminhos espacialmente ou temporalmente otimizados, mas também é necessário que os robôs se comportem de uma maneira *socialmente aceitável*. A necessidade de comportamentos socialmente aceitáveis cruza muitos domínios e está especialmente preocupada com o nível de aceitação ou desconforto das pessoas em compartilhar o meio ambiente com agentes autônomos. Um espaço pessoal individual representa a distancia espacial que uma pessoa percebe e se relaciona em uma interação considerando o seu entorno. No entanto, na maioria das situações cotidianas, as pessoas estão interagindo entre si em grupos de duas ou mais pessoas. Neste caso, os espaços sociais individuais devem ser considerados em um novo *espaço social coletivo* que englobe todo o grupo.

Este trabalho propõe uma metodologia para avaliar a associação de indivíduos no ambiente com base em sua posição e orientação corporal, para identificá-los e classificá-los em grupos. Em seguida, apresentamos um modelo matemático baseado no uso de uma função de densidade para determinar eficientemente o espaço social de grupos e indivíduos de acordo com as diferentes zonas interpessoais, garantindo que o espaço individual também seja respeitado nas formações em grupos.

1.1 Contextualização

Se comparado com humanos, robôs possuem a capacidade de realizar tarefas repetitivas com bastante precisão e velocidade, sem a desvantagem de apresentar desgaste físico ou mental. Tais características contribuem para a diminuição da necessidade de atribuir pessoas para execução de trabalhos braçais repetitivos ou atividades que coloquem em risco a sua saúde. Em consequência, a aplicação de dispositivos robóticos no setor de manufatura industrial tornou-se uma realidade consolidada. Em 1959 a Unimation foi a primeira empresa responsável por produzir um manipulador industrial, denominado Unimate (Figura 1.1a). Mais precisamente, Unimate foi o primeiro braço robótico ou manipulador da indústria. No entanto, apesar da evolução tecnológica desses dispositivos ao longo dos anos, uma desvantagem fundamental dos manipuladores comerciais é a sua falta de mobilidade. Os manipuladores possuem um alcance limitado de movimentação, restritos a operar somente nas regiões ao entorno de onde foram fixados [58].

A robótica móvel surge então como uma alternativa para inserção de flexibilidade na execução de tarefas, agregando mobilidade aos dispositivos robóticos. Seu desenvolvimento teve início no final de 1960 com o projeto Shakey no Stanford Research Institute (Figura 1.1b). Com a introdução de novas concepções como os conceitos de percepção, mapeamento, planejamento de movimentos e arquitetura de controle, o desenvolvimento de robôs adquire finalidades mais abrangentes, ganhando aplicabilidade fora das fábricas. Por consequência, a utilização de robôs na execução de tarefas cotidianas tende a se tornar cada vez mais factível. Em contrapartida, o surgimento deste novo segmento suscita na necessidade de elaboração de novas técnicas que permitam agentes atuarem em ambientes na presença de humanos, sem oferecer nenhum tipo de risco, como colidir com um indivíduo ou invadir o seu espaço de interação interpessoal.



(a) Unimate (1959). (b) Shakey (1960). (c) kismet (1990).

Figura 1.1: Exemplos de robôs que agregaram avanços tecnológicos e revolucionaram a histórica da robótica: (a) Primeiro manipulador da indústria. (b) Início do desenvolvimento de robôs moveis. (c) Início das pesquisas em interação humano-robô.

Como reflexo dos avanços ocorridos no campo da robótica, em meados da década

de 1990, novas iniciativas de pesquisa surgiram, seguindo, em princípio, o novo paradigma de inteligência artificial, visando robôs com corpos e mentalidade humana [17], sendo o mais famoso o Cog do *Massachusetts Institute of Technology* (*Instituto de Tecnologia de Massachusetts*, MIT), um robô com a parte superior do torso humanoide, mais tarde acompanhado pelo Kismet, um robô do mesmo laboratório, que consistia em um rosto articulado que expressava "emoções"(Figura 1.1c). Um reflexo desse novo campo emergente que originou a área de estudos em HRI. Segundo Goodrich and Schultz [25], HRI é um campo de estudo dedicado à compreensão, concepção e avaliação de sistemas robotizados para uso por ou com seres humanos.

No processo de interação entre humanos e robôs as distâncias sociais também devem ser levadas em consideração. O estudo do espaço social e do conjunto de regras que regem as interações pessoais é denominado por *proxemics*. Este termo foi cunhado pela primeira vez pelo antropólogo Edward Hall [26], que possuía como interesse compreender como humanos utilizam o espaço na comunicação. Hall descreve a distância interpessoal de uma pessoa em quatro zonas distintas, onde as distâncias relativas entre pessoas podem variar de acordo com o lugar ou a depender da cultura em que se está inserido. O conjunto de regras sociais não-verbais relativas ao proxemic estão especialmente preocupados com o nível de aceitação ou desconforto das pessoas em compartilhar o seu espaço pessoal. Nesse sentido, ao inserirmos robôs autônomos e pessoas no mesmo ambiente compartilhado, as mesmas regras sociais passam a reger as interações entre os humanos e robôs.

Em geral, as metodologias propostas para tornar mais receptível as formas de interação entre humanos e robôs, principalmente em termos de navegação socialmente aceitável, tendem a modelar os comportamentos socialmente aceitáveis utilizando técnicas de aprendizado de máquina ao invés de defini-los matematicamente. Embora essas abordagens consigam gerar um modelo a partir da extração automática de correlações entre os dados de comportamentos sociais demonstrados, esses sistema possuem um comportamento estocástico, não sendo possível determinar como será o seu comportamento em ambientes imprevisíveis do mundo real. Este trabalho possui como objetivo propor uma nova abordagem baseada em grafo para identificar possíveis associações de indivíduos no ambiente de acordo com sua posição e orientação corporal. Posteriormente, com base nessa detecção dos grupos de interação, propomos uma representação do espaço social coletivo baseada em uma função de densidade, por intermédio da junção de funções Gaussianas assimétricas para delimitação do espaço socialmente aceitável, de forma que um robô móvel consiga realizar tarefas na presença de pedestres, minimizando qualquer impacto social para os indivíduos presentes no ambiente.

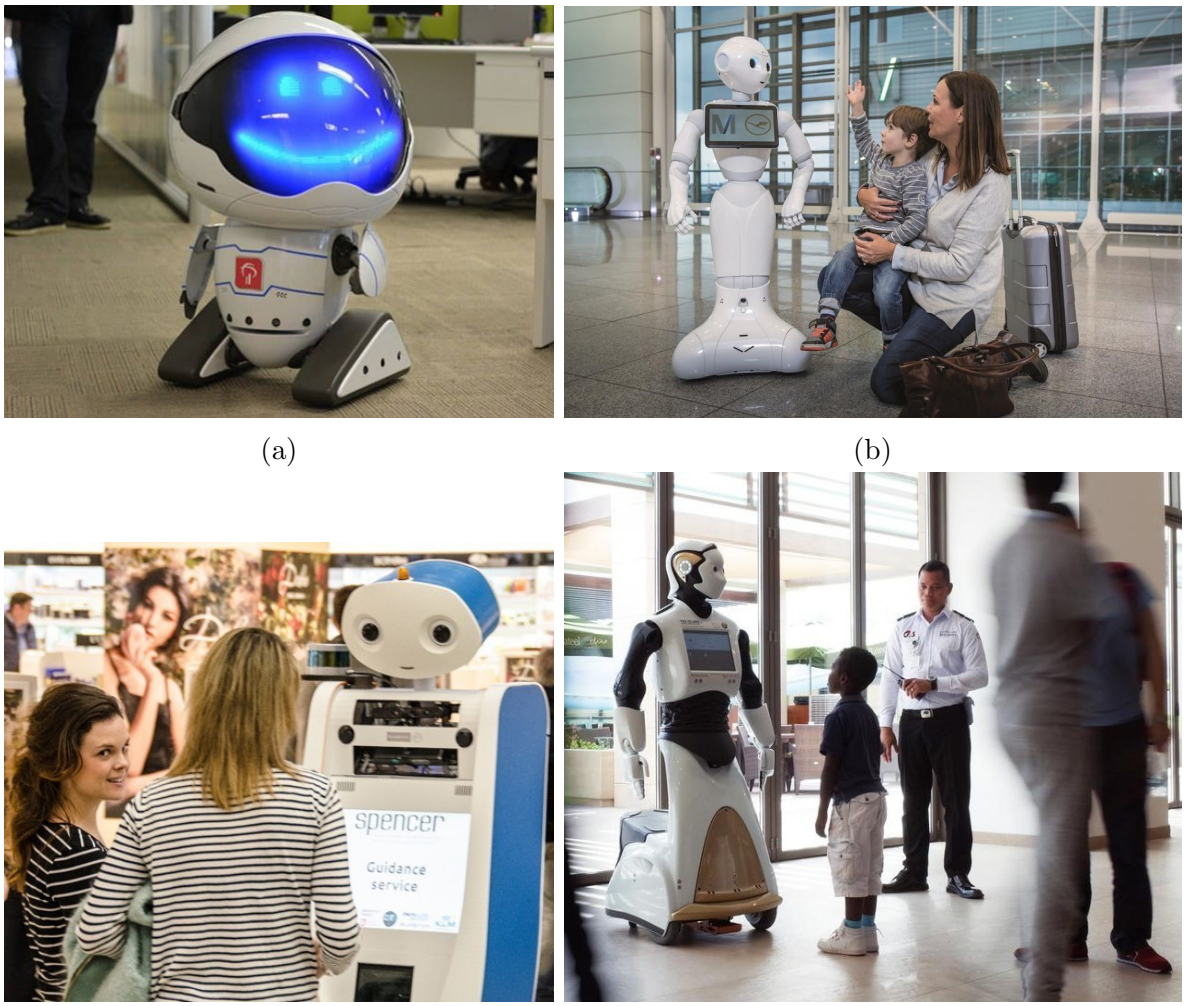
1.2 Motivação

Atualmente o desenvolvimento de robôs de serviço e companhia está recebendo bastante popularidade. Porém, estes equipamentos além de possuírem a necessidade de navegar entre obstáculos regulares, devem considerar também os humanos, sendo uma tarefa bastante complexa, principalmente quando situados em ambientes altamente dinâmicos, a exemplo de aeroportos, hospitais ou museus.

Geralmente, robôs móveis possuem um sistema de navegação que implementam algum método de desvio de colisões. Entretanto, a maioria não considera que na presença de humanos, não agir de acordo com a expectativa social causará desconforto para as pessoas presentes no ambiente. Este é um tópico de pesquisa ainda incipiente, sendo encontrado na literatura específica publicações recentes que discorrem sobre o assunto [36, 51, 8], explorando os problemas ainda em aberto desta vertente, onde a maioria das abordagens relacionadas a HRI se restringem a lidar apenas com uma pessoa ou pequenos grupos de indivíduos. Nesse trabalho, examinamos uma forma alternativa de identificar grandes grupos de interação e delimitar as diferentes zonas de conforto social considerando os indivíduos de forma coletiva. A determinação desse limites sociais proporciona a possibilidade de tratar os grupos de pessoas como entidades únicas, garantindo flexibilidade no uso de abordagens desenvolvidas anteriormente no contexto da navegação autônoma por entre grupos de pessoas. Além disso, a determinação de fronteiras tem grande aplicabilidade em sistemas multi-agente, em cenários como vigilância e monitoramento de multidões (por exemplo, um show ao ar livre ou um jogo de futebol), sendo melhor considerar as pessoas de forma mais coletiva em vez de individualmente.

Pesquisas nesse contexto ajudam a viabilizar cada vez mais aplicações envolvendo robôs autônomos na presença de seres humanos, e consequentemente na utilização desses para desempenhar atividades do cotidiano. Por exemplo, podemos considerar os seguintes robôs como aplicações já existentes no mundo real: Link 237, esse robô possui a função de recepcionista em uma das agências do banco Bradesco, cumprimentando e interagindo de forma amigável com os clientes; Pepper, desenvolvido em parceria entre as empresas Aldebaran Robotics e SoftBank, a principal funcionalidade desse robô é a de reconhecer emoções e interagir de modo correspondente com as pessoas, sua utilização está sendo testada pela Lufthansa em conjunto com o Aeroporto de Munich na tarefa de fornecer informações e responder a perguntas dos passageiros; Spencer, robô projetado pela Örebro University e encomendado pela empresa *Koninklijke Luchtvaart Maatschappij* (*Companhia Real de Aviação*, KLM) para atuar no Aeroporto de Schiphol localizado em Amsterdã, suas funções englobam guiar passageiros de um portão de embarque para outro, procurar clientes que estão perdendo voos e mostrar dados gerais da passagem em seu monitor; Chip Candroid, o robô social humanoide do banco *Commonwealth Bank of*

Australia (CBA) passou a integrar a empresa Air New Zealand ajudando os clientes com seus *check-ins* no portão de embarque do Aeroporto de Sydney.



(a) (b)

(c) (d)

Figura 1.2: Algumas aplicações existentes que envolvem a tarefas de interação social entre humanos e robôs: (a) Link 237, robô recepcionista que interage com clientes em agência bancária. (b) Pepper, robô social utilizado no Aeroporto de Munich capaz de detectar emoções e interagir de modo correspondente. (c) Spencer, robô capaz de localizar e guiar passageiros no aeroporto. (d) Chip Candroid, robô social humanoide que atua tando no CBA quanto no Aeroporto de Sydney.

1.3 Problema

Existe uma extensa quantidade de possíveis interações que podem ocorrer entre humanos e robôs, neste trabalho o escopo estará limitado as interações provenientes da

ação de navegação. Quando um robô se desloca por ambientes compostos por pessoas, a execução desta atividade pode causar desconforto aos observadores, seja quando o robô se interpõe no caminho dos pedestres, realiza uma tarefa de forma muito lenta, navega muito próximo a pessoas, ou se movimenta com velocidades elevadas. O principal problema reside em como instituir ao robô um processo de tomada de decisão similar ao que seria adotada por um ser humano.

Dado um ambiente com pessoas:

1. *Como identificar os diferentes grupos de interação?*
2. *Como determinar a associação entre estes indivíduos e seu espaço pessoal?*

A metodologia de identificação dos diferentes grupos de interação proposta, visa identificar a presença de grupos de interação em ambientes dinâmicos, desconsiderando a existência de obstáculos e assumindo o conhecimento prévio da posição e orientação dos indivíduos. Esses grupos de interação são formados por quatro zonas distintas que se dividem em distância Pública, Social, Pessoal e Íntima. (Figura 1.3).



(a) Distância pública.



(b) Distância social.



(c) Distância pessoal.



(d) Distância intima.

Figura 1.3: Diferentes tipos de zonas interpessoais definido por Hall [26], onde as distâncias definidas para cada zona podem variar de acordo com o lugar ou a depender da cultura em que se está imergido.

A distância pública corresponde ao espaço que caracteriza quão próximo sentamos ou permanecemos de pé em relação a uma pessoa pública ou palestrante durante uma fala

em público (Figura 1.3a). A distância social define o espaço de interações entre conhecidos, como quando se conversa com clientes no trabalho (Figura 1.3b). A distância pessoal qualifica o espaço de comunicação entre amigos ou familiares (Figura 1.3c). Por fim, a distância íntima caracteriza o espaço de interação entre pessoas que são muito próximas, possibilitando ações como abraçar, tocar ou sussurrar (Figura 1.3d). Portanto, é primordial o desenvolvimento de metodologias eficientes para examinar a cena e detectar essa associação entre os indivíduos. Usualmente, pesquisas relacionadas utilizam os conceitos de proxemics. Entretanto, este modelo de regras não descreve a orientação ou trajetória a ser seguida. Além de considerar as regras de proxemics, está dissertação parte da hipótese que as poses dos indivíduos podem ser utilizadas para delimitar as opções de trajetórias a serem seguidas. Pretende-se assim, dentro de uma multidão de pessoas (Figura 1.4a), identificar as interações estabelecidas entre todos os pares de indivíduo (Figura 1.4b), criando posteriormente uma região de conforto que obedeça as convenções sociais definidas de forma a respeitar cada indivíduo unitário (Figura 1.4c), gerando por fim, uma função implícita que permita uma tomada de decisão mais apropriada ao cenário enfrentado pelo robô (Figura 1.4d).

Na abordagem de determinação do espaço social referente ao indivíduo ou aos grupos de interação detectados no passo anterior, propomos uma representação baseada em uma função de densidade, por intermédio da junção de funções Gaussianas assimétricas para delimitação do espaço socialmente aceitável. Dessa forma, propomos uma metodologia de representação flexível para diferentes distâncias sociais, onde é possível extrair as regiões delimitantes do espaço interpessoal de interação dos indivíduos.

Dentre as possíveis aplicabilidades da representação das fronteiras de conforto social em grupos de interação, adotamos como exemplificação da metodologia proposta o problema da navegação de um robô móvel na presença de pedestres, avaliando por meio de simulações a representação do espaço social coletivo em ambiente estático e livre de obstáculos, onde o planejamento de trajetória do robô deveria evitar situações de colisão ou invasão do espaço pessoal de interação dos indivíduos. Sendo assim possível demonstrar nessa aplicabilidade, que a nossa abordagem gera comportamentos mais socialmente favoráveis do que métodos que não consideram tais restrições.

1.4 Objetivo

O objetivo dessa dissertação consiste no desenvolvimento de um método baseada em grafos, utilizando uma nova abordagem matemática para definir um modelo de interação entre pares de indivíduos de acordo com a sua posição e orientação corporal, e

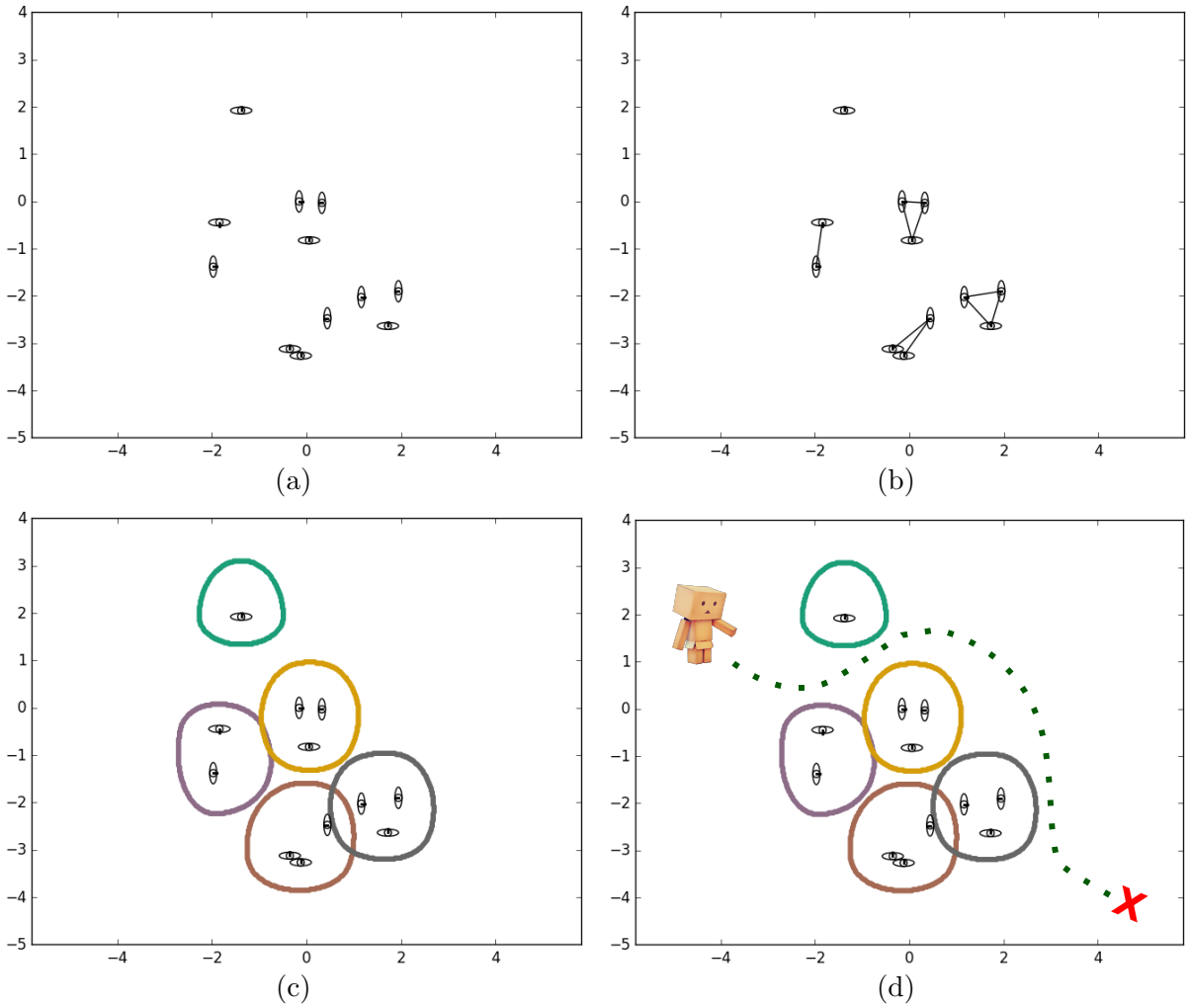


Figura 1.4: Ilustração das diferentes etapa: (a) Indivíduos espalhados pelo ambiente; (b) Identificação dos grupos de interação; (c) Determinação do espaço Pessoal; (d) exemplo de uso em uma tarefa de navegação.

com base nesse modelo de detecção dos grupos de interação, criar uma representação do espaço social coletivo de acordo com as zonas sociais especificadas nas regras de proxemics.

Nessa abordagem, pretendi-se obter agrupamentos de interação entre indivíduos com diferentes níveis de cardinalidade de acordo com a zona interpessoal especificada, estimando posteriormente uma representação do espaço interpessoal individual e coletivo dos indivíduos, de forma a delimitar uma região de conforto social. 15

Para o escopo do trabalho, será assumido um conhecimento global prévio da posição e orientação de todos os indivíduos. Assim, o trabalho terá um maior enfoque na estratégia de identificação dos grupos de interação e seleção das regiões referentes ao espaço social coletivo. Todos os ambientes de teste serão desprovidos de obstáculos e divididos entre cenários dinâmicos e estáticos. Os cenários dinâmicos serão utilizados durante os procedimentos experimentais para avaliar a determinação da associação de indivíduos e a representação do espaço social coletivo, utilizando *datasets* sintéticos e do mundo real. Já o cenário estático, será utilizado como demonstrativo do método proposto aplicado na

interação humano-robô, onde um robô deverá por meio de simulações executar uma determinada tarefa na presença de pessoas sem adentrar o espaço interpessoal dos indivíduos.

1.5 Contribuições

As principais contribuições desta dissertação são:

1. **Detecção de grupos**: Propomos uma metodologia baseada em grafos para identificar possíveis associações de indivíduos no ambiente, onde definimos matematicamente um modelo de interação de acordo com a posição e orientação corporal entre os pares de indivíduos presentes no ambiente, sendo possível segregar diferentes grupos de interação de acordo com as zonas sociais especificadas nas regras de proxemics. Em consequência da análise de interações ser realizada considerando pares de indivíduos, a nossa metodologia garante escalabilidade na determinação de um grande número de indivíduos pertencentes a um mesmo grupo de interação.

2. **Representação do espaço social**: Com base na detecção dos grupos de interação, propomos uma representação do espaço social coletivo a partir da junção de funções Gaussianas assimétricas para delimitação do espaço socialmente aceitável. A determinação de tais limites sociais oferece a possibilidade de tratar os grupos de forma eficiente como entidades únicas, delimitando o acesso indevido a sua zona de conforto e garantindo flexibilidade no uso de abordagens previamente desenvolvidas no contexto de grupos de pessoas.

1.6 Organização do Texto

O restante deste trabalho está organizado da seguinte forma: O Capítulo 2 apresenta uma revisão da literatura especifica sobre o espaço pessoal e a identificação de grupos. A metodologia proposta é apresentada em detalhes no Capítulo 3 e validada por um conjunto de experimentos em um ambiente simulado, cujos resultados são mostrados e discutidos no Capítulo 4. Em conclusão, o Capítulo 5 apresenta orientações sobre a elaboração de trabalhos futuros.

Capítulo 2

Trabalhos Relacionados

Neste trabalho, investigamos o problema de identificar corretamente a associação de indivíduos no ambiente, a fim de organizá-los em grupos. Estes grupos são utilizados para a determinação de um espaço social coletivo, o que é útil em diferentes cenários. Vários trabalhos na literatura tratam de partes separadas desse problema, e cada aspecto fundamental é discutido nas seções a seguir.

Existem muitos tipos possíveis de interação entre robôs e seres humanos, enquanto o termo HRI pode principalmente evocar imagens de robôs falando com seres humanos ou realizando a entrega de objetos, o mesmo também compreende atividades relativas a navegação de robôs [36]. Na navegação, a interação com os seres humanos acontece tanto na navegação autônoma consciente do ser humano [59, 51], como na abordagem de pessoas [53, 7] ou acompanhamento de humanos [12].

É importante ressaltar que para realizar uma navegação socialmente aceitável, o robô deverá executar um comportamento similar ao que seria adotado por um ser humano. Neste sentido, é importante prover ao robô a percepção de determinadas convenções sociais adotadas pelos seres humanos para que o mesmo consiga interagir pelo ambiente de forma natural e sociável.

2.1 Espaço Social

Convenções sociais espaciais são uns dos aspectos mais pesquisados para tentar caracterizar e descrever a forma que um indivíduo se relaciona com as pessoas ao seu entorno. Diferentes trabalhos sugerem que as mesmas zonas interpessoais observadas na interação humano-humano também podem ser consideradas em cenários de interação humano-robô. Essa noção de espaço social pessoal é um dos aspectos mais utilizados em diversos tópicos de pesquisa HRI. Hall [27] observou que a distância interpessoal de uma pessoa pode ser dividida basicamente em quatro zonas distintas, onde as distâncias relativas entre as pessoas podem variar de acordo com o local ou a depender da cultura

em que estão imersas [4]. Estas quatro zonas são classificadas em espaço Público, Social, Pessoal e Íntimo (Figura 2.1).

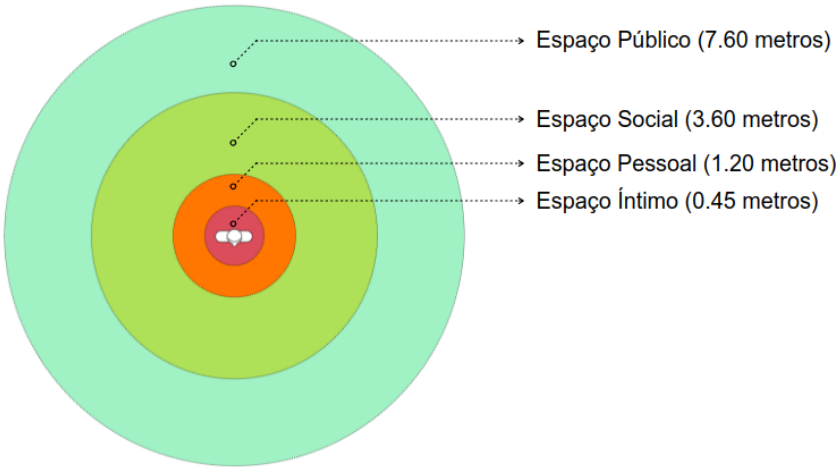


Figura 2.1: Zonas de reação pessoal, revelando quanto espaço as pessoas geralmente precisam em várias situações sociais.

Recentemente, foi realizado um estudo baseado na teoria clássica de Hall (1966), tentando relacionar as distâncias sociais, pessoais e íntimas preferidas em 42 países [1]. Os participantes tinham que responder a uma tarefa gráfica para avaliar o quão perto uma Pessoa B poderia se aproximar, para que eles se sentissem confortáveis ao interagir com: a) um estranho, b) um conhecido, e c) uma pessoa próxima. A Figura 2.2 mostra os resultados gráficos da comparação média para a distância interpessoal em todos os 42 países avaliados. Os valores exatos obtidos para o Brasil foram: $101.21\,\text{cm} \pm (40.05)$ para a distância Social; $78.00\,\text{cm} \pm (35.27)$ para a distância Pessoa; e $54.04\,\text{cm} \pm (40.50)$ para a distância Íntima. Os demais resultados correspondentes aos outros países podem ser encontrados nos materiais suplementares disponibilizados por Al-Hamdani [1] em `http://journals.sagepub.com/doi/suppl/10.1177/0022022117698039`.

Pesquisas em ciências cognitivas mostram que o espaço pessoal geralmente pode ser modelado de forma oval, assumindo uma largura maior em direção à frente de uma pessoa [37, 49, 67]. As principais técnicas utilizadas para modelar o formato do espaço pessoal conforme definido na literatura se diversificam em abordagens que empregam formas geométricas pré-definidas, funções Gaussianas assimétricas e aprendizado de máquina.

Considerando inicialmente as formas pré definidas, Ohki et al. [43] propôs um método simplificação, onde o espaço pessoal dos indivíduos foi definido como uma forma retangular constante, adotando para o comprimento da frente, comprimento traseiro e largura os respectivos valores de 1.5m, 0.4m e 0.4m. Aplicando o mesmo conceito de utilização do espaço interpessoal como uma forma geométrica pré definida, Nakauchi and Simmons [40] apresentou o *design* e a implementação de um robô que permanece em linha tanto quanto as pessoas fazem. O sistema proposto emprega visão estéreo para reconhecer filas de pessoas e usa o conceito de espaço pessoal para modelar o comportamento social.

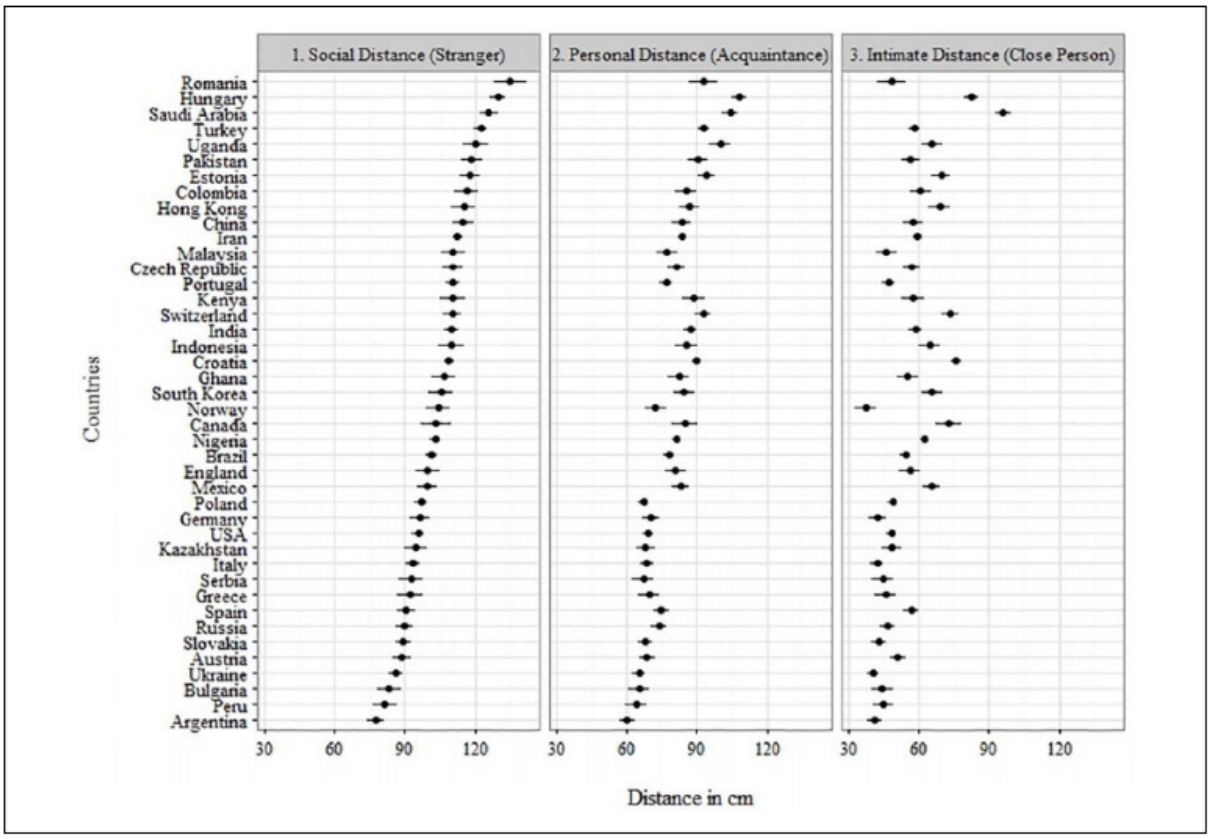


Figura 2.2: Valores médios (em centímetros) para as distâncias social, pessoal e íntima em todas as 42 nações avaliadas. *Nota.* A não-sobreposição dos intervalos de confiança entre dois países indica diferenças significativas na média. As médias para a distância interpessoal considerando estranhos estão classificadas de forma ordenada [1].

O modelo do espaço pessoal é baseado em medidas de pessoas formando filas, esse espaço pessoal é usado tanto para detectar o fim de uma fila quanto para determinar a quantidade de espaço a ser deixada entre o robô e a pessoa a sua frente. O tamanho do espaço pessoal foi então calculado realizando experimentos com 10 indivíduos. O tamanho médio do espaço pessoal estimado a partir desses experimentos é ilustrado na Figura 2.3.

Seguindo a representação do espaço pessoal por intermédio de Gaussianas, Amaoka et al. [3] propõe um modelo matemático para o conceito de espaço pessoal (Figura 2.4) e o aplicou para simular a comunicação não-verbal entre agentes em mundos virtuais. A distância entre duas pessoas reflete o tipo de relacionamento. Primeiramente é modelado o espaço em torno de um agente como uma função de distribuição de probabilidade que reflete em cada ponto do espaço a importância desse ponto para o agente. O agente atualiza dinamicamente esta função de acordo com (1) sua relação com o outro agente, (2) sua orientação do rosto, e (3) a evolução do relacionamento ao longo do tempo como um agente estranho pode se tornar um amigo.

De forma análoga a representação anterior, Kirby [34] desenvolveu uma *framework* de navegação para tarefas sociais físicas humano-robô, como navegar pelas multidões ou esperar em fila. A *framework* denominada *Constraint-Optimizing Method for Person-*

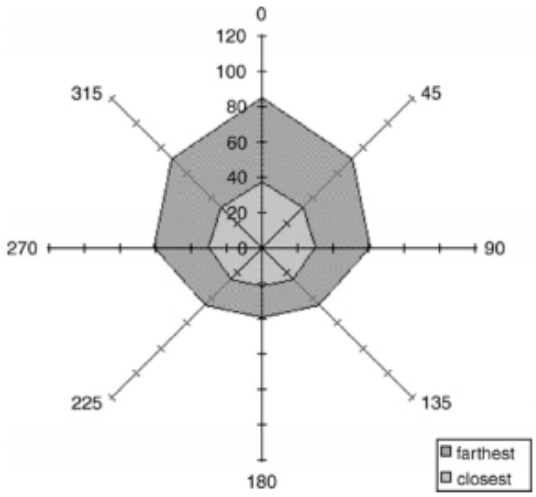


Figura 2.3: Tamanho real do espaço pessoal quando as pessoas estão em fila [40].

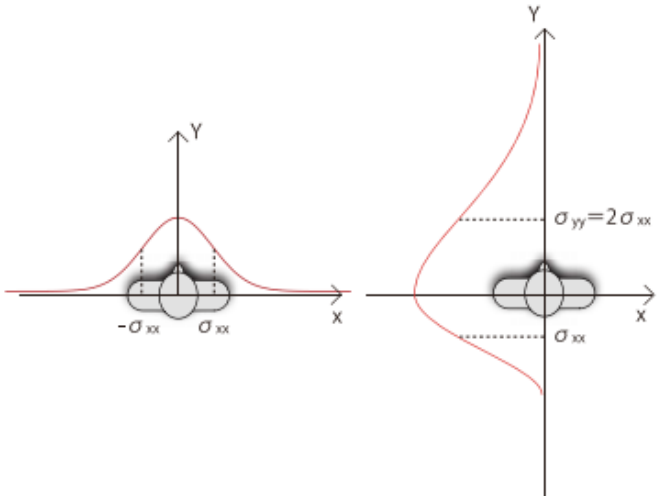
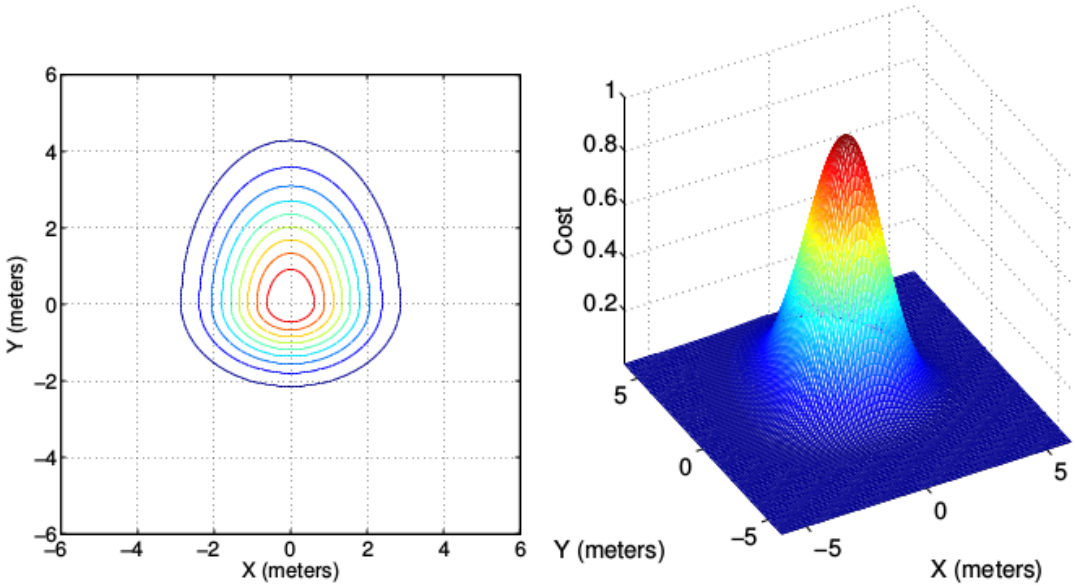


Figura 2.4: Modelo do espaço pessoal com base na orientação do rosto. O espaço pessoal na área frontal de uma pessoa é mais amplo que as áreas traseiras e laterais [3].

Acceptable NavigatION (COMPANION) é uma generalização para representar convenções sociais como componentes de um problema de otimização de restrições. Convenções sociais, como espaço pessoal e tendência para a direita são descritos como funções de custo matemático, além de possuir uma representação variável de acordo com a velocidade de deslocamento da pessoa (Figura 2.5). Estes custos são então utilizados no planejamento do caminho, juntamente com métricas mais típicas relacionadas à tarefa, como a distância mais curta.

Papadakis et al. [46] propôs um método com base no uso de funções de densidade de probabilidade normal não-estacionárias para o espaço de indivíduos e no tratamento de interações espaciais multi-pessoas através do mapeamento social. A *framework* probabilística é capaz de instanciar diversos modelos de espaçamento social e representar dimensões distintivas na interação humano-robô. (Figura 2.6).

Alternativamente as representações descritas anteriormente, Mead and Matarić



(a) Contour map (b) Surface plot

Figura 2.5: Custo do espaço pessoal para uma pessoa que se desloca a 1.0m/s ao longo do eixo Y positivo (para cima) [34].

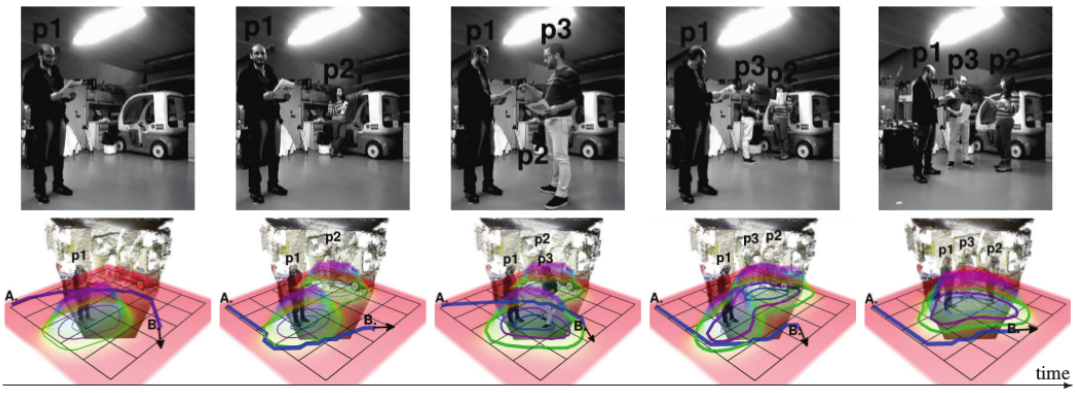


Figura 2.6: Ilustração do mapeamento das interações sociais de três participantes ($\mathbf{p}_1$, $\mathbf{p}_2$ e $\mathbf{p}_3$) usando a abordagem proposta por Papadakis et al. [46] em vários intervalos de tempo. O plano de trajetória sobreposto entre os pontos **A** e **B** corresponde à trajetória que melhor respeita a sensibilidade social coletiva reconstruída.

[38] apresentou uma *framework* computacional de proxemics com base em modelos probabilísticos baseados em dados de como os sinais sociais (fala e gesto) são produzidos (por um ser humano) e percebidos (por um robô). A *framework* e os modelos foram implementados como sistemas autônomos de comportamento proxemic para robôs sociáveis, incluindo: (1) um método baseado em amostragem para a estimativa do estado objetivo de proxemic do robô em relação aos parâmetros de distância e orientação humano-robô, (2) um controlador proxemic reativo para a realização do estado objetivo, e (3) um planejador de trajetória baseado em custos para maximizar as taxas automáticas de reconhecimento de fala e reconhecimento de gestos ao longo de um caminho para o estado objetivo.

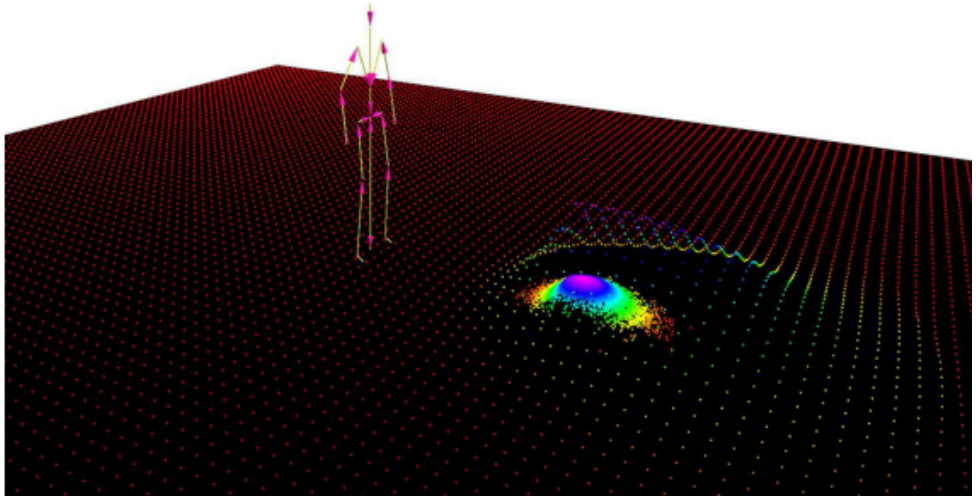


Figura 2.7: O método de estimativa do estado proxemic objetivo usa uma abordagem baseada em amostragem para estimar como o ser humano produz sinais sociais, e como o robô os percebe usando modelos baseados em dados. As estimativas baixas e altas são indicadas pela cor (vermelho a magenta) e altura (baixa a alta) de pontos amostrados no plano do chão (mostrado em preto). O agrupamento de pontos denota o impacto do processo de reescalonamento. O método usa essas estimativas para selecionar uma ação de interação que maximize o desempenho esperado [38].

2.2 Percepção de Grupos de Pessoas

A seguir serão apresentadas pesquisas que buscam desenvolver sistemas automatizados com capacidade de identificar grupos de interação. Os principais trabalhos possuem o enfoque na análise da postura corporal e nas regras de proxemics através da análise dos dados de câmeras exocêntricas, fixadas de cima para baixo. As metodologias propostas se subdividem em abordagens baseadas em Hough Voting, modelagem de grafos ou na utilização de aprendizado de máquina.

2.2.1 Grupos de Interação

A percepção de interações sociais em grupos de pessoas também é uma informação de grande utilidade para um robô, pois para viabilizar um comportamento colaborativo, se torna necessário prover ao robô uma compreensão da dinâmica de grupo. Expressões da atitude de alguém em relação a uma situação social são manifestadas através de uma multiplicidade de pistas comportamentais não-verbais, tais como a postura corporal, gesto, olhar, expressão facial, análise conversacional e proxemics [71] (Figura 2.8).

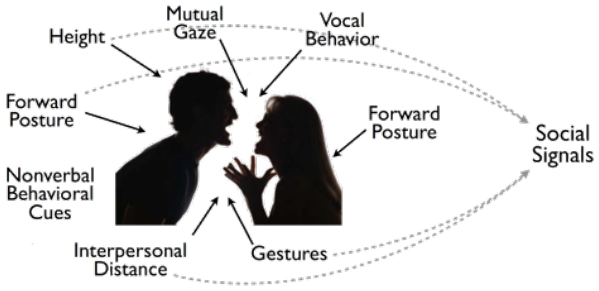
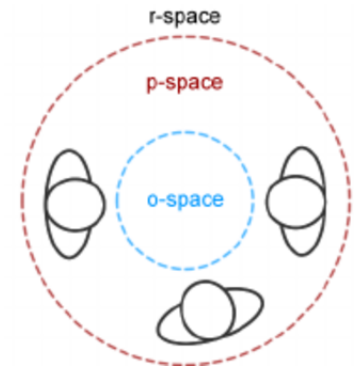


Figura 2.8: Exemplo da multiplicidade de pistas comportamentais não-verbais expressada na atitude de alguém em relação a uma situação social, tais como a postura corporal, gesto, olhar, expressão facial, análise conversacional e proxemics [71].

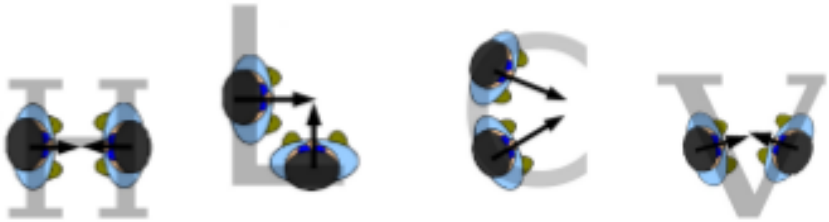
Durante as interações sociais, uma característica importante de ser observada é a disposição espacial dos indivíduos perante o grupo. O arranjo espacial durante uma interação social objetiva atender a duas necessidades principais [15]: a primeira é prover a todas as pessoas envolvidas a possibilidade de participar da interação, a segunda é estabelecer uma separação entre um grupo de interação e indivíduos externos, no caso de haver outras pessoas no ambiente. Os padrões formados durante as interações sociais, que descrevem como as pessoas ajustam sua posição e orientação para interagir de forma conjunta e gerenciar conjuntamente sua atenção, são atribuídos ao sistema F-formation concebido por Kendon [32].

F-Formation é uma formação sócio-espacial que caracteriza grupos de duas ou mais pessoas que estabelecem e mantêm uma relação espacial e orientaçional. O padrão F-formation segue três espaços sociais (Figura 2.9a). O *o-space* é um espaço convexo vazio ao qual todo mundo reunido tem acesso direto, fácil e igual. O *p-space* é a área em torno do *o-space*, onde os indivíduos estão posicionados. O *r-space* é atribuído a toda área além do *p-space*. Kendon [32] também descreve como as pessoas costumam se estabelecer sob a forma de um círculo, elipse, ferradura, lado a lado ou forma de L, para que elas possam ter acesso fácil e preferencial entre si, enquanto excluem distrações exteriores ao grupo (Figura 2.9b). Um levantamento e discussão mais aprofundado sobre algoritmos de percepção de robôs para entender a dinâmica de grupo humano no contexto de HRI é

apresentado em Taylor and Riek [61].



(a) Padrão sócio-espacial [16].



(b) Tipos de formação considerando um par de interlocutores [41].

Figura 2.9: Arranjos espaciais de F-formation humana: (a) representação sócio-espacial do padrão F-Formation [32]. (b) representação dos F-formations mais frequentes para grupos de 2 pessoas [11].

Em Cristani et al. [14], a detecção de F-Formations foi realizada empregando elementos relacionados a proxemics e estimativas de orientação da cabeça. A abordagem baseia-se na modelagem do *o-space* como uma distribuição Gaussiana aleatória, e utilizou uma estratégia de transformada de Hough para representar uma modelagem precisa da definição formal de F-Formations. Como principais características apresentadas, as pessoas devem estar razoavelmente próximas umas das outras, devidamente orientadas para o *o-space*, onde o mesmo deve estar vazio, permitindo que os indivíduos realizem contato visual (Figura 2.10).

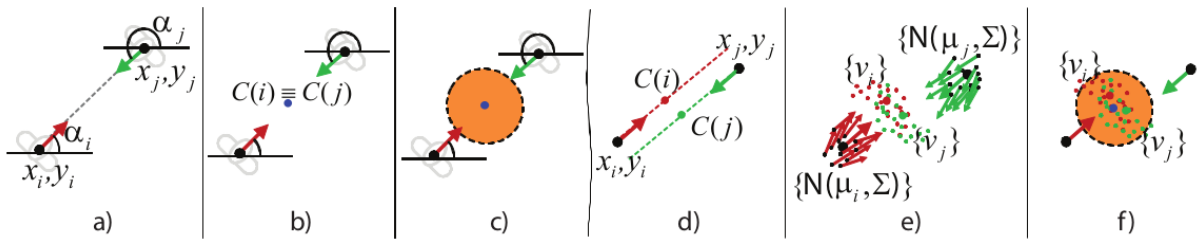


Figura 2.10: Exemplificação da abordagem proposta. (a-c) Dois indivíduos posicionados frontalmente a um voto de distância fixa para um mesmo centro da circunferência do *o-space*. (d) Os indivíduos não estão de forma exatamente frontal em casos reais. (e-f) Várias posições e orientações da cabeça são desenhadas a partir de distribuições Gaussianas de modo a lidar com a incerteza de cenários reais [14].

Alternativamente, em Hung and Kröse [31], um método de estimar F-Formations é proposto usando um algoritmo de agrupamento em grafos, individualizando o clique máximo em grafos ponderados, onde cada nodo no grafo representa uma pessoa e as arestas entre eles medem a afinidade entre pares (Figura 2.11).

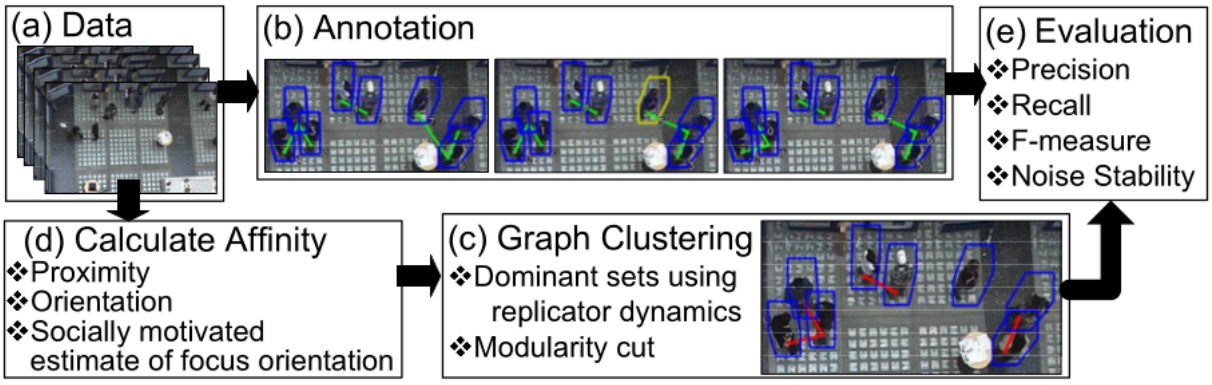


Figura 2.11: Fluxograma resumindo a abordagem. Linhas verdes: F-formations anotados, Caixa amarela: associado, Linhas vermelhas: detecções [31].

Setti et al. [54] comparou a abordagem baseada em Hough com a estratégia baseada em grafos, destacando os pontos fortes e fracos de ambos em diferentes cenários do mundo real e também analisou quais aspectos do problema são importantes para abordar a detecção de F-formation. O estudo conclui que, embora o método de Cristani et al. [14] seja mais estável usando informações de orientação da cabeça na presença de ruído, o método de Hung and Kröse [31] obteve um melhor desempenho na presença de informações posicionais. Setti et al. [55] também propôs uma extensão multi-escala da abordagem baseada em Hough, modelando explicitamente as F-formations para lidar com grupos de cardinalidades diversas. Sendo capaz de melhorar a estratégia original (Figura 2.12).

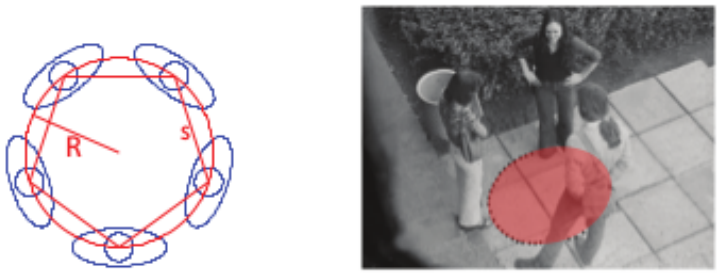


Figura 2.12: Exemplo de disposição poligonal de 5 pessoas em um F-formation (Esquerda) e um real F-formation com o relacionado *o-space* destacado em vermelho (Direita) [55].

Seguindo nas abordagens baseadas em grafos, Tran et al. [62] aplicou um algoritmo semelhante de agrupamento inspirado no princípio do conjunto dominante para descobrir grupos de interação social. Por definição, um agrupamento deve ter alta homogeneidade interna e deve ter alta heterogeneidade entre as entidades do agrupamento e as entidades externas. Da mesma forma, os grupos de interação social devem ter fortes interações

dentro de seus membros e devem ter interações mais fracas com os que estão fora do grupo. Para medir quantitativamente a extensão da relação semântica ou a interação entre as pessoas na cena, foi modelado um grafo ponderado não direcionado construído usando todas as pessoas na cena como vértices e as conexões entre elas foram ponderadas usando a relação ou interação medida. A interação baseia-se na aproximação da atenção de um indivíduo como uma elipse centrada em um deslocamento fixo na frente do indivíduo. Quanto mais se sobrepõem as elipses de atenção relacionadas a dois indivíduos, maior a interação entre eles (Figura 2.13). Já Setti et al. [56] propôs uma *framework* com base no corte em grafos para a detecção de F-formation baseadas nas estimativas de posição e orientação dos alvos, onde o *o-space* de uma F-formation é encontrado atribuindo-lhe os indivíduos cujos segmentos transacionais se sobrepõem, sem se concentrar em um arranjo particular. A abordagem é iterativa, e começa assumindo um número arbitrariamente elevado de F-formation: depois disso, uma otimização Hill Climbing alterna entre atribuir indivíduos a grupos usando a otimização baseada em corte de grafos e atualizando os centros das F-formations, podando grupos não suportados de acordo com um comprimento de descrição mínimo anterior. As iterações continuam até a convergência do algoritmo, o que é garantido.

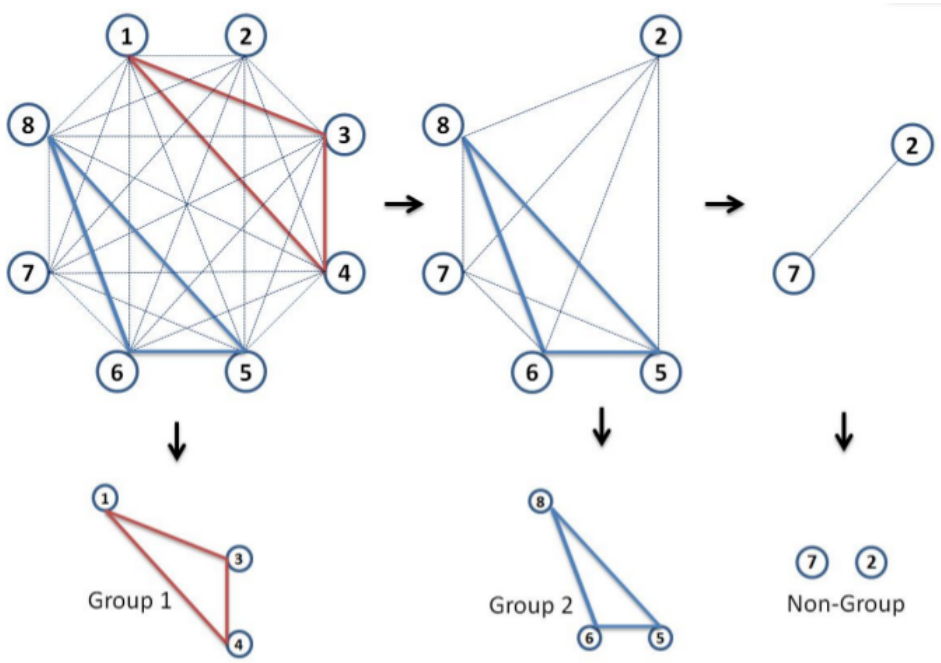


Figura 2.13: Ilustração da descoberta do grupo interagindo: Dois grupos são descobertos e duas pessoas que não pertencem a qualquer grupo. As pessoas não pertencentes ao grupo são eliminadas na análise da atividade grupal. [62].

Englobando as metodologias fundamentadas em aprendizado de máquina, Vascon et al. [66] desenvolveu uma *framework* com teoria dos jogos, apresentando uma modelagem estatística da incerteza associada à posição e orientação das pessoas. Especificamente, foi empregado uma representação da afinidade entre pares de pessoas candidatas,

expressas como uma distância entre distribuições sobre a região de atenção orientada mais plausível. Além disso, é realizada a integração das informações temporais em vários *frames* para suavizar o ruído na orientação da cabeça e apresentar estimativas usando noções da teoria evolutiva dos jogos *multi-payoff* (Figura 2.14). Em Varadarajan et al. [65], também é apresentada uma *framework* unificada para inferir as estimativas das orientações da cabeça, corpo e F-formation nos cenários sociais. O trabalho é inspirado pela *framework* de estimativa acoplada da pose da cabeça e do corpo, proposta em Chen and Odobez [10]. No entanto, tendo como diferencial a comparação da estimativa de pose da cabeça e corpo com a detecção de F-formation. Especificamente, a *framework* de aprendizagem proposta explora os dados rotulados e não rotulados através de uma regularização múltipla para aprender a gama de orientações de cabeça-corpo conjuntamente possíveis, explora as restrições posicionais e baseadas em pose, relacionando interações para descobrir F-formations e iterativamente refina estimativas de pose de interações baseada no conhecimento de F-formation e vice-versa (Figura 2.15).

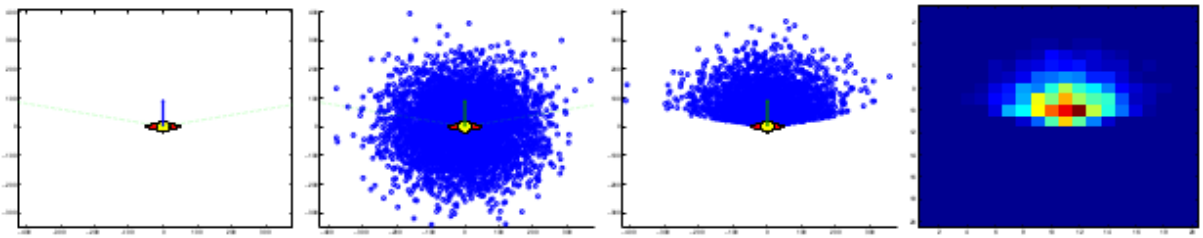


Figura 2.14: Na figura é mostrado o processo de geração do tronco: a) dado a posição e orientação da i-ésima pessoa um cone de abertura $\alpha = 160°$ é sobre imposto, b) o conjunto Gaussiano 2D de amostras é gerado, c) apenas as amostras biologicamente viáveis são mantidas, d) *bin* do espaço em uma grade de 20×20 para obter a representação final do histograma h_i [66].

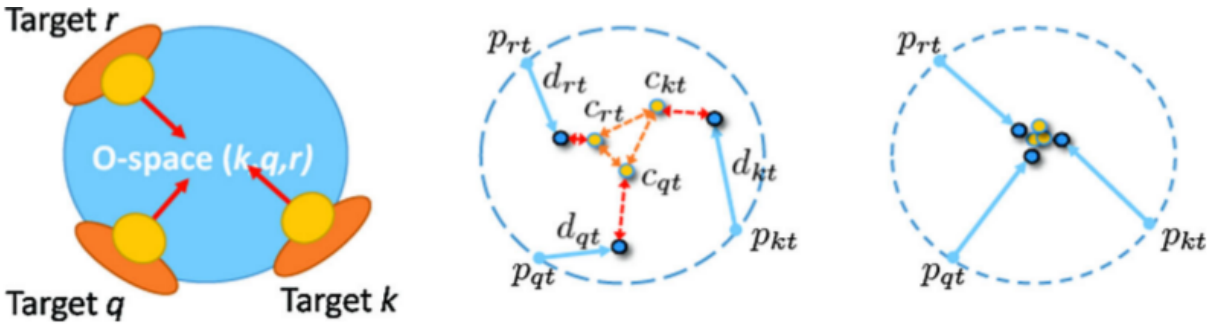


Figura 2.15: (Esquerda) *O-space* do F-formation envolvendo três alvos k, r e q com a sua pose do corpo. (Centro) vetores de direção $d_{(.)}$ obtidos através do regressor de pose do corpo são mostrados usando setas azuis, enquanto $c_{(.)}$ (pontos amarelos) denotam os locais do centro eleitos. Ao minimizar a função de perda, é refinado a pose do corpo e as estimativas de F-formation para chegar a menor configuração de perda (Direita), onde os centros votados para cada *cluster* alvo no centroide *o-space* [65].

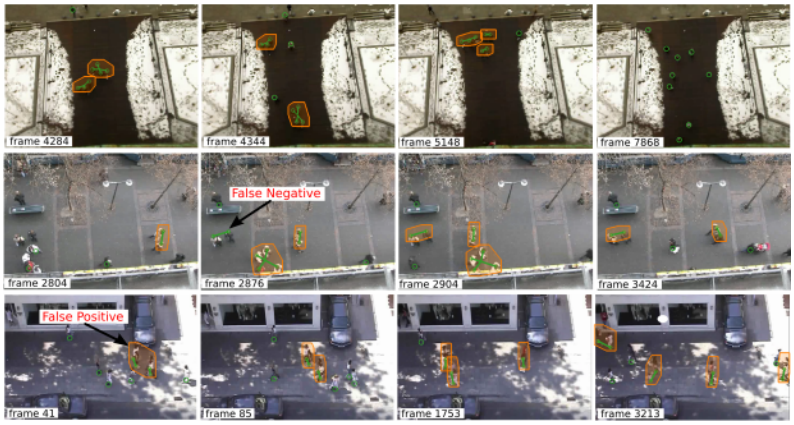


Figura 2.16: Resultados qualitativos para os *datasets eth* (primeira linha), *hotel* (segunda linha) e *zara01* (último linha). O *ground truth* da posição dos indivíduos e grupos é mostrado com círculos e segmentos verdes, respectivamente. Os grupos estimados são representados com o invólucro convexo laranja [74].

2.2.2 Rastreamento de Grupos Sociais

Na subseção anterior foram apresentados os principais trabalhos relacionados a detecção de F-formations. Entretanto, a definição formal do F-formation é adequada apenas para grupos estacionários e não está definida para grupos em movimento, caso que não pode ser completamente ignorado na análise de multidões. Descreveremos a seguir alguns trabalhos que realizam a detecção e rastreamento de grupos sociais por intermédio da extração de características comportamentais e ambientais dos indivíduos pertencentes a multidão.

Yamaguchi et al. [72] apresentou um modelo comportamental de pedestres para estimar propriedades pessoais ocultas e melhorar o desempenho do rastreamento em cenários realistas, onde a avaliação da estimativa do destino e do grupo social, juntamente com o erro de predição comportamental, sugere que é possível obter uma estimativa razoável de informações pessoais não observáveis apenas a partir de dados puramente comportamentais e ambientais (Figura 2.17). No mesmo contexto, Zanotto et al. [74] apresentou um método não supervisionado destinado a detectar formações de grupo em sequências de vídeos, que se baseia em um processo de inferência on-line (*Dirichlet Process Mixture Models*). O *framework* variacional sequencial proposto permite realizar inferências em tempo real, enquanto as restrições sociais baseadas nas regras de proxemics asseguram a produção de hipóteses de grupo adequadas consistentes com a percepção humana (Figura 2.16). De forma mais abrangente, no algoritmo de detecção proposto por Shao et al. [57], foram introduzidas várias propriedades independentes da cena (por exemplo, coletividade, estabilidade e conflito) para grupos de multidões com base no aprendizado da transição coletiva anterior (Figura 2.18).

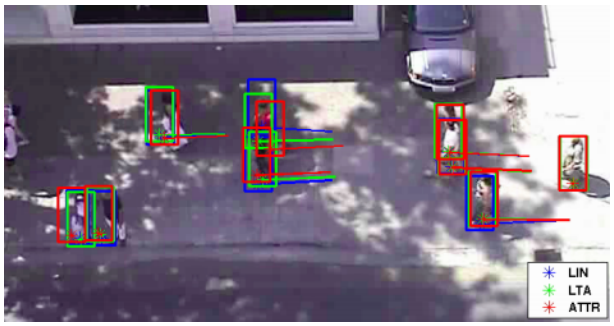


Figura 2.17: Exemplo do rastreador. A linha indica o deslocamento do quadro anterior. Menos de 0,625 fps, é difícil encontrar uma correspondência entre quadros sem prévia [72].

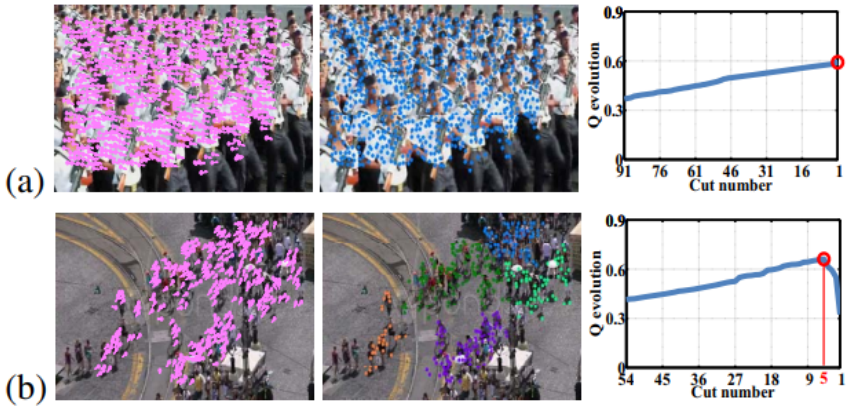


Figura 2.18: (a) e (b) mostram um grupo uniforme e não uniforme, respectivamente. Da esquerda para a direita, é demonstrado a detecção de grupo coerente original, os subgrupos obtidos através de agrupamentos adicionais e o número ideal de cortes inferidos pela função de modularidade [57].

Distintivamente das abordagens supracitadas, o algoritmo proposto por Ge et al. [22] utiliza uma técnica baseada em agrupamento, capaz de detectar pequenos grupos de indivíduos que viajam juntos. Esses grupos são descobertos por agrupamento hierárquico ascendente usando uma distância simétrica de Hausdorff generalizada (Figura 2.19).

Em um recente trabalho, Solera et al. [60] propôs um novo algoritmo para detectar grupos sociais em multidões por meio de um procedimento de agrupamento de correlação nas trajetórias das pessoas. A afinidade entre os membros pertencentes a multidão é aprendida utilizando uma técnica supervisionada de aprendizado de máquinas, através de uma formulação *on-line* do *framework Structural Support Vector Machine* (SSVM). A metodologia proposta foi avaliada contra os trabalhos de detecção de grupos sociais citados anteriormente, superando seus resultados e alcançando assim o estado da arte. (Figura 2.20).

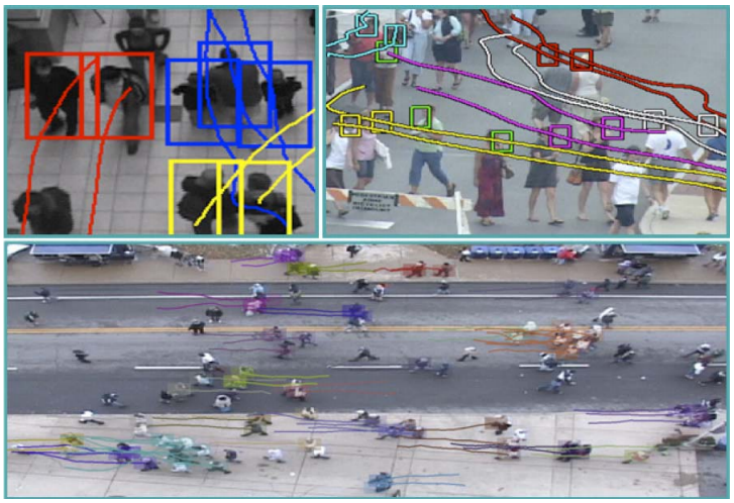


Figura 2.19: Grupos pequenos são prevalentes em cenas de pedestres. O algoritmo proposto detecta grupos de pessoas que viajam juntas através de agrupamento hierárquico em trajetórias extraídas automaticamente de vídeo de multidões em várias condições [22].

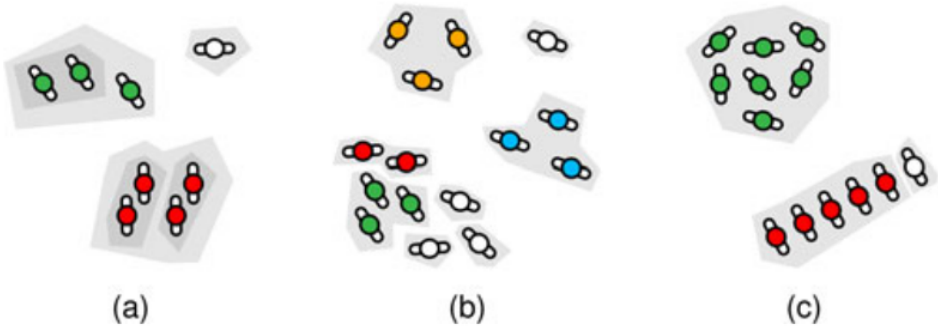


Figura 2.20: Destaques das propriedades dos grupos sociais: (a) coerência hierárquica, (b) invariância de densidade e (c) transitividade. [60].

2.3 Aplicações de Robôs em Contextos Sociais

Conforme discutido na subseção anterior, o problema de identificar e associar corretamente grupos de pessoas é um desafio em si, e aplica técnicas de diferentes áreas como a visão computacional e o aprendizado de máquina (Seção 2.2). Outra questão é como determinar e representar os limites sociais levando em consideração esta nova *entidade combinada*. A inclusão do espaço social dos indivíduos (preceito abordado na Subseção 2.1) durante o planejamento de trajetórias, corrobora para a criação de rotas mais naturais, ou seja, condizentes com o caminho que um ser humano provavelmente executaria. Desta forma, contribuindo para que as pessoas presentes no ambiente possuam uma sensação maior de conforto em relação a movimentação realizada pelo robô durante seu processo de navegação e aumentando assim a sociabilidade da interação humano-robô. Algumas publicações se concentram em diversos outros aspectos que não possuem relação

direta com o conforto, a naturalidade e a sociabilidade do planejamento de trajetórias realizado pelo robô na presença de multidões. Muitas vezes, o foco principal é permitir novas formas de interação humano-robô, como mover-se em formação ou otimizar outros aspectos para mover-se entre seres humanos, objetivando eficiência de tempo ou robustez. A seguir, discutiremos as mais diferentes aplicações que fazem menção a navegação social, abordagem de pessoas, atividades de escolta e tarefas relativas a seguir pessoas.

2.3.1 Navegação Autônoma

Com o objetivo de permitir que robôs naveguem eficientemente em multidões, [30] aplica o conceito de *Inverse Reinforcement Learning* (IRL), capturando aspectos da navegação de multidões, utilizando a densidade e direções médias do fluxo como *features* para aprender o comportamento da movimentação humana (Figura 2.21). No entanto, como abordado no trabalho desenvolvido por Trautman and Krause [63] (Figura 2.22), uma vez que o ambiente supera um certo nível de complexidade, o planejador decide que todos os caminhos de avanço são inseguros e o robô congela no local (ou executa manobras desnecessárias) para evitar colisões. Neste sentido, Trautman and Krause [63] estudam a navegação segura de um robô móvel através de multidões de agentes dinâmicos com trajetórias incertas, onde o seu principal *insight* é que os agentes dinâmicos resolvem o problema do robô congelado ao se envolverem em "evasão de colisão conjunta": Eles criam caminhos cooperativos para criar trajetórias viáveis. Para lidar com o problema de navegar com segurança através de multidões densas foi desenvolvido o *Interacting Gaussian Processes* (IGP), um modelo estatístico não paramétrico baseado em processos Gaussianos de saída dependente, acoplado através de um potencial de interação não-linear. Os autores demonstram como a navegação neste modelo é naturalmente moldada como uma tarefa de inferência, que pode ser aproximadamente resolvida usando amostragem de importância.

De forma mais complementar, no artigo proposto por Kruse et al. [35] é considerado que um robô deve possuir ao menos movimentos legíveis, para que assim seres humanos sejam capazes de prever suas metas e intenções, considerando que uma pessoa que se interponha no caminho de um robô pode muito bem se afastar, sem gerar nenhum tipo de desconforto caso seja abordada pelo robô que queira passar. Consequentemente, são apresentadas medidas para utilizar o existente *Human-Aware Navigation Planner* (HANP) em situações altamente dinâmicas e colaborativas. Ao introduzir a noção de planejamento otimista, HANP encontra soluções mais aceitáveis e as mais potencialmente aceitáveis para o movimento do robô na presença de humanos. O planejamento preditivo acrescenta

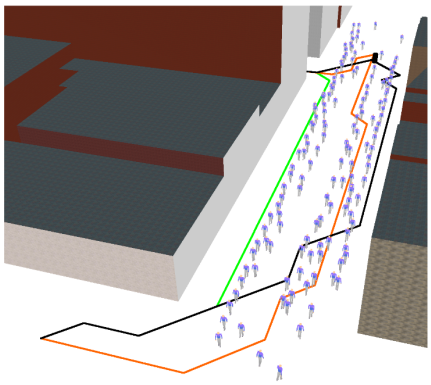


Figura 2.21: Uma visão do caminho do robô (preto), um caminho humano (laranja / cinza escuro) e o caminho mais curto (verde / cinza claro). Há dois fluxos de multidão, cada um movendo-se ao longo do lado direito da passarela. Observa-se que o robô se move com o fluxo no lado direito e depois cruza o fluxo de multidão que se aproxima de maneira semelhante ao caminho humano no final de sua jornada [30].

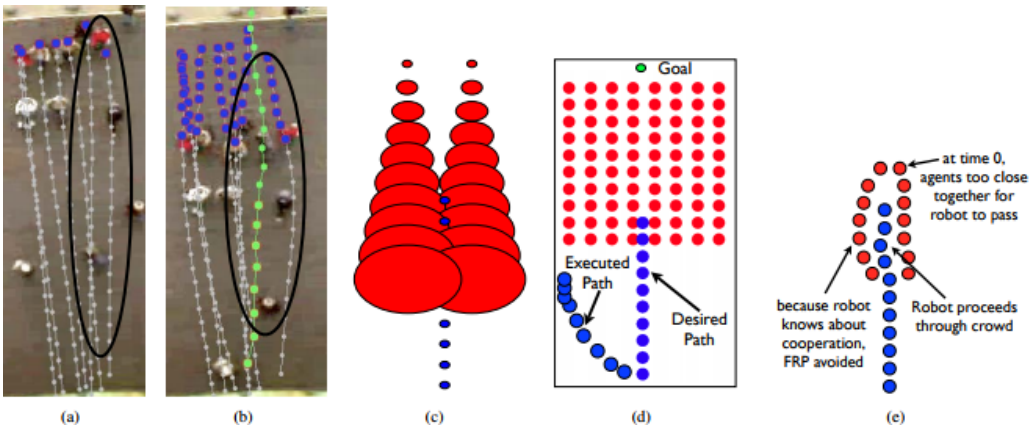


Figura 2.22: (a, b) Evidências empíricas da evasão de colisão conjunta: círculos azuis (representando a posição atual) sobre as linhas cinzas os pedestres que estão se movendo para baixo e os círculos negros são a área de interesse. Em (a), os pedestres azuis ainda não viram a pessoa verde; Suas trajetórias projetadas (em cinza) são muito estreitas. Em (b), os pontos verdes com o cerco vermelho são a posição atual do pedestre movendo-se para cima, e todos os pedestres ajustaram suas trajetórias para criar espaço - perceber o quão grande a previsão cinza se tornou. É este comportamento de prevenção de colisão conjunta capturada no artigo. (c-e) Ilustração do problema do robô congelado. Agentes de multidão dinâmicos em vermelho viajando para baixo, o robô que se pretende controlar em azul. Os múltiplos pontos indica vários pontos ao longo de uma trajetória. (c) explosão de incerteza devido a previsão não corrigida. (d) Mesmo com uma previsão perfeita, o espaço para a navegação do robô pode não existir. (e) Modelagem da prevenção de colisões cooperativa resolve o problema do robô congelado. [63].

custos específicos para explicar o movimento previsto de um ser humano no tempo de planejamento. Entretanto, utilizando essas medidas os planos gerados não são garantidos para ser livre de colisões, mas essa função é delegada ao controlador de movimento para criar um comportamento seguro, esperando em frente de outros agentes (Figura 2.23).

O trabalho proposto por Müller et al. [39] apresenta uma ideologia similar de

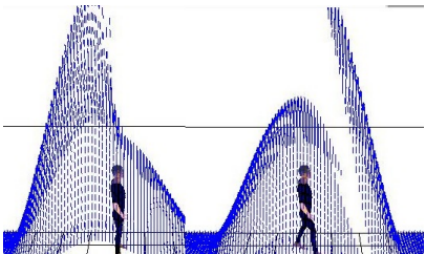


Figura 2.23: Os custos em frente do humano em movimento aumentou em comparação com o humano parado. Para um humano de pé, evita-se mover atrás do humano, atribuindo altos custos a essa área. Para os humanos em movimento, é preferido evitar mover em sua direção de movimento [35].

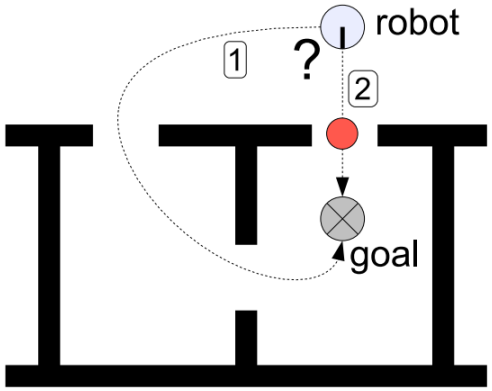


Figura 2.24: Uma pessoa bloqueia um caminho significativamente mais curto causando um desvio para o robô. Alternativamente, o robô pode tentar espantar esta pessoa e seguir o caminho mais curto [39].

atribuir legibilidade de movimentos para robôs móveis, gerando um comportamento de movimento humano para robôs em ambientes populados. No entanto, ao invés de planejar um caminho em torno das pessoas, a metodologia proposta tenta identificar e seguir indivíduos ou grupos de pessoas que parecem avançar para o mesmo objetivo. Isto é realizado por um planejamento iterativo e uma abordagem de seleção de pessoa com base no algoritmo A* que determina quais os indivíduos que o robô deve seguir. Além disso, o robô considera ações para expulsar as pessoas que bloqueiam passagens estreitas em direção ao objetivo desejado ou para evitar desvios significativos (Figura 2.24). Ainda no mesmo seguimento, mas empregando uma metodologia que começa a considerar o espaço interpessoal dos indivíduos, Cosgun et al. [13] propôs uma abordagem de planejamento de caminho que antecipa a reação das pessoas ao robô com base no modelo de força social, a partir da extração de um planejador estático usando o algoritmo de busca A*, minimizando os custos de segurança, distúrbios e o comprimento do caminho, bem como empregando um planejador dinâmico que aprimora partes do caminho que estiverem próximos de pessoas (Figura 2.25).

Em Rios-Martinez et al. [50], foi formulado um método estocástico de otimização adaptativa para minimizar o desconforto dos seres humanos no ambiente, com restrições

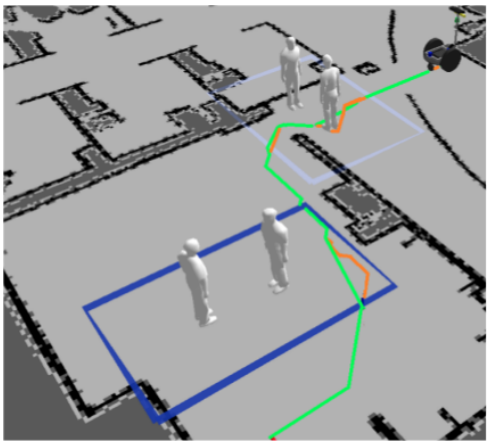


Figura 2.25: O plano estático resultante e os planos dinâmicos são mostrados em verde e laranja, respectivamente [13].

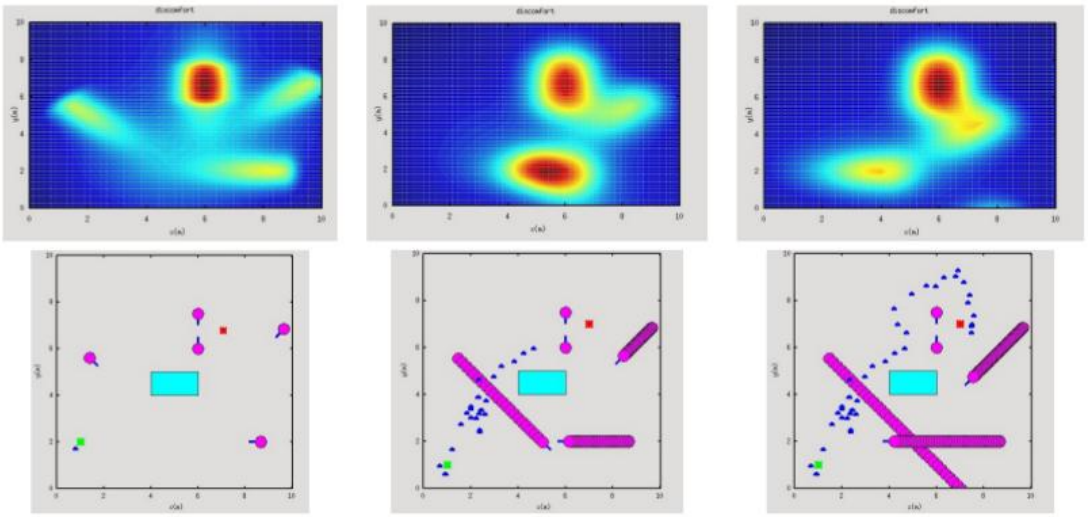


Figura 2.26: Simulação do robô navegando em um ambiente povoado por pessoas em três momentos diferentes. Três seres humanos andando e dois conversando. A função de desconforto é mostrada no topo. As pessoas são representadas por círculos, a posições do robô por pequenos triângulos, em verde e vermelho a posição inicial e final respectivamente [50].

sobre o espaço pessoal e o processo de informação das pessoas, bem como as zonas de interação. O processo de otimização é obtido usando o algoritmo *Cognitive-based Adaptive Optimization* (CAO). Este método de otimização é útil em particular quando a expressão analítica da função de otimização é desconhecida, mas os valores numéricos estão disponíveis para qualquer configuração de estado (Figura 2.26).

Em uma publicação mais recente, Pérez-Higueras et al. [47] apresentou uma abordagem para ensinar um robô comportamentos de navegação social usando demonstrações. Para este fim, um método baseado em IRL foi implementado e vinculado com um *Optimal Rapidly-exploring Random Trees* (RRT*) para aprender os pesos de sua função de custo, de modo que o planejador se comporte de forma semelhante aos comportamentos demonstrados (Figura 2.27).

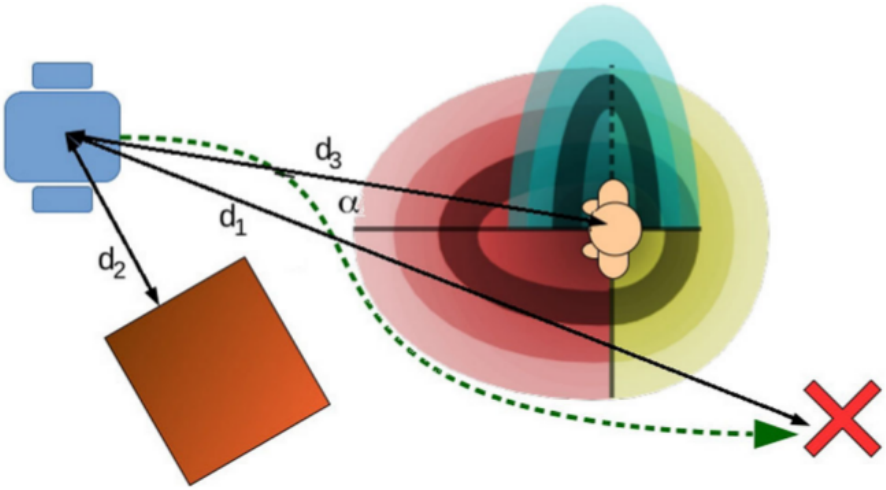


Figura 2.27: Algumas das *features* empregadas na função de custo social aprendida. d_1, distância para o objetivo. d_2, distância para o obstáculo mais próximo. d_3, distância das pessoas para o robô. α, ângulo entre a frente da pessoa e a localização do robô. As três funções Gaussianas implantadas sobre as pessoas também são representadas [47].

2.3.2 Abordando Pessoas

Usualmente em ambientes repletos por pessoas, surge a necessidade de executar tarefas na qual acabam requerendo que um determinado indivíduo aborde um grupo de pessoas e inicie uma interação. Para o caso de um robô de serviço incorporado ao mesmo cenário, as habilidades requeridas para a realização das mesmas tarefas não seriam diferentes. Fornecer este tipo de aptidão a um robô é um desafio ainda explorado por pesquisas relacionadas a HRI.

Gómez et al. [24] apresentou um modelo de espaço pessoal para os indivíduos e grupos humanos utilizando uma mistura de duas funções Gaussianas, e desenvolveu um *framework* teórico para uma série de sub-problemas introduzidos do planejamento de caminho social usando um método de *fast-matching squares*. Também foi proposto um modo estendido para abordar grupos de pessoas. No entanto, este estudo considera apenas indivíduos estáticos, sem considerar informações sobre grupos sociais humanos (Figura 2.28).

Em Truong and Ngo [64], um *framework* de conforto e segurança humana é proposto, permitindo que um robô de serviço móvel navegue de forma segura e sociável em ambientes sociais. O *framework* proposto leva em consideração estados humanos (posição, orientação, movimento e poses das mãos) e informações de interação social relativas ao robô em consideração para modelar o espaço pessoal estendido e o espaço de interação social, respectivamente, a combinação da qual resulta em uma zona social dinâmica. O *framework* de conforto e segurança humano baseado em *Dynamic Social Zone* (DSZ) é

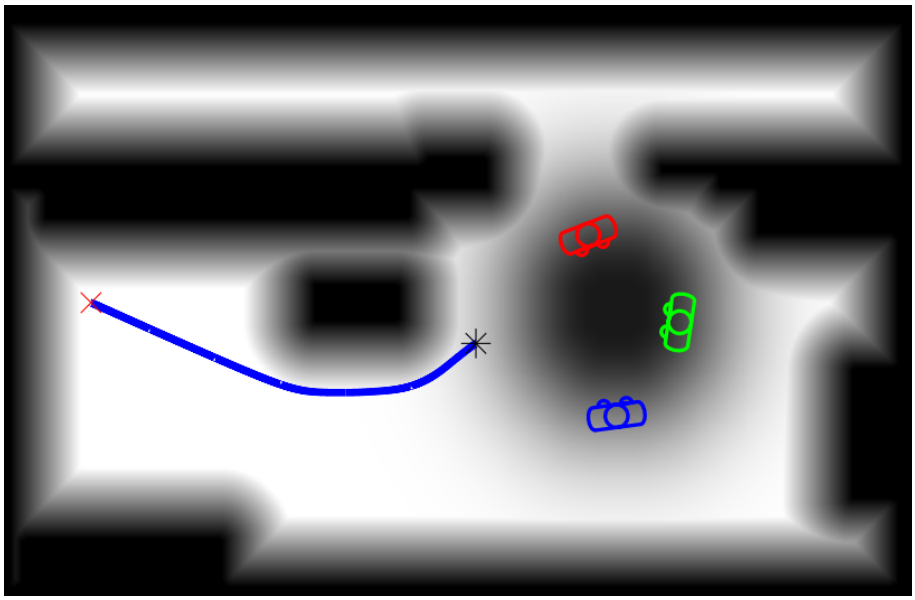


Figura 2.28: Resultados para a solução de observação, abordando um grupo de humanos [24].

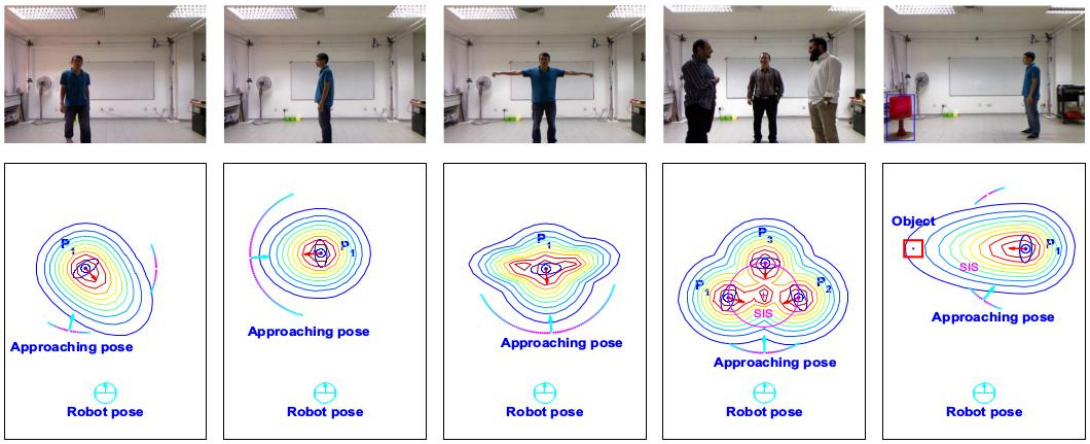
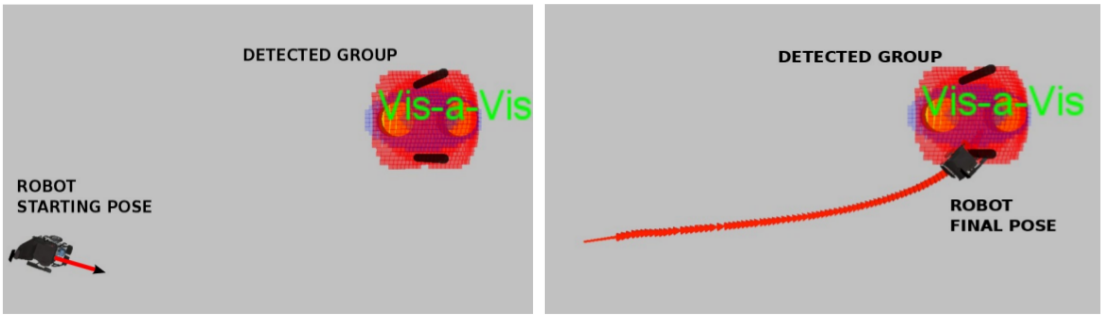


Figura 2.29: Ilustração do resultado do *framework* de conforto e segurança humano em diferentes cenários: a primeira linha corresponde aos cenários, a segunda linha apresenta o resultado da zona social dinâmica e a postura aproximada do robô correspondente aos cenários na primeira linha [64].

capaz de estimar uma pose objetivo de aproximação do robô para um ser humano ou um grupo de humanos de forma socialmente aceitável (Figura 2.29).

Em Narayanan et al. [42] foi proposto uma estratégia de movimento robótico socialmente compatível que poderia ser empregada por robôs sociais, tais como humanoides, robôs de serviço ou cadeiras de rodas inteligentes, para se aproximar e se juntar a grupos humanos em interação e então se tornar uma parte equitativa da interação (Figura 2.30).

Vázquez et al. [69] explorou métodos de aprendizado de reforço on-line para controlar a orientação de um robô durante conversas multipartidárias simuladas. Tendo como principal pressuposto, que o robô deve se voltar para o falante para transmitir a atenção e manter a consciência do foco de atenção (Figura 2.31).



(a) Início. (b) Fim.

Figura 2.30: Pose inicial e final do robô em relação a um grupo detectado para um caso simulado [42].

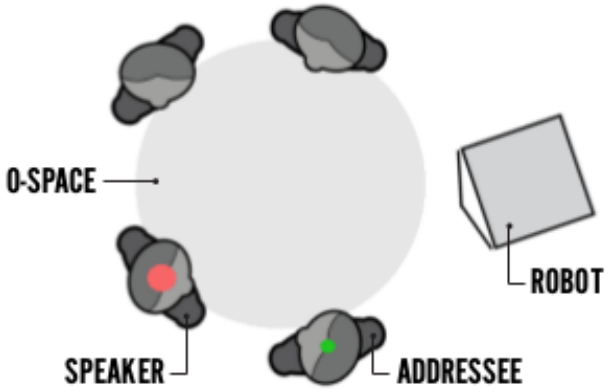


Figura 2.31: Conversação de grupo simulada entre um robô e quatro pessoas. Os círculos vermelho e verde em cima dos agentes identificam o falante e os destinatários, respectivamente. O grande círculo cinza representa o *o-space* do grupo F-formation [69].

2.3.3 Escoltando Pessoas

A atividade de orientar pessoas por ambientes restritos, como hospitais, museus e hotéis, possui uma grande demanda nos dias atuais. Na literatura específica existe uma gama de trabalhos sendo desenvolvidos que tentam solucionar os problemas relativos a liderança de pessoas [6, 18] e até mesmo grupos de animais [68] utilizando robôs autônomos.

Pandey and Alami [45] apresentou uma abordagem para evoluir o processo de orientação do comportamento típico do robô para um comportamento centrado no ser humano, que proporciona ao humano a flexibilidade para decidir sobre a forma como ele deseja ser guiado. No *framework* proposto, é o robô que se ajusta e se adapta à atividade humana, ao mesmo tempo que mantém a tarefa conjunta, e iniciar as estratégias de reengajamento no caso de ausência temporária do humano a partir do compromisso conjunto de orientação (Figura 2.32).

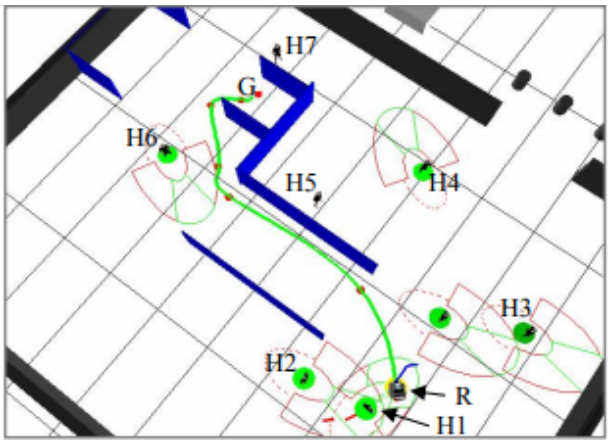
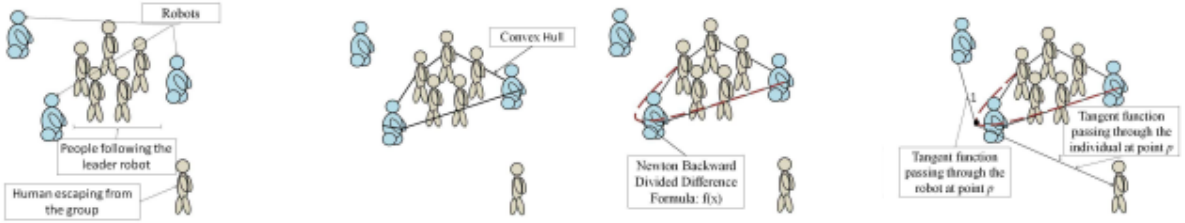


Figura 2.32: Robô aproximou-se do humano H1, para iniciar o processo de orientação. O trajeto verde mostra o caminho planejado para o objetivo após o compromisso conjunto de orientação ter sido estabelecido com H1. [45].



(a) (b) (c) (d)

Figura 2.33: (a) Representação de um ambiente com pessoas e robôs. (b) Computação do contorno convexo. (c) Interpolação do contorno convexo com fórmula *Newton Backward Divided Difference* (d) Cálculo da trajetória para resgatar o indivíduo, essa trajetória é composta por duas tangentes da função $f(x)$ no ponto p: (1) passando pelo robô pastor. (2) passando pelo indivíduo que esta escapando do grupo [21].

Garrell and Sanfeliu [21] deu um passo a frente e apresentou um método para otimizar localmente a atribuição de tarefas aos robôs para fazer suas missões. A atribuição de robôs é feita através da análise do trabalho mínimo necessário para realizar essa tarefa, onde a função a minimizar baseia-se, por um lado, pelo movimento do robô e, por outro lado, pelo impacto de tais movimentos sobre o deslocamento das pessoas. O primeiro termo leva em consideração o trabalho necessário para mover um robô de sua origem para um destino, enquanto o segundo termo analisa o impacto que os robôs têm sobre as pessoas a serem movidas, e sua computação usa a formulação de Helbing and Molnar [28] [29] (Figura 2.33).

Feil-Seifer and Matarić [20] considerou o problema de um robô movendo-se para um objetivo junto com um parceiro, demonstrando que um modelo aprendido pode ser usada para ponderar trajetórias do sistema de navegação de um robô para o movimento autônomo. O artigo apresenta a realização de um sistema de navegação consciente de pessoa que não requer ajuste de parâmetros *ad hoc*, e nenhuma entrada além de um pequeno conjunto de exemplos de treinamento (Figura 2.34).



Figura 2.34: Visão da câmera aérea com uma sobreposição de pose do robô e eixos mostrando a posição detectada do parceiro social [20].

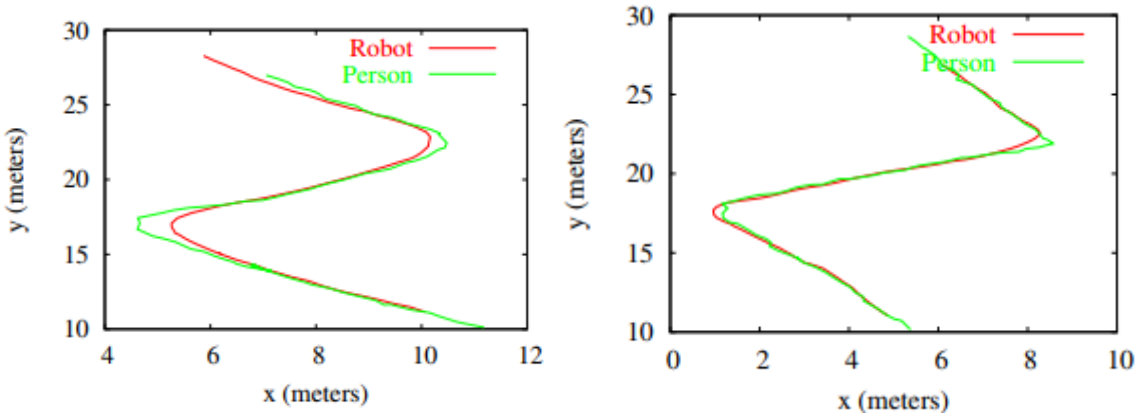
2.3.4 Seguindo Pessoas

Dentre as atividades de navegação mencionadas anteriormente que envolvem HRI, a tarefa de seguir pessoas é a que oferece um menor nível de complexidade agregado se comparada com as outras, mas nem por isto a torna menos importante.

Gockley et al. [23] apresentou um rastreador pessoal baseado em laser e descreveu dois métodos diferentes para um robô acompanhar uma pessoa: seguindo a direção e seguindo o caminho. Ambos os comportamentos foram analisados e se identificou que, embora suas performances quantitativas sejam muito semelhantes, elas são percebidas de forma diferente pelas pessoas (Figura 2.35). Especificamente, os observadores dos comportamentos do robô sentiram que o comportamento de seguimento de direção parecia mais natural e humano, e mais estreitamente alinhados as suas expectativas.

Cosgun et al. [12] apresentou um método para acompanhamento autônomo de pessoas em ambientes semelhantes a escritórios para um robô de telepresença (Figura 2.36). A abordagem proposta leva em conta a dinamicidade da pessoa e planeja os movimentos do robô pesquisando trajetórias futuras. A estrutura de planejamento é flexível e permite codificar diferentes convenções sociais para tarefas que requerem interações espaciais. Entretanto, a abordagem consegue rastrear apenas uma pessoa, não sendo robustas em situações com vários indivíduos.

De forma alternativa, no trabalho proposto por Young et al. [73] é explorado uma forma de solucionar os conjuntos de tarefas em ambientes cotidianos no qual pessoas devem direcionar ou conduzir um robô à medida que caminham em estreita proximidade um com o outro. Foi apresentado uma interface de coleira de cachorro que permite que uma pessoa conduza um robô simplesmente segurando a coleira, seguindo uma metáfora



(a) Seguindo a direção. (b) Seguindo o caminho.

Figura 2.35: Caminhos da pessoa e do robô em torno de cantos, para cada uma das duas abordagens. O robô corta drasticamente os cantos quando não seguir o caminho exato da pessoa. Deve-se observar que cada caminho mostrado é de aproximadamente $15m$ de comprimento. [23].



Figura 2.36: Usuário remoto e o acompanhante caminhando juntos enquanto conversam [12].

de interação de coleira de cachorro.

2.4 Contextualização

Conforme apresentado nas seções anteriores, existem diferentes abordagem para solucionar o problema da HRI de forma a permitir que esta interação seja o mais socialmente aceitável possível. Neste trabalho, propomos uma nova abordagem baseada em grafo para identificar possíveis associações de indivíduos no ambiente de acordo com sua posição e orientação corporal. Posteriormente, com base nessa detecção dos grupos de interação, propomos uma representação do espaço social coletivo baseada em uma função de densidade, por intermédio da junção de funções Gaussianas assimétricas para delimitação do espaço socialmente aceitável, esta é uma tarefa bastante complexa uma vez que o modelo de convenções sociais adotadas pelos seres humanos possui uma alta dinamicidade, podendo se modificar a depender do contexto local ou cultural.

A Tabela 2.1 apresenta um resumo dos principais trabalhos encontrados na literatura relacionados ao problema, destacando quais características específicas abordam, e ao final contextualizando o problema abordado neste trabalho. Pode-se observar pela distribuição das características na Tabela 2.1, uma quantidade significativa de trabalhos abordando a representação do espaço social de grupos considerando como critério de agrupamento social a relação de interação entre os indivíduos. Entretanto, esses trabalhos em sua maioria consideram um conjunto com poucos indivíduos, além de desconsiderarem a possibilidade de interações em diferentes zonas sociais. Neste ponto, nossa abordagem busca contribuir para a criação de um modelo de detecção dos grupos, que estão interagindo ou compartilhando o mesmo espaço de interação, que se aproximem das configurações observadas no cotidiano, ou seja, grupos com diferentes cardinalidades de indivíduos em distâncias interpessoais distintas.

Para este trabalho, destaca-se também a utilização da abordagem de navegação ponto-a-ponto como uma aplicabilidade do método proposto, onde pretendemos futuramente viabilizar a abordagem para as demais atividades.

Tabela 2.1: Alguns dos principais trabalhos encontrados na literatura, destacando as caracteristicas mais relevantes abordadas por cada um deles.

Trabalhos	**Espaço Social**		**Agrupamentos Sociais**		**Navegação**			
	Indivíduos	***Grupos***	***Interações***	***Grupos***	***Ponto-a-Ponto***	***Abordagem***	***Escolta***	***Follow***
[40]		•		•	•			
[34]	•				•			
[14]		•	•					
[60]		•		•				
[50]		•	•		•			
[46]		•	•	•	•			
[47]		•	•		•			
Truong and Ngo [64]		•	•	•		•		
Vázquez et al. [69]			•			•		
Pandey and Alami [45]			•				•	
Gockley et al. [23]								•
[12]			•					•
Este Trabalho		•	•	•	•			

Capítulo 3

Metodologia

Nesta seção é apresentada e discutida a metodologia proposta para o problema de determinação dos grupos de interação e seu espaço social coletivo.

Primeiramente, é abordado a representação do espaço interpessoal para o caso no qual apenas um único indivíduo é considerado, onde apresentamos a modelagem utilizada para delimitar o espaço interpessoal segundo as diferentes zonas sociais definidas nas regras de proxemics. Em seguida, o problema é estendido para o caso de múltiplos indivíduos no mesmo ambiente, onde apresentamos o modelo matemático proposto para identificação do *Level of Interaction* (*Nível de Interação*, LoI) entre pares de indivíduos. Por fim, abordamos a proposição do espaço social coletivo considerando os grupos de interação como uma entidade única, onde as diferentes zonas sociais são definidas com base em uma função de densidade.

3.1 Espaço Social Atômico

Inicialmente é apresentado o algoritmo utilizado para definir o espaço interpessoal de um único indivíduo, o qual segue o modelo de representação por intermédio de Gaussianas (como discutido no Capítulo 2.1). Posteriormente, é proposto uma modelagem baseado em grafos para identificação das possíveis interações entre pares de indivíduos, onde os dados de posição e orientação são utilizados para computar de forma determinística o LoI entre esses indivíduos.

No trabalho desenvolvido por Kirby [34], é apresentado um algoritmo capaz de modelar uma função Gaussiana assimétrica que define o custo dos pesos relativo ao espaço social de uma pessoa. Tendo como base o modelo matemático proposto, será apresentado a seguir uma reformulação para inclusão das distâncias interpessoais inerentes às regras de proxemics.

Em geral, uma função Gaussiana assimétrica bidimensional é expressa como

$$f(x,y) = A\ exp(-(a(x-x_c)^2 + 2b(x-x_c)(y-y_c) + c(y-y_c)^2)), \tag{3.1}$$

onde o coeficiente A é a altura do pico da curva, x_c e y_c representam a posição do centro do pico, e a matriz

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \tag{3.2}$$

é positiva definida, sendo que os parâmetros para a equação geral

$$a = \frac{cos^2\theta}{2\sigma^2} + \frac{sin^2\theta}{2\sigma_s^2}, \qquad b = \frac{sin\ 2\theta}{4\sigma^2} + \frac{sin\ 2\theta}{4\sigma_s^2}, \qquad c = \frac{sin^2\theta}{2\sigma^2} + \frac{cos^2\theta}{2\sigma_s^2} \tag{3.3}$$

controlam a rotação da função assimétrica bidimensional pelo ângulo θ no sentido horário e σ, σ_s controlam a largura na direção x e y da função (Algoritmo 1).

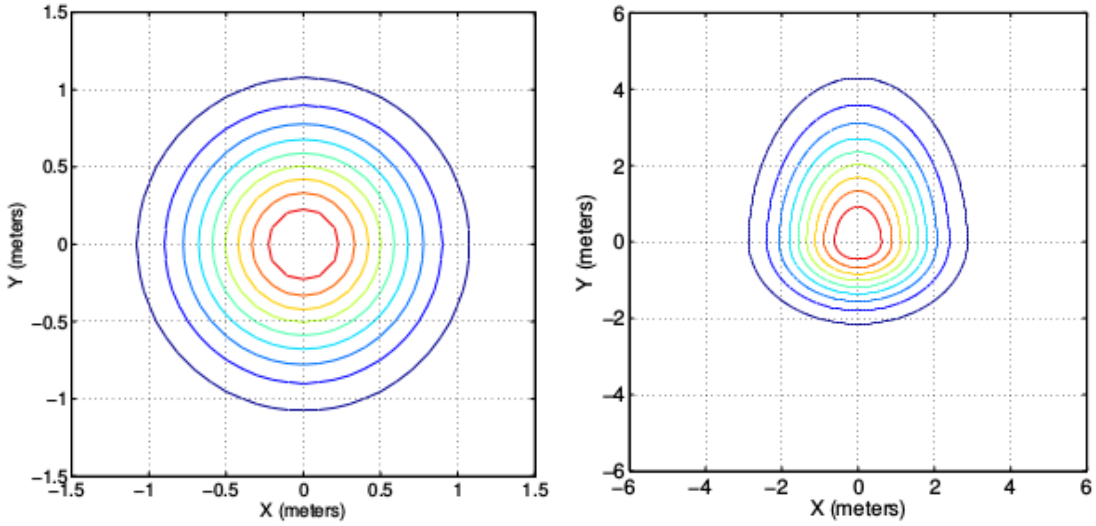
Algoritmo 1 Algoritmo para calcular o valor em (x, y) de uma função Gaussiana assimétrica centrada em (x_c, y_c), com uma rotação de θ e desvio padrão de σ_h , σ_s , e σ_r [34].

$\alpha \leftarrow$atan2$(y - y_c, x - x_c) - \theta + \pi/2$;
Normalize α;
$\sigma \leftarrow (\alpha \leq 0\ ?\sigma_r : \sigma_h)$;
$a \leftarrow (cos\,\theta)^2/(2\sigma^2) + (sin\,\theta)^2/(2\sigma_s^2)$;
$b \leftarrow sin(2\theta)/(4\sigma^2) - sin(2\theta)/(4\sigma_s^2)$;
$c \leftarrow (sin\,\theta)^2/(2\sigma^2) + (cos\,\theta)^2/(2\sigma_s^2)$;
return $exp(-(a(x - x_c)^2 + 2b(x - x_c)(y - y_c) + c(y - y_c)^2))$;

Estudos na literatura mencionam que um espaço pessoal individual é geralmente maior na direção da orientação do rosto, ou seja, as pessoas são mais sensíveis ao seu espaço frontal do que nas regiões laterais e da retaguarda [37, 49, 67]. O modelo proposto em Kirby [34] é limitado para representar as mudanças no formato do espaço social considerando como único fator a velocidade de deslocamento do indivíduo (Figura 3.1). Neste trabalho, aproveitamos a modelagem do espaço pessoal individual como uma *Asymmetric Gaussian Function* (*Função Gaussiana Assimétrica*, AGF). Entretanto, usamos a AGF considerando uma velocidade constante de deslocamento $\nu = 1.0$m/s (Figura 3.1b), mesmo para pessoas estáticas, para obter um espaço pessoal maior na área da frente. É importante ressaltar que essa função de custo foi projetada para corresponder aproximadamente ao espaço pessoal mantido nos Estados Unidos, estes valores podem sofrer alterações se considerado para diferentes culturas e grupos de familiaridade.

Uma vez que o espaço social definido pela Gaussiana é assimétrico, a distância interpessoal a qual iremos estender de forma a representar as diferentes zonas sociais, deve ser definida segundo o eixo da Gaussiana de maior extensão, cujo caso o eixo está alinhado em direção ao tronco do sujeito (orientação θ). Nesta direção, o desvio padrão da função Gaussiana é definido como

$$\sigma_h = \max(2\nu, \frac{1}{2}). \tag{3.4}$$



(a) Mapa de contorno para $\nu = 0$ m/s. (b) Mapa de contorno para $\nu = 1.0$ m/s.

Figura 3.1: Representação da mudança no formato do espaço social segundo a velocidade de deslocamento do indivíduo [34].

O desvio padrão para os lados (direção $\theta \pm \frac{\pi}{2}$) é dado por:

$$\sigma_s = \frac{2}{3}\sigma_h, \tag{3.5}$$

e para a retaguarda por

$$\sigma_r = \frac{1}{2}\sigma_h. \tag{3.6}$$

Logo, podemos representar formalmente o espaço pessoal como

$$\mathcal{G}(\cdot) = \text{AGF}(x, y, \theta, \sigma_s, \sigma_r, \sigma_h). \tag{3.7}$$

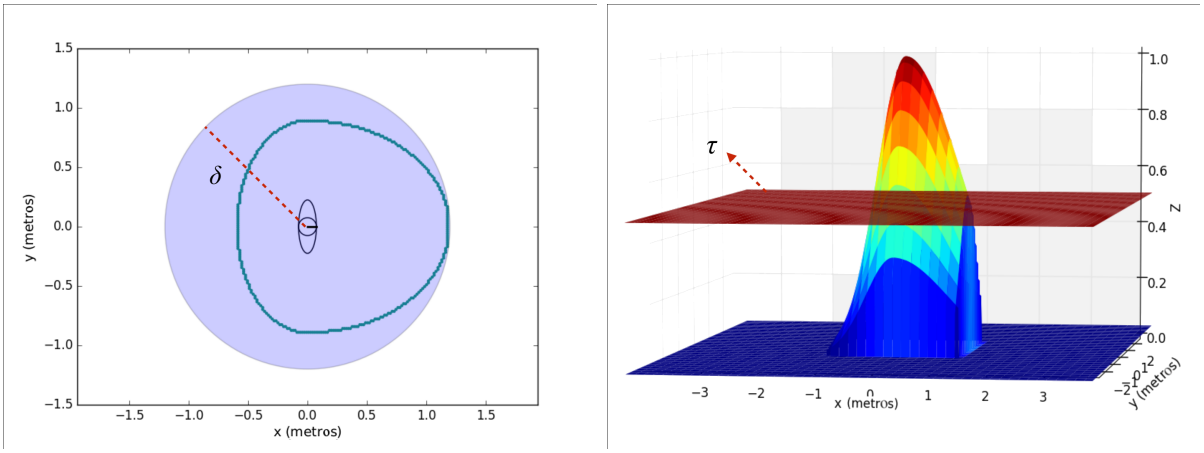
Além disso, propõe-se uma reformulação para incluir as distâncias interpessoais de proxemics [27], visando assim não mais modificar o *formato*, mas sim o *tamanho* do espaço social. Consideremos então um limiar τ que será responsável por particionar a função assimétrica supracitada em quatro zonas distintas:

1. Distância íntima $[0 - 0,45]$ m;
2. Distância pessoal $[0,45 - 1,20]$ m;
3. Distância social $[1,2 - 3,6]$ m;
4. Distância pública $[> 3,6]$ m.

Para este caso específico, uma função Gaussiana será segmentada no seu ponto mais amplo. Assim, encontramos o valor de interseção entre uma Gaussiana simétrica unidimensional e a AGF para determinar o ponto de corte desejado. Isto é realizado

substituindo a notação Gaussiana padrão $(x - \mu)$ pela distância interpessoal desejada representada por δ e o σ por σ_h, o desvio padrão da AGF em seu eixo de maior amplitude (frente). O limiar τ (Figura 3.2b) que segmentaria a função de custo de acordo com uma das zonas interpessoais δ previamente definidas (Figura 3.2a) será dada por

$$\tau = e^{-\frac{\delta^2}{2\sigma_h^2}} . \tag{3.8}$$



(a) Representação do limiar δ.

(b) Representação do limiar τ.

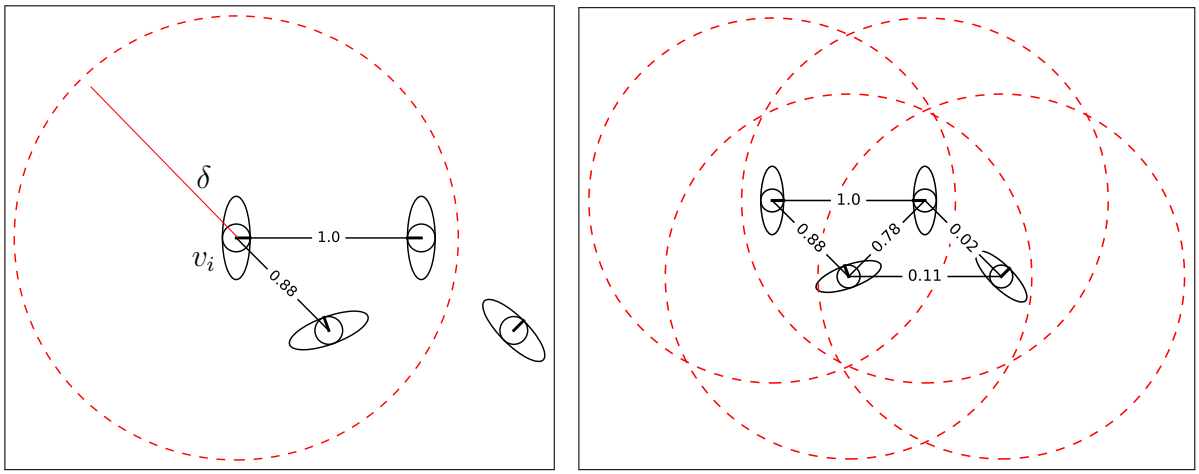
Figura 3.2: Definição dos limiares que modificarão o formato do espaço interpessoal segundo a zona social desejada (e.g. Íntima, Pessoal, Social ou Pública).

3.1.1 Determinação de Interação entre Indivíduos

A detecção de possíveis interações entre os indivíduos no ambiente é modelada seguindo uma abordagem baseada em grafos, construído sobre as regras de proxemics e estimativas de posição e orientação do corpo. Assuma que $G = \langle V, E \rangle$ seja um grafo não-dirigido, com vértices $v_i \in V$ o conjunto de pessoas a serem segmentadas como um grupo de interação, e arestas $(v_i, v_j) \in E$ correspondendo a pares de pessoas vizinhas interagindo na mesma zona de proxemics (de acordo com o limiar δ).

Conforme demonstrado na Figura 3.3, o conjunto de arestas E é inicialmente construído conectando um vértice semente inicial v_i com todos os indivíduos vizinhos pertencentes a mesma zona de proxemics δ (linhas vermelhas tracejadas). Todos os indivíduos localizados dentro da mesma zona social formam um grupo de interação em potencial. Posteriormente, cada vértice candidato $(v_i, v_j) \in E$ recebe um peso correspondente $w((v_i, v_j))$ (Equação 3.9), sendo uma medida não-negativa que representa o LoI entre os elementos

vizinhos v_i e v_j (Figura 3.3a). Este procedimento ocorre sucessivamente até que cada vértice em V seja analisado, gerando o resultado final mostrado na Figura 3.3b.



(a) (b)

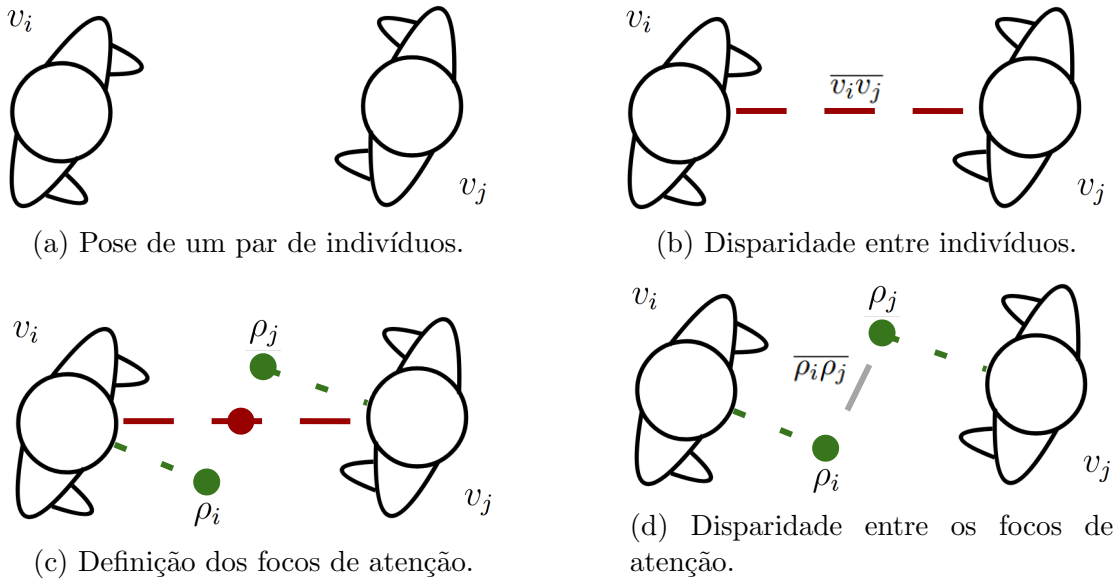
Figura 3.3: Determinação do grupo. (a) O círculo tracejado vermelho representa a região de busca de um indivíduo v_i dada uma distância social δ. O LoI para os indivíduos candidatos a estabelecer uma interação social é calculado de acordo com a Equação 3.9. (b) Grafo de interação completo para todos os indivíduos na cena.

Como mencionado, atribuímos para cada aresta uma função de peso para estimar o LoI entre indivíduos vizinhos v_i e v_j. Da mesma forma que a abordagem realizada por Cristani et al. [14], esta função baseia-se no foco de atenção estabelecido para cada par de pessoas de acordo com a orientação do corpo. No entanto, na sua abordagem, várias poses são amostradas a partir de distribuições Gaussianas para lidar com a incerteza de cenários reais. À medida que nossa metodologia visa aplicações voltadas para a interação humano-robô, optamos por uma abordagem mais direta e determinística, onde desconsideramos o processo de reamostragem das poses dos indivíduos.

Dado como entrada a posição e orientação de cada pessoa, podemos expressar a relação entre todos os pares de nós de pessoas na cena por uma função ponderada normalizada entre 0 e 1. Especificamente, esse intervalo representa a distância de disparidade no foco de atenção entre um par de candidatos para estabelecer uma interação social (Figura 3.4). O peso de uma aresta é então dado por

$$w((v_i, v_j)) = \left(1 - \frac{\|\rho_i - \rho_j\|}{2\,\|v_i - v_j\|}\right)^k, \tag{3.9}$$

onde, $\|\cdot\|$ indica a norma euclidiana, v_i e v_j representam a posição de um par de indivíduos que compartilham o mesmo espaço pessoal, onde o centro de cada pessoa é utilizada como coordenada de referência de forma que a distância euclidiana entre dois indivíduos nunca será igual a 0. Os parâmetros ρ_i e ρ_j são definidos como a posição respectiva do foco de atenção atribuído a cada indivíduo, e o k é um termo atribuído inicialmente ao valor 1. O termo de regularização k é definido empiricamente como mostrado no Algoritmo 2.



(a) Pose de um par de indivíduos.

(b) Disparidade entre indivíduos.

(c) Definição dos focos de atenção.

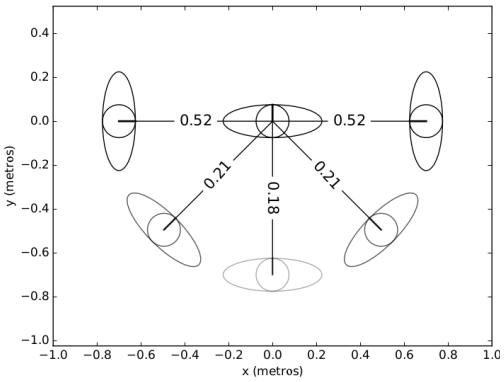
(d) Disparidade entre os focos de atenção.

Figura 3.4: Ilustração da distância de disparidade no foco de atenção entre um par de indivíduos v_i e v_j.

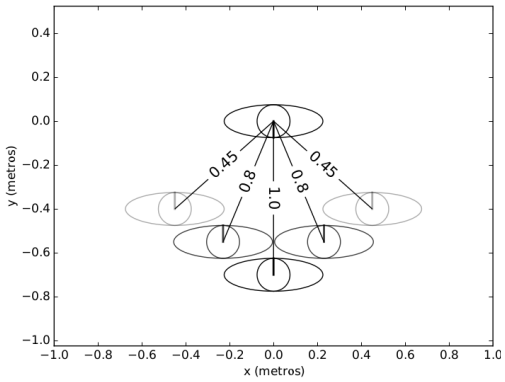
Algoritmo 2 Interação(v_i, v_j)

1: $k \leftarrow 1.0$;
2: **if** $\overline{v_i v_j} \cap \overline{\rho_i \rho_j}$ **then**
3: $\quad k \leftarrow 2.0$;
4: **end if**
5: **if** $\|\rho_i - \rho_j\| \geq \min(\|v_i - \rho_j\|, \|v_j - \rho_i\|)$ **then**
6: $\quad k = k + 0.5$;
7: **end if**
8: **return** $w((v_i, v_j))$ ▷ Equação (3.9)

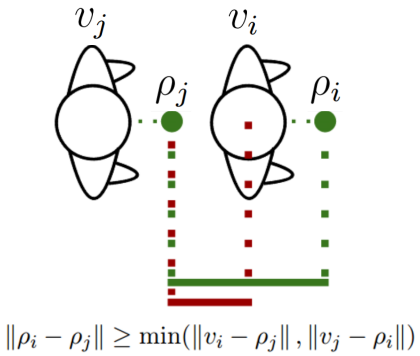
Inicialmente, o ponto central entre (v_i, v_j) é definido como ponto referencial. Em seguida, este centroide é transladado em direção ao eixo de orientação correspondente a cada um dos indivíduos, onde ρ_i and ρ_j são respectivamente definidos como o principal centro de foco. Posteriormente, a distância absoluta entre os focos de atenção é calculada. A adição de duas restrições torna-se necessária para trazer robustez ao modelo, para os casos em que um elemento não tenha acesso direto ao foco de atenção do outro (Figura 3.5a), e para lidar com configurações convergentes como em uma F-formation face-a-face (Figura 3.5b). A primeira restrição, onde nossa variável de restrição assume o valor de $k = k + 0.5$, visa garantir que o LoI gradualmente diminua sua relevância sempre que a distância euclidiana $\|\rho_i - \rho_j\|$ atingir um valor maior do que o $\min(\|v_i - \rho_j\|, \|v_j - \rho_i\|)$ (Figura 3.5c). A segunda restrição, onde nossa variável de restrição assume o valor de $k = 2.0$, produz o mesmo efeito de diminuir o LoI sempre que a linha produzida pelos pontos ρ_i e ρ_j interceptar a linha produzida pelos pontos v_i e v_j (Figura 3.5d). Os valores de atribuição para a variável de restrição k foram obtidos experimentalmente, utilizando como critério de decisão a obtenção de pesos do LoI inferiores a 0.5 para as configurações definidas como inválidas segundo as condicionais supracitadas.



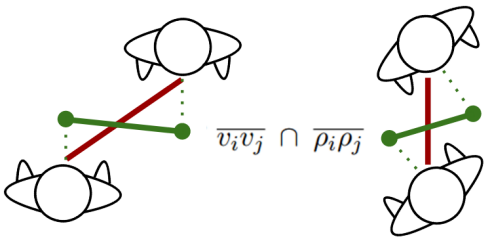
(a) Exemplo de oclusão entre focos de atenção.



(b) Exemplo de dispersão do focos de atenção.



(c) Exemplo de oclusão entre focos de atenção.



(d) Exemplo de dispersão do focos de atenção.

Figura 3.5: Exemplificação das circunstâncias onde se faz necessário a adição de restrições ao modelo matemático proposto.

Depois de calcular todos os pesos das arestas pertencentes ao grafo, um corte nas arestas do grafo será executado dado um certo limiar τ_w. Como resultado final, serão obtidas todas as pessoas vizinhas que estiverem interagindo na mesma zona de proxemics através dos subgrafos de G. Nosso método demonstrou resultados satisfatórios, conseguindo identificar as principais configurações de F-formation. Com apenas uma função de custo e duas restrições, somos capazes de considerar formações lado-a-lado, face-a-face, canto-a-canto, entre outros (e.g. N, H, V, L, C e I), ilustrado na Figura 3.6.

3.2 Espaço Social Coletivo

Ao considerarmos a presença de pessoas em um determinado ambiente, estejam essas organizadas individualmente ou em grupos, devemos determinar os espaços sociais para cada uma dessas entidades. Neste trabalho, uma vez que o espaço pessoal é representado por uma AGF, usaremos uma função de densidade para obter suas fronteiras e uma estratégia de Mistura de Gaussiana para determinar o espaço social combinado de

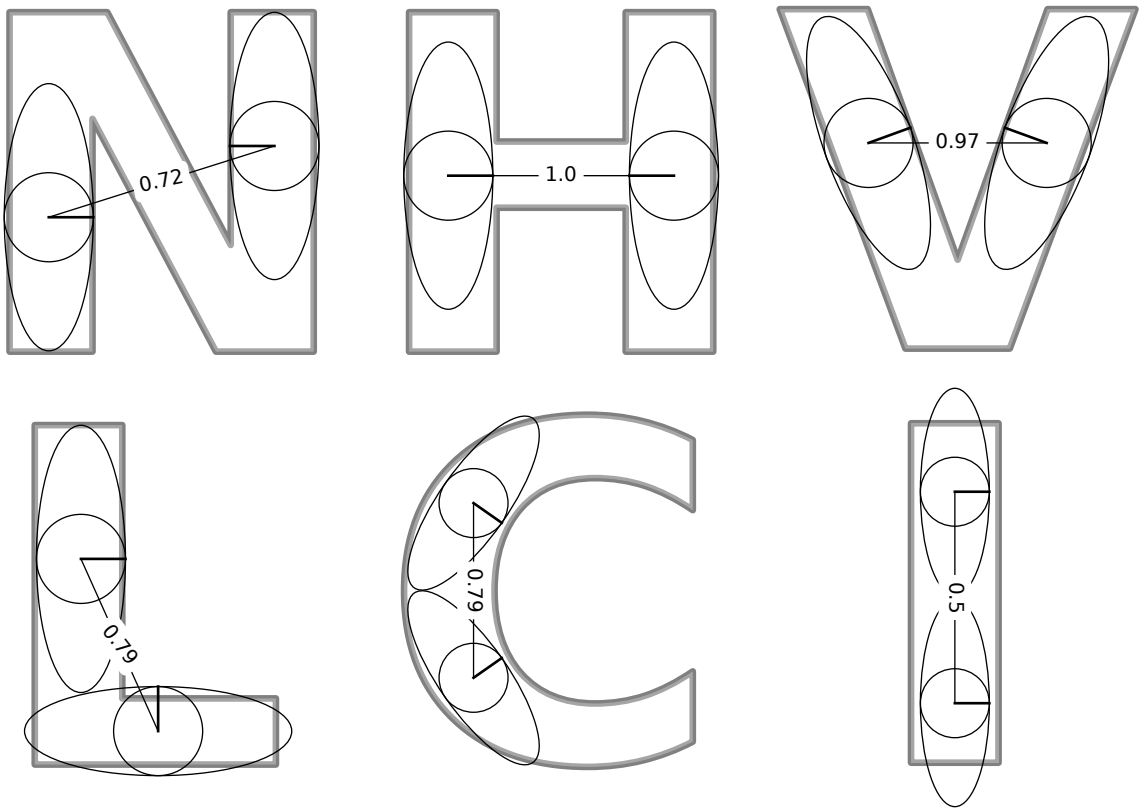


Figura 3.6: Exemplo de valores de peso para o nível de interação de acordo com a Equação 3.9 considerando diferentes configurações de F-formation.

um grupo. Isto é realizado usando um método inspirado em [70].

Inicialmente, calculamos a AGF para cada vértice nos subgrafos G' pertencente a G, onde os subgrafos correspondem aos grupos de interação detectados em um conjunto de indivíduos. Em um vértice *singleton*, caso no qual apenas um único indivíduo é considerado, a AGF é gerada apenas nos arredores de cada vértice, dentro dos limites correspondentes à distância interpessoal selecionada. Para os casos em que o subgrafo não é formado por um vértice *singleton*, a AGF é gerada na área correspondente a

$$\max(\delta, \iota) \tag{3.10}$$

com

$$\iota = \max_{(v_i, v_j) \in E(G')} \|v_i - v_j\| \,, \tag{3.11}$$

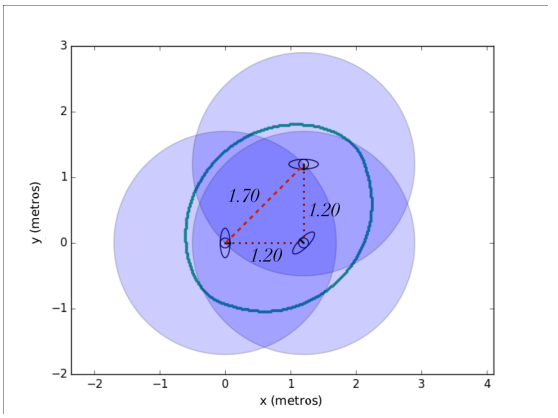
onde $E(G')$ denota todos os pares de arestas pertencentes ao subgrafo G', ou seja, um grupo de indivíduos interagindo. Em seguida, é preciso definir o limite do espaço social coletivo. Isto é, identificar os pesos que correspondem a mistura de Gaussianas expressa como

$$\Gamma = \sum_{i}^{V(G')} \phi_i \mathcal{G}_i(\cdot) \,, \tag{3.12}$$

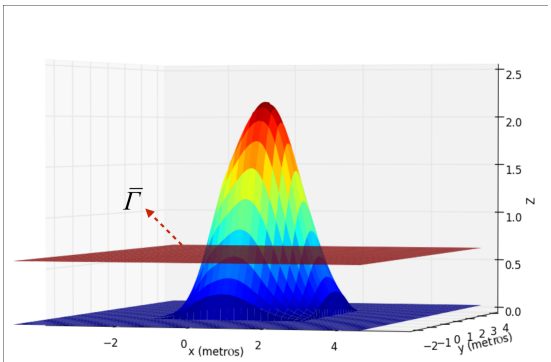
onde $V(G')$ denota todos os vértices pertencentes ao subgrafo G', equivalente ao número de funções Gaussianas que estão sendo misturadas, e ϕ_i representa o peso da mistura da i^{th}

AGF. Assim, uma vez que a função Gaussiana é definida apenas no espaço correspondente à zona social de todos os indivíduos pertencentes ao grupo, calculamos o peso médio $\bar{\Gamma}$ que corresponderá ao limite que inclui a maioria dos indivíduos pertencentes ao grupo, para ser usado na determinação do limiar h (Equação 3.13).

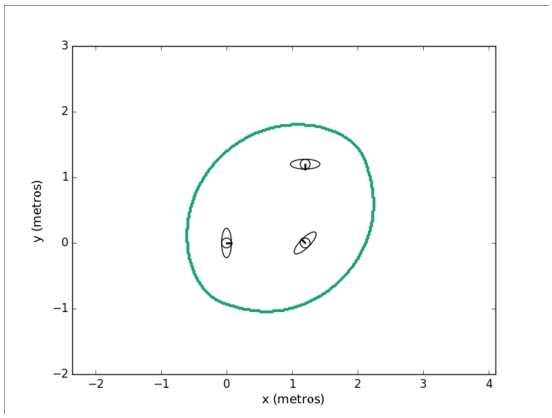
A Figura 3.7 exemplifica esta etapa apresentando como instância a determinação do espaço social coletivo para três indivíduos que estão localizados no mesmo ambiente. Nesse exemplo, buscamos identificar os grupos que estejam compartilhando a mesma zona de interação Pessoal $\delta = 1.2$m. A Equação 3.10 é utilizada para determinar a área na qual a AGF será computada, sendo o resultado da expressão $\max(1.2, 1.7)$ ilustrado pela região sombreada na Figura 3.7a. Para cada indivíduo pertencente ao grupo, uma AGF será computada respeitando a região delimitante. Posteriormente, é realizado o processo de mistura das Gaussianas (Equação 3.12), onde o peso médio $\bar{\Gamma}$ é então utilizado como limiar para definir o ponto de corte (Figura 3.7b). Dessa forma, podemos obter os pontos de borda referentes ao espaço social coletivo (Figura 3.7c).



(a) Região para cálculo da AGF.



(b) Corte da mistura de Gaussianas.



(c) Delimitação do espaço social coletivo.

Figura 3.7: Exemplo das etapas de definição do espaço social coletivo para $\delta = 1.2$m.

Finalmente, podemos formalizar a escolha do limite responsável pela segmentação de pessoas de acordo com seu espaço social da seguinte forma:

$$h = \begin{cases} \tau & \text{subgrafo } \textit{singleton} \\ \bar{\Gamma} & \text{caso contrario} \end{cases} \tag{3.13}$$

Com as restrições impostas a h, garantimos que o espaço social atômico alinhado ao torso do indivíduo nunca estará a uma distância menor do que a distância interpessoal selecionada. E para o espaço social coletivo, devido à mistura de Gaussianas, o limite selecionado será o mais próximo possível do indivíduo, mas nunca menos do que a distância pessoal.

Capítulo 4

Experimentos

O objetivo dessa seção consiste em expor as vantagens das duas principais contribuições desse trabalho: a abordagem baseada em grafos para identificar possíveis associações de indivíduos no ambiente, onde definimos matematicamente um modelo de interação de acordo com a posição e orientação corporal entre os pares de indivíduos, sendo possível segregar diferentes grupos de interação de acordo com as zonas sociais especificadas nas regras de proxemics e a representação do espaço social coletivo a partir da detecção dos grupos de interação, onde a determinação de tais limites sociais oferece a possibilidade de tratar os grupos como entidades únicas, delimitando o acesso indevido a sua zona de conforto interpessoal. Assim, espera-se que a metodologia proposta consiga identificar com uma maior acurácia grupos de interação, estimada através da média harmônica entre *precision* e *recall*, e que as restrições de navegação impostas sobre os indivíduos gerem comportamentos mais sociais do que métodos que não consideram tais restrições.

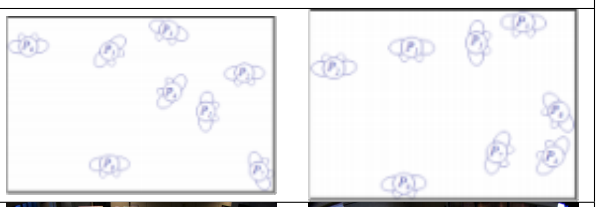
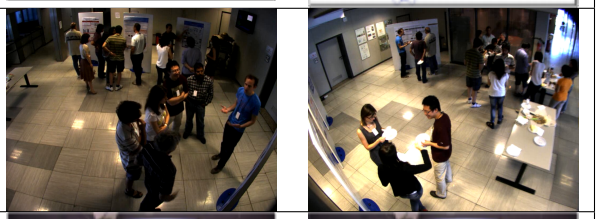
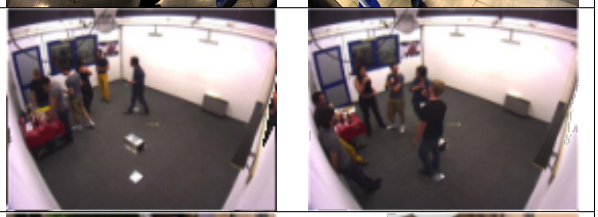
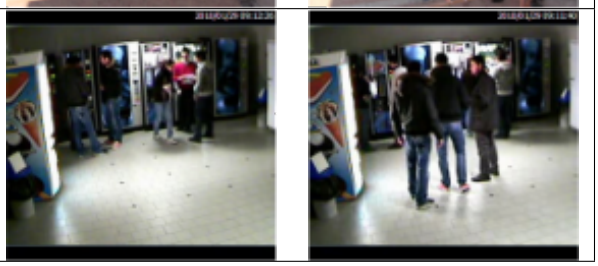
Inicialmente, avaliamos como nosso método executa a tarefa de detectar interações entre indivíduos, onde comparamos o desempenho do algoritmo proposto em relação às cinco abordagens no estado da arte focadas na identificação de grupos de interação. Visando uma comparação mais completa, os algoritmos foram avaliados em cinco conjuntos de *datasets* publicamente disponíveis, que apresentam dados sintéticos e reais incluindo diferentes tarefas comportamentais com anotações da pose e relação de interação entre os indivíduos, cujas principais características estão resumidas na Tabela 4.1 [56]. Na Tabela 4.2 podem ser observados exemplos visuais correspondentes a cada um dos *dataset*. Na parte preliminar, descrevemos os *datasets* empregados como *benchmark*. Em seguida, apresentamos a capacidade da nossa abordagem para agrupar pessoas de acordo com o nível de interação entre os indivíduos e simultaneamente definir uma zona de conforto considerando o espaço interpessoal. Analisamos os pontos fortes e fracos da nossa abordagem para cada cenário testado.

4.1 Datasets

Tabela 4.1: Principais características dos *datasets* [56].

Dataset	Tipo de dados	Detecção	Qualidade de detecção
Synthetic	sintético	–	perfeita
SALSA	real	manual/semi-automático	alta
Cocktail Party	real	automático	alta
Coffee Break	real	automático	baixa
GDet	real	automático	muito baixa

Tabela 4.2: Imagens correspondentes aos *datasets* utilizados.

Dataset	**Frames de amostra**
Synthetic	
SALSA	
Cocktail Party	
Coffee Break	
GDet	

Synthetic[1] [14]: Um conjunto gerado artificialmente de 10 situações diferentes, cada uma sendo repetida 10 vezes, resultando em um total de 100 *frames*. Os dados foram anotados para obter o *ground truth* dos F-formations, onde as posições e orientações são livres de ruído.

Synergetic sociAL Scene Analysis (SALSA)[2] [2]: Este é um conjunto de dados mais recente, composto por duas gravações de um evento social (30 minutos cada) coletadas em um espaço interno regular envolvendo 18 participantes. A primeira parte consistiu em uma sessão de apresentação de pôsteres, onde quatro estudantes de pós-graduação apresentaram quatro estudos de pesquisa. Uma quinta pessoa presidiu a sessão de pôster. Na segunda, todos os participantes foram autorizados a interagir livremente sobre alimentos e bebidas durante um coquetel. O SALSA também fornece as anotações de *ground truth*, que foram realizadas manualmente ou semi-automaticamente durante toda a duração do evento. As anotações foram produzidas em duas etapas. No primeiro passo, usando uma ferramenta dedicada de anotação de cena multi-vista, a posição, a orientação da cabeça e do corpo de cada alvo foi anotado a cada 45 *frames*. No segundo passo, as posições anotadas e as orientações da cabeça/corpo foram usadas para deduzir F-formations. As anotações foram novamente realizadas a cada 45 *frames*.

Cocktail Party[3] [75]: Este *dataset* contém cerca de 30 minutos de gravações de vídeo de um coquetel em um ambiente de laboratório de 30 m^2 envolvendo 7 indivíduos. O *dataset* foi gravado utilizando quatro câmeras (15Hz, 1024 $\times$ 768) sincronizadas e posicionadas nos cantos da sala. Grupos em um *frame* a cada 5 segundos foram anotados manualmente por um especialista, resultando em um total de 320 *frames* rotulados para avaliação. Neste conjunto de dados, a informação de proxemics é estimada automaticamente, portanto, erros podem estar presentes.

Coffee Break[4] [14]: O *dataset* apresenta um cenário de *coffee-break* em um evento social realizado ao ar livre, com um máximo de 14 indivíduos organizados em grupos de 2 ou 3 pessoas cada. As imagens são registradas a partir de uma única câmera com resolução de 1440 $\times$ 1080. Considerando os dados de *ground truth*, um psicólogo anotou os vídeos que indicam os grupos presentes nas cenas, para um total de 119 quadros divididos em duas sequências. Todas as anotações de proxemics provêm de ferramentas automáticas de visão computacional.

[1] *Dataset* disponível em http://profs.sci.univr.it/~cristanm/ssp/.
[2] *Dataset* disponível em http://tev.fbk.eu/salsa.
[3] *Dataset* disponível em http://tev.fbk.eu/resources.
[4] *Dataset* disponível em http://profs.sci.univr.it/~cristanm/ssp/.

GDet[5] [5]: O *dataset* é composto por 5 subsequências de imagens adquiridas por 2 câmeras de baixa resolução (352 × 328), com uma série de *frames* que variam de 17 a 132, para um total de 403 *frames* anotados. Os vídeos foram adquiridos com duas câmeras monoculares, localizadas em ângulos opostos da sala. As anotações de proxemics seguem o mesmo protocolo que no *dataset* Coffee Break.

4.2 Determinação do Grupo

A metodologia proposta foi comparada com cinco abordagens no estado-da-arte:

1. Hough Voting (HVFF-lin) [14];
2. Sua variante não-linear (HVFF-ent) [54];
3. Sua extensão multi-escala (HVFF-ms) [55];
4. A abordagem *game-theoretic* (CTCG) [66];
5. E a abordagem do corte de grafo (GCFF) [56].

Como medidas de precisão, adotamos as métricas propostas em Cristani et al. [14] e estendida em Setti et al. [54]. Nesse sentido, um grupo é considerado corretamente estimado se pelo menos $\lceil (T \cdot |G|) \rceil$ dos seus membros são detectados corretamente pelo algoritmo e se não mais do que $\lceil (1 - T) \cdot |G| \rceil$ falsos sujeitos são identificados, onde $|G|$ é a cardinalidade do grupo rotulado G, e $T \in]0, 1]$ é um limiar arbitrário, chamado limiar de tolerância. Em particular, nos concentramos em dois valores, $T = 1$ (Tabela 4.3) e $T = 2/3$ (Tabela 4.4). As taxas de *False Positive* (*Falso Positivo*, FP) e *False Negative* (*Falso Negativo*, FN) são derivadas subtraindo o *True Positive* (*Verdadeiro Positivo*, TP) da cardinalidade dos grupos detectados e dos grupos de *ground truth*, respectivamente. Com base nessas métricas, calculamos por *frame* as métricas *precision* e *recall*:

$$precision = \frac{TP}{TP + FP} \ , \ recall = \frac{TP}{TP + FN} \ , \tag{4.1}$$

e o F_1-*score*, definido como a média harmônica do *precision* e *recall*:

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \ , \tag{4.2}$$

no qual a média desses valores sobre todos os *frames* expressa o resultado final obtido.

[5] *Dataset* disponível em http://profs.sci.univr.it/~cristanm/ssp/.

Neste caso, a métrica precision mede a porcentagem dos indivíduos estimados que foram classificados corretamente como um grupo de interação. Em contrapartida, o recall informa qual a porcentagem dos indivíduos que deveriam ter sido estimados foram realmente selecionados. Os problemas referentes a detecção visual não fazem parte do escopo deste trabalho. Nesse sentido, utilizamos as anotações de *ground truth* da posição e orientação do corpo dos indivíduos considerando apenas os quadros anotados. Empiricamente adotamos como parâmetro para nossa abordagem um limiar $\tau_w = 0.5$ para cortar o grafo. Nestes experimentos consideramos uma distância social fixa de $\delta = 1.20$m, limiar definido por Hall [27] como distância de transição da zona Pessoal para Social, onde partimos da premissa que a maior parte dos *datasets* possuam interações sendo formadas nessas zonas sociais.

A Tabela 4.3 apresenta os resultados com $T = 1$, que considera um grupo como correto se e somente se todos os seus membros forem detectados corretamente. As comparações para o *dataset* SALSA foram realizadas apenas considerando o nosso método, já que no trabalho apresentado por Alameda-Pineda et al. [2] o limiar de tolerância $T = 1$ não foi considerado. Podemos observar que os melhores resultados foram obtidos para o *dataset* Synthetic, o qual em virtude de ter sido gerado artificialmente apresenta uma qualidade de detecção perfeita. Para os *datasets* Cocktail Party, Coffee Break e GDet, é possível constatar que obtivemos os menores níveis de desempenho nos índices de recall. Isto significa que mesmo para o melhor caso (Coffee Break), se considerarmos todos os indivíduos que deveriam ter sido estimados, apenas 50% deles foram selecionados. Podemos inferir que o baixo desempenho produzido foi decorrente da seleção de uma distância social δ inferior ao comportamento social estabelecido em cada um dos *datasets*, restringindo assim a detecção dos grupos de interação mais esparsos e consecutivamente aumentando a quantidade de falsos negativos.

A Tabela 4.4 mostra os resultados obtidos considerando o limiar $T = 2/3$, correspondente a encontrar pelo menos 2/3 dos membros de um grupo e não mais do que 1/3 de falsas interações. A acurácia dos métodos comparativos para o *dataset* Synthetic, Cocktail Party, Coffee Break e GDet foi extraída de Setti et al. [56]. Da mesma forma, os dados de precisão dos métodos selecionados para o *dataset* SALSA foram extraídos de Alameda-Pineda et al. [2]. É possível verificar que nossa metodologia conseguiu os melhores resultados para o *dataset* real SALSA e manteve um limiar de acurácia comparável em relação ao *dataset* Synthetic. Por apresentar um limiar de tolerância mais flexível, podemos notar uma melhoria da acurácia em relação aos resultados anteriormente discutidos. Entretanto, as considerações realizadas a respeito dos índices de acurácia do recall se mantém, sendo realizados novos procedimentos experimentais para apurar as condições de baixo desempenho identificadas.

Para uma melhor avaliação visual, os dados referentes ao F_1-score apresentados nas tabelas anteriores foram sintetizados na Figura 4.1, onde os *datasets* foram organizados

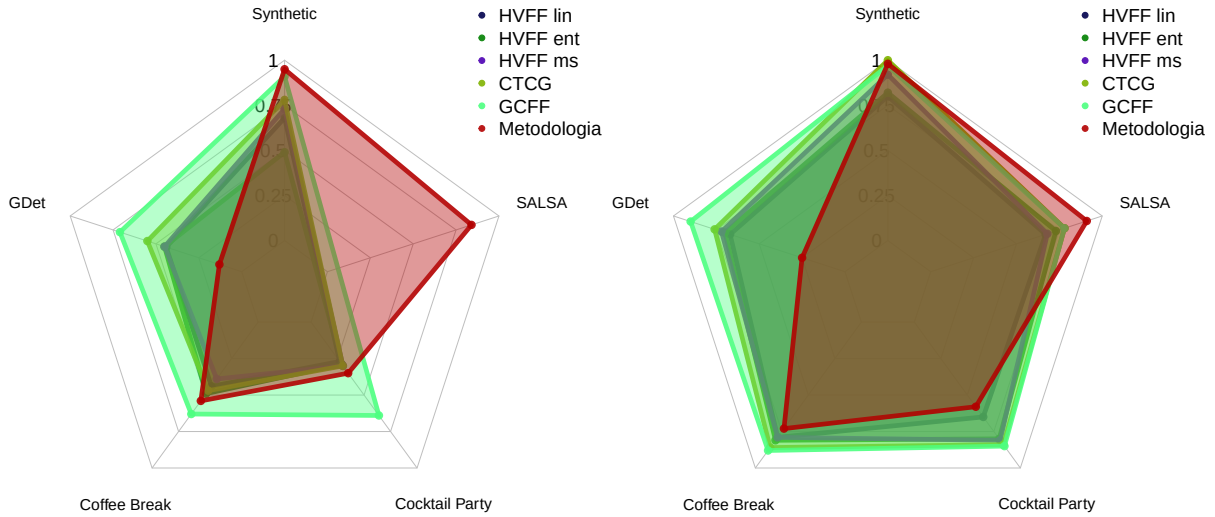
Tabela 4.3: Avaliação da determinação de grupo considerando $T = 1$ e valores de comparação retirados de Setti et al. [56]. Nestes experimentos consideramos uma distância social fixa de $\delta = 1.20$m, que se encontra na zona de proxemics Pessoal.

	Synthetic			SALSA			Cocktail Party			Coffee Break			GDet		
	Prec.	Rec.	F_1	Prec.	Rec.	F_1	Prec.	Rec.	F_1	Prec.	Rec.	F_1	Prec.	Rec.	F_1
HVFF lin [14]	0.64	0.73	0.68	-	-	-	0.26	0.27	0.27	0.41	0.47	0.44	0.43	0.45	0.44
HVFF ent [54]	0.47	0.52	0.49	-	-	-	0.28	0.30	0.29	0.47	0.52	0.49	0.44	0.45	0.45
HVFF ms[55]	0.72	0.73	0.73	-	-	-	0.30	0.30	0.30	0.40	0.38	0.39	0.44	0.45	0.45
CTCG [66]	0.78	0.78	0.78	-	-	-	0.31	0.28	0.30	0.46	0.47	0.47	0.51	0.60	0.55
GCFF [56]	0.91	0.91	0.91	-	-	-	**0.63**	**0.65**	**0.64**	0.61	**0.64**	**0.63**	**0.73**	**0.68**	**0.71**
Metodologia	**0.97**	**0.94**	**0.95**	**0.84**	**0.86**	**0.84**	0.37	0.34	0.35	**0.63**	0.50	0.55	0.14	0.13	0.13

Tabela 4.4: Avaliação da determinação de grupo considerando $T = 2/3$ e valores de comparação retirados de Setti et al. [56], Alameda-Pineda et al. [2]. Nestes experimentos consideramos uma distância social fixa de $\delta = 1.20$m, que se encontra na zona de proxemics Pessoal.

	Synthetic			SALSA			Cocktail Party			Coffee Break			GDet		
	Prec.	Rec.	F_1	Prec.	Rec.	F_1	Prec.	Rec.	F_1	Prec.	Rec.	F_1	Prec.	Rec.	F_1
HVFF lin [14]	0.75	0.86	0.80	0.59	0.74	0.67	0.59	0.74	0.65	0.73	0.86	0.79	0.66	0.68	0.67
HVFF ent [54]	0.79	0.86	0.82	0.66	0.80	0.73	0.78	0.83	0.80	0.76	0.86	0.81	0.69	0.71	0.70
HVFF ms[55]	0.90	0.94	0.92	0.61	0.76	0.68	0.81	0.81	0.81	0.83	0.76	0.79	0.71	0.73	0.72
CTCG [66]	**1.00**	**1.00**	**1.00**	**0.89**	0.70	0.78	**0.86**	0.82	0.84	0.83	0.89	0.86	0.76	0.76	0.76
GCFF [56]	0.97	0.98	0.97	0.82	0.85	0.78	0.84	**0.86**	**0.85**	**0.85**	**0.91**	**0.88**	**0.92**	**0.88**	**0.90**
Metodologia	**1.00**	0.96	0.98	**0.89**	**0.95**	**0.92**	0.59	0.59	0.58	0.82	0.67	0.73	0.26	0.24	0.25

em sentido horário segundo sua ordem decrescente de qualidade de detecção. Como pode ser observado, a metodologia proposta carece de desempenho para o limiar de tolerância $T = 1$ (Figura 4.1a) em conjuntos de dados reais, principalmente devido a inconsistências temporais na detecção automática, sendo assim capaz de obter os melhores resultados no conjunto de dados sintéticos. Para o limiar de tolerância $T = 2/3$ (Figura 4.1b), também obtivemos os melhores resultados quanto maior a qualidade de detecção do *dataset*.



(a) Limiar de tolerância $T = 1$.

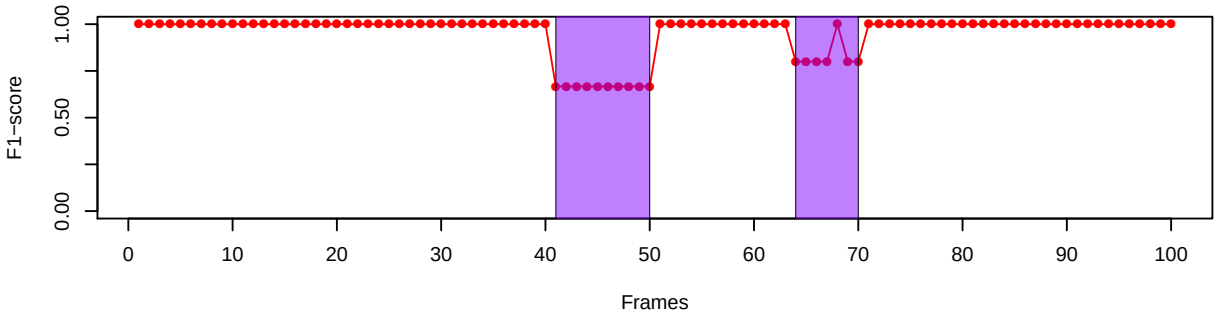
(b) Limiar de tolerância $T = 2/3$.

Figura 4.1: Gráfico de radar contendo as variáveis quantitativas referentes ao F_1-score para os *datasets* organizados em sentido horário segundo sua ordem decrescente de qualidade de detecção.

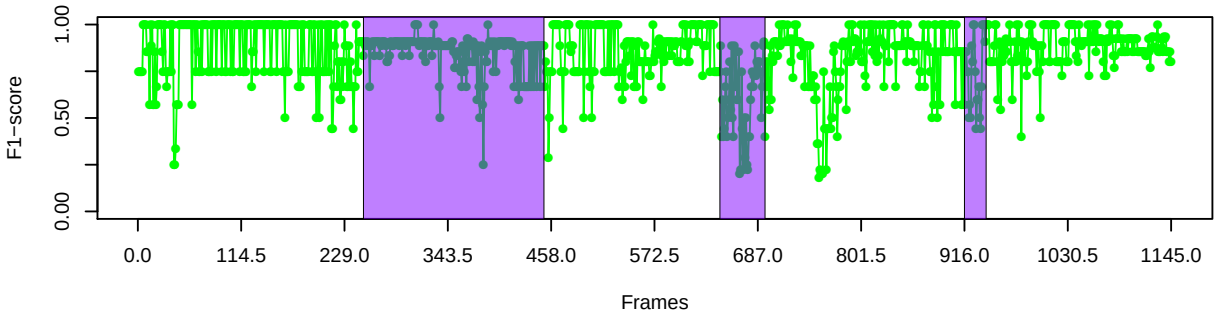
Tendo em vista identificar especificamente onde nossa metodologia não obteve resultados satisfatórios, apresentamos nas Figuras 4.2 e 4.3 a distribuição temporal dos valores de F_1-score obtidos por *frame*, onde as circunstâncias de baixo desempenho (sombreamento roxo) foram investigadas mais a fundo. Identificamos que as duas principais causas de baixo desempenho são: (i) distâncias sociais diferentes usadas para a classificação, e (ii) uma inconsistência temporal na informação de posição e orientação entre *frames*, decorrente da incerteza de detecção nos conjuntos de dados anotados que está associado com a qualidade de detecção dos *datasets*. Quanto menor for a qualidade de detecção, maior será a incerteza relativa a informação de posição e orientação dos indivíduos. É importante salientar que a detecção dos indivíduos não faz parte do escopo desse projeto, em todos os experimentos foram utilizados os dados de pose dos indivíduos fornecidos por cada *dataset*.

Para ilustrar melhor os casos de classificação errônea, usamos o código *script* de exibição disponível públicamente, fornecido por Setti et al. [56] [6]. Podemos observar que a distribuição temporal para o *dataset* Synthetic (Figura 4.3a) apresenta uma queda de

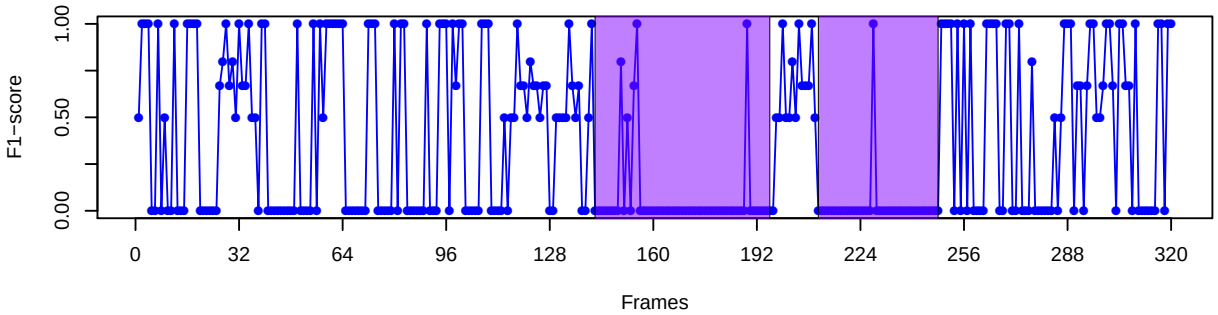
[6]Os códigos associados estão Públicamente disponíveis em http://profs.sci.univr.it/~cristanm/ssp/.



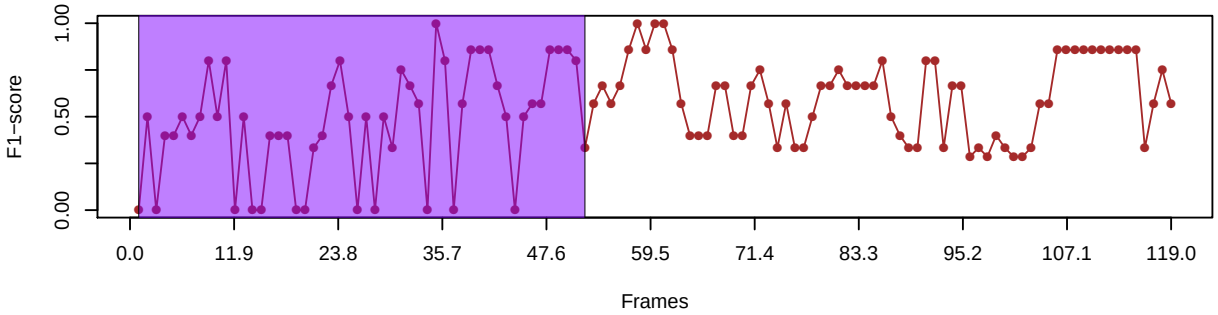
(a) Synthetic.



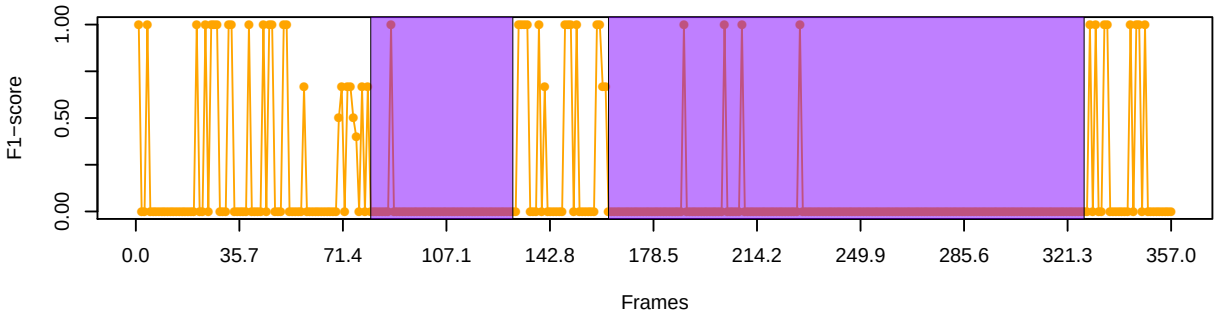
(b) SALSA.



(c) Cocktail Party.

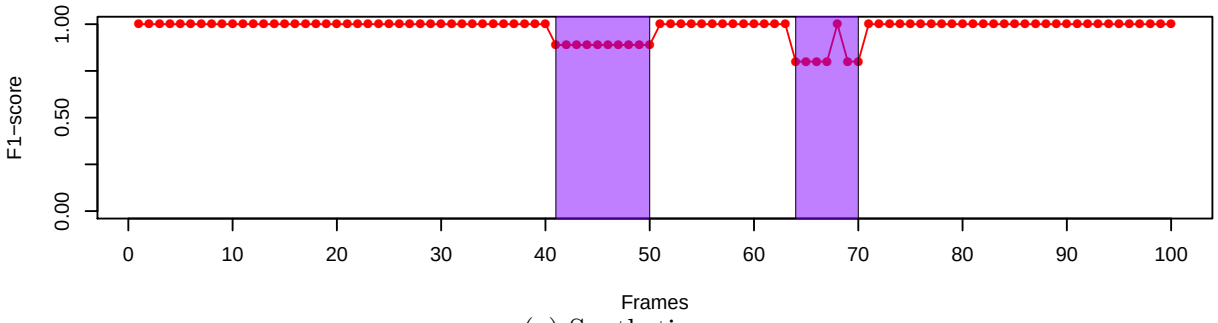


(d) Coffee Break.

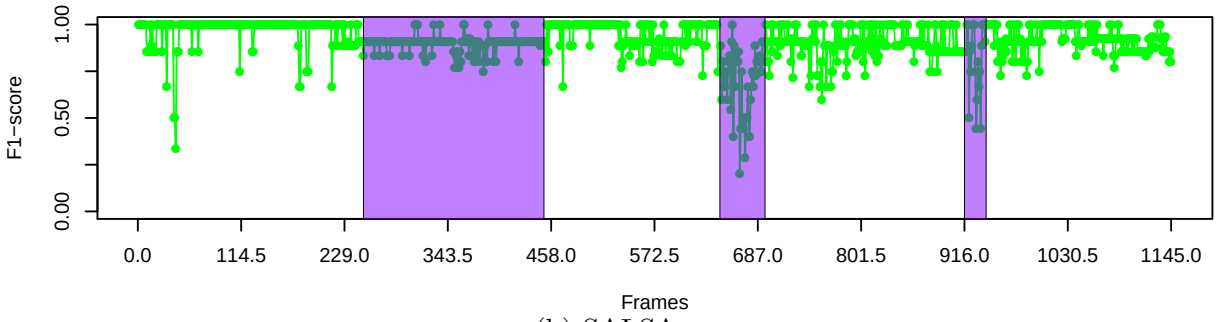


(e) GDet.

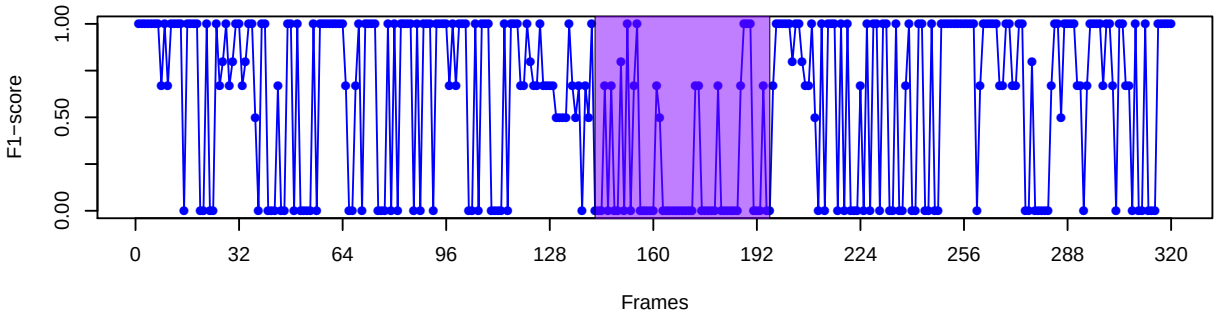
Figura 4.2: Distribuição temporal dos valores de F_1-score obtidos por *frame* para $T = 1$. As circunstâncias mais desfavorecidas são destacadas no sombreado roxo.



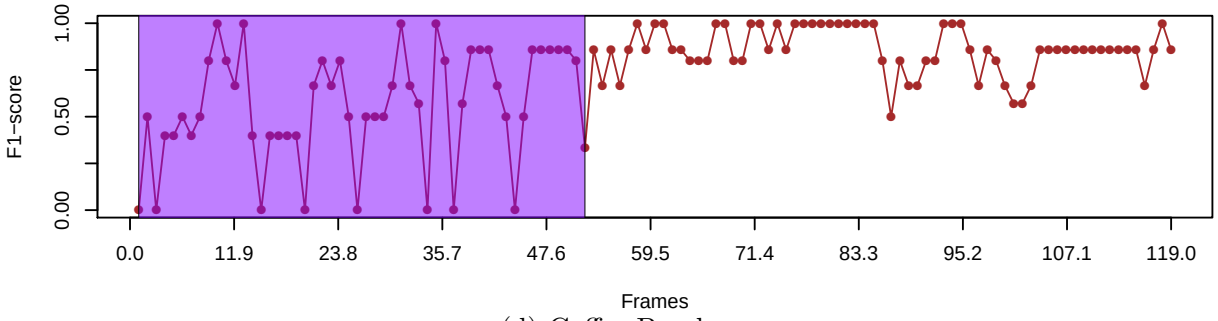
(a) Synthetic.



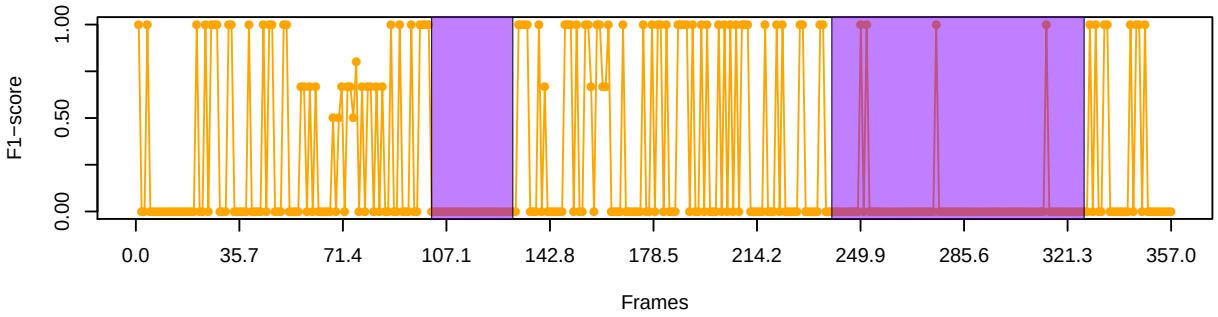
(b) SALSA.



(c) Cocktail Party.



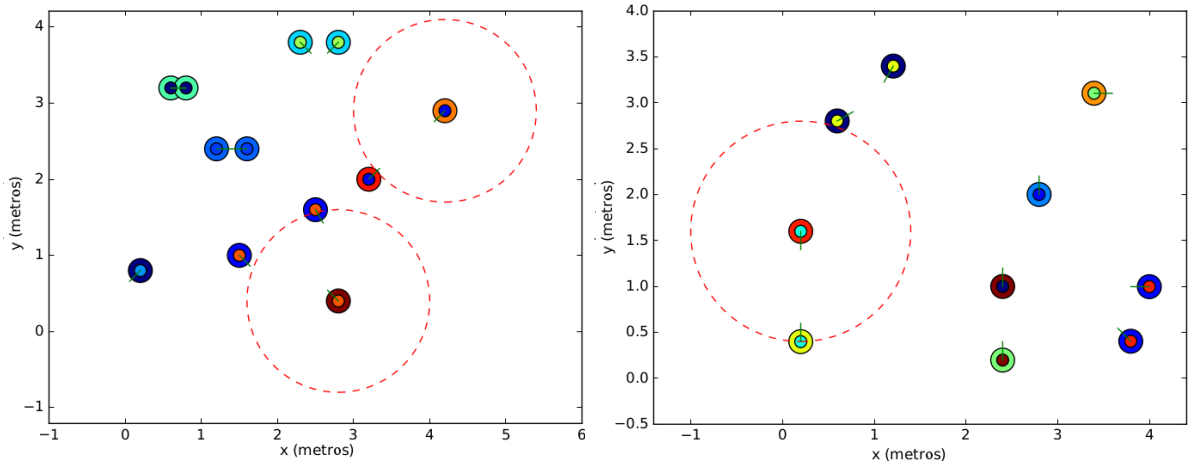
(d) Coffee Break.



(e) GDet.

Figura 4.3: Distribuição temporal dos valores de F_1-score obtidos por *frame* para $T = 2/3$. As circunstâncias mais desfavorecidas são destacadas no sombreado roxo.

desempenho em dois momentos (*frames* $40-49$ e $63-69$). Como pode ser visto na Figura 4.4, a divergência de nossa classificação de grupos de interação em relação ao *ground truth* é devido ao fato de que os indivíduos estão posicionados a uma distância maior do que o espaço pessoal de $\delta = 1.20$m (círculos tracejados vermelhos). O *frame* 67 (Figura 4.3a) parece ter um ligeiro deslocamento de posição, sendo uma exceção e agrupando os dois indivíduos na mesma zona pessoal.



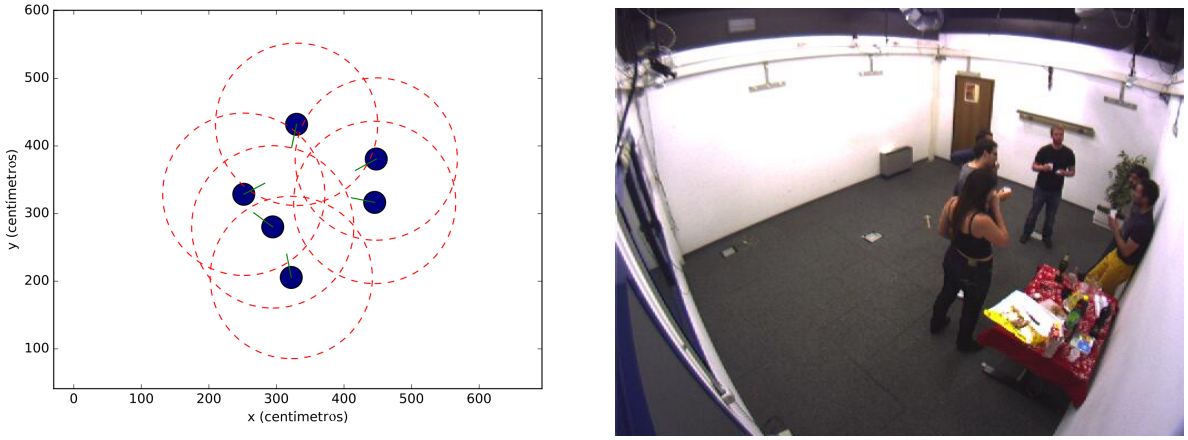
(a) *Frames* 40 a 49. (b) *Frames* 63 a 69.

Figura 4.4: *Frame* de exemplo que ilustra casos de classificação errônea de grupos interativos considerando o *dataset* sintético. Os círculos tracejados vermelhos representam os espaços pessoais dos indivíduos. A cor do círculo externo representa a nossa classificação, e a cor interna o *ground truth*.

A questão da divergência de distância social também ocorre nos conjuntos de dados do mundo real. Por exemplo, a Figura 4.5b mostra o *frame* 167 do *dataset* Cocktail Party. A classificação de *ground truth* (Figura 4.5a) classifica os seis indivíduos como parte do mesmo grupo. No entanto, de acordo com a nossa área social assumida, os mesmos não estão dentro da mesma zona social. Portanto, é fundamental definir as zonas de proxemics corretas para serem usadas em um determinado contexto.

A Figura 4.6a corresponde ao *frame* 440 no *dataset* SALSA, mais especificamente referente ao Poster Session. Observamos uma divergência na classificação de nossos grupos (círculos vermelho escuro), onde nosso método infere que todos os 5 indivíduos compartilham o mesmo foco de atenção de acordo com a configuração corporal apresentada. Por fim, a Figura 4.6b mostra uma cena pertencente à parte do vídeo Cocktail Party. Nessa situação, podemos visualizar o momento exato em que o *ground truth* apresenta uma mudança na posição e orientação dos indivíduos (isso pode ser o resultado de dados ruidosos). Como nossa abordagem não realiza uma avaliação temporal da cena, os ruídos da detecção podem influenciar diretamente no desempenho do algoritmo.

Cada um dos fatores previamente identificados como responsáveis pela redução de desempenho na identificação dos grupos de interação foram apurados individualmente. Nesse sentido, primeiramente fizemos uma avaliação aumentando gradualmente a distância



(a) Group classification. (b) *Frame* 167.

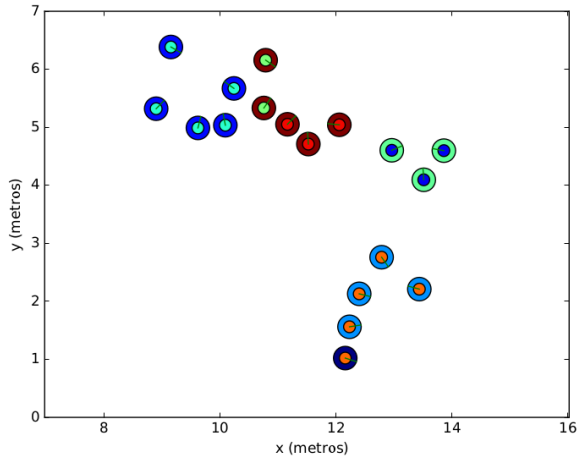
Figura 4.5: Exemplo de classificação incorreta de grupos interagindo considerando o *dataset* Cocktail Party. Os círculos azuis representam a classificação segundo o *ground truth*, e os círculos tracejados vermelhos a distância social.

social (ainda mantendo a coerência com o ambiente), o que consequentemente resulta em agregar mais indivíduos em grupos, onde a maximização da métrica de F-score foi utilizada como critério para determinação da melhor distância social. Como mostrado na Tabela 4.5, o impacto em nossos resultados é positivo. O *dataset* Cocktail Party, por exemplo, considera claramente uma zona Social (em vez de Pessoal). O impacto direto da variação da distância social sobre a acurácia das métricas utilizadas pode ser visualizada na Figura 4.7.

Tabela 4.5: Resultados considerando diferentes distâncias sociais para cada *dataset* avaliado, onde a maximização da métrica F_1-score foi utilizada como critério para determinação da melhor distância δ.

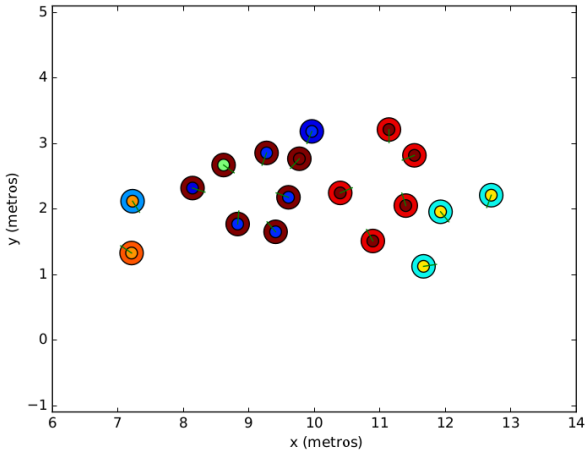
	T = 1				T = 2/3			
	Prec.	Rec.	F_1	δ	Prec.	Rec.	F_1	δ
Synthetic	**1.00**	**0.98**	**0.99**	1.25m	**1.00**	0.98	0.99	1.22m
SALSA	**0.85**	**0.92**	**0.88**	1.43m	0.88	**0.97**	**0.92**	1.33m
Cocktail Party	**0.76**	**0.77**	**0.75**	2.37m	**0.90**	**0.94**	**0.91**	2.51m
Coffee Break	**0.65**	**0.95**	**0.76**	3.27m	0.82	0.80	0.79	2.04m
GDet	0.39	0.38	0.38	3.49m	0.60	0.60	0.59	3.49m

Como principais características a serem observadas, podemos destacar o comportamento observado nas Figuras 4.7a e 4.7b, onde a partir de um determinado limiar o acréscimo da distância social implica no decréscimo da métrica Precision e consecutivamente do F_1-score. A ocorrência deste comportamento se deve ao fato de tanto o *dataset* Synthetic quanto o SALSA possuírem uma grande quantidade de grupos distintos aglomerados em regiões espacialmente circunjacentes, de forma que a utilização de uma distância social elevada resulta na mesclagem de grupos distintos.





(a) SALSA Poster Session: *Frame* 440.



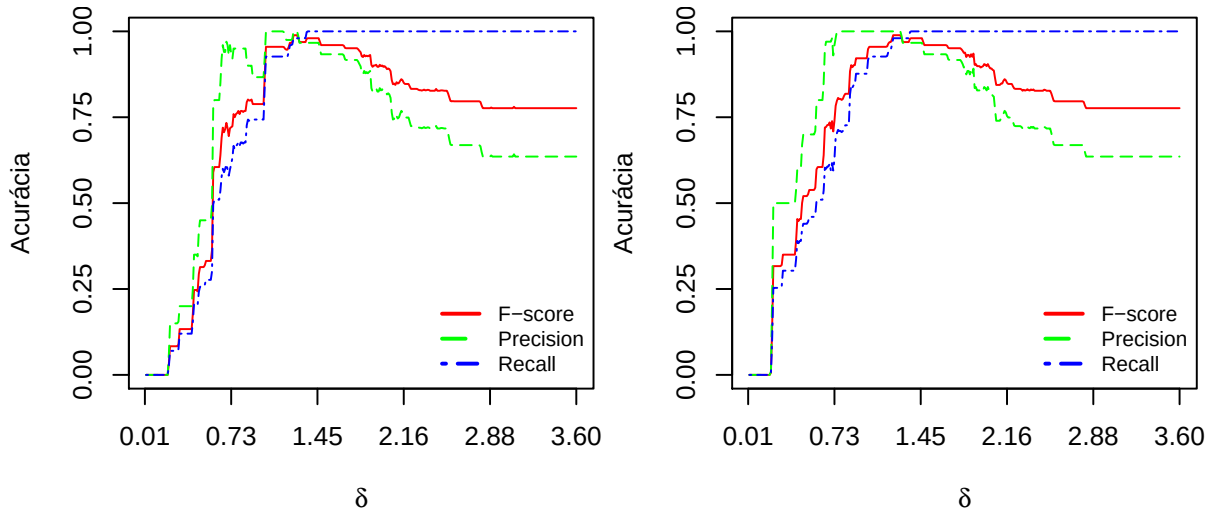


(b) SALSA Cocktail Party: *Frame* 270.

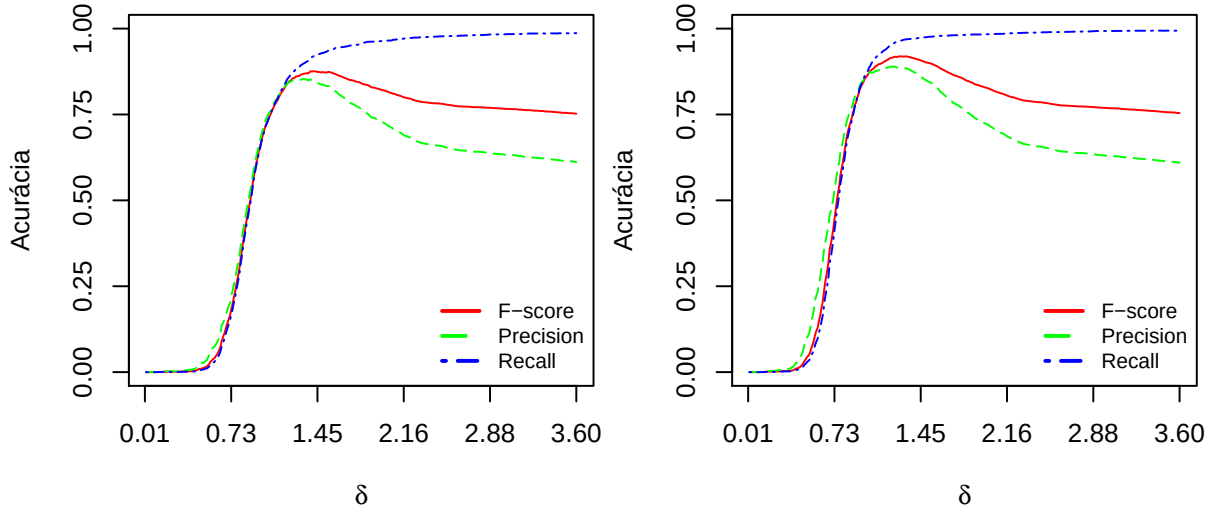
Figura 4.6: Exemplo com *frames* ilustrando casos de classificação errônea em grupos interativos considerando o *dataset* SALSA. A cor do círculo externo representa a nossa classificação, e a cor interna o *ground truth*.

Para os *datasets* apresentados nas Figuras 4.7c e 4.7d, podemos verificar uma mudança de comportamento das métricas, onde a métrica precision segue melhorando sua acurácia em conjunto com a métrica recall. Podemos atribuir o comportamento apresentado na Figura 4.7c ao fato de que o *dataset* Cocktail Party foi registrado em um ambiente interno de tamanho reduzido, apresentando grupos de indivíduos mais coesos (Figura 4.5b), o que implica na redução da possibilidade de mesclagem entre grupos distintos como ocorrido com os *datasets* anteriormente citados. Já para o *dataset* Coffee Break, os melhores níveis de acurácia foram obtidos para distâncias sociais acima de $2.88m$. Podemos inferir que por ser o único *dataset* em ambiente ao ar livre, a ocorrência de grupos de interação mais esparsos seja maior que nos demais *datasets*.

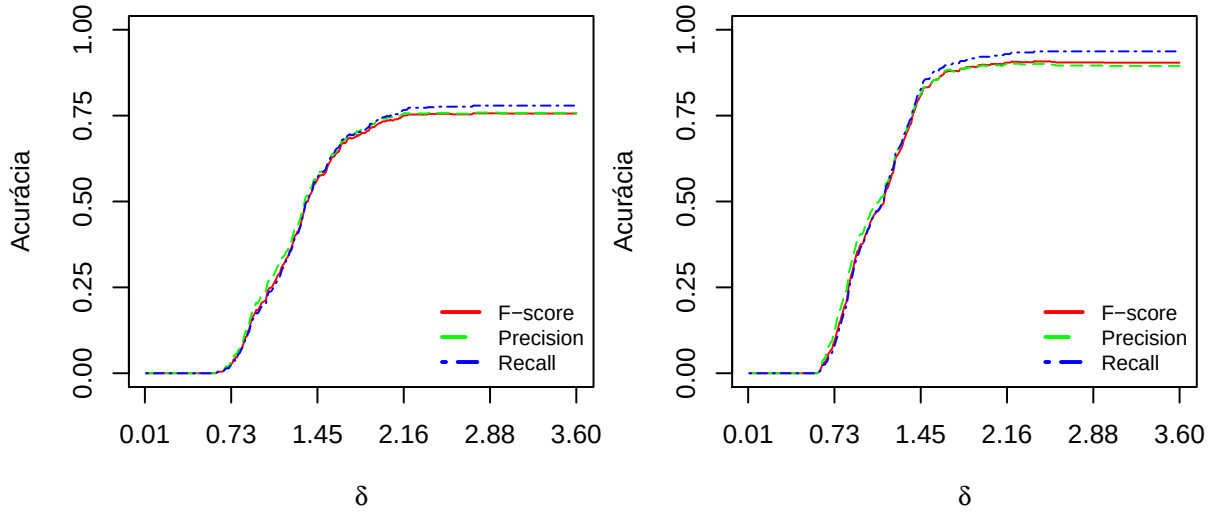
Uma vez que as melhores distâncias sociais já foram identificadas, o próximo passo a ser realizado é averiguar o impacto da existência de inconsistências temporais na informação entre *frames*. Partindo do pressuposto que toda interação é estabelecida ou



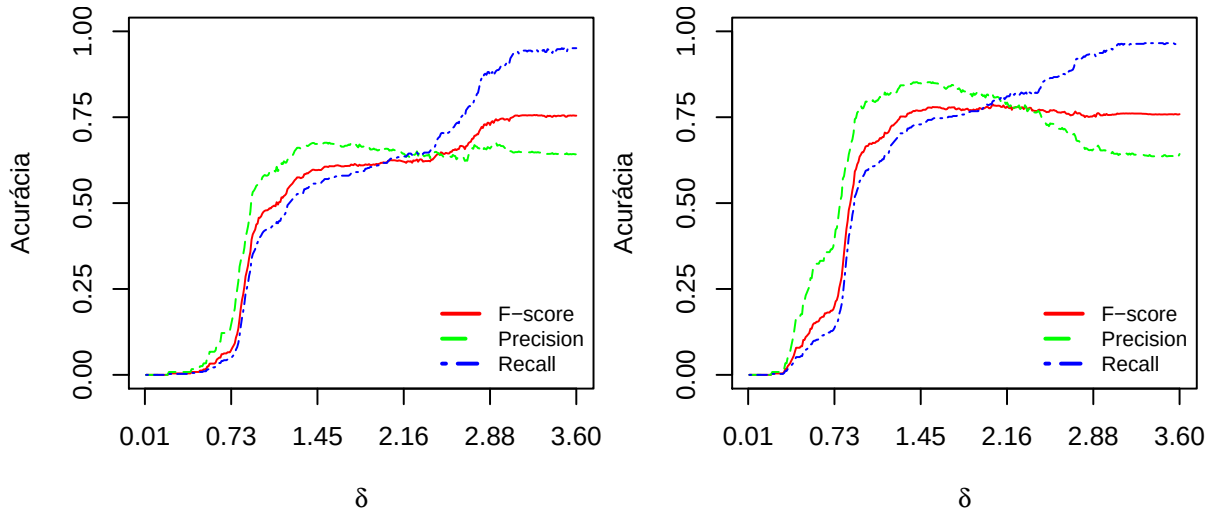
(a) Resultados para o *dataset* Synthetic com os limiares de tolerância $T = 1$ e $T = 2/3$.



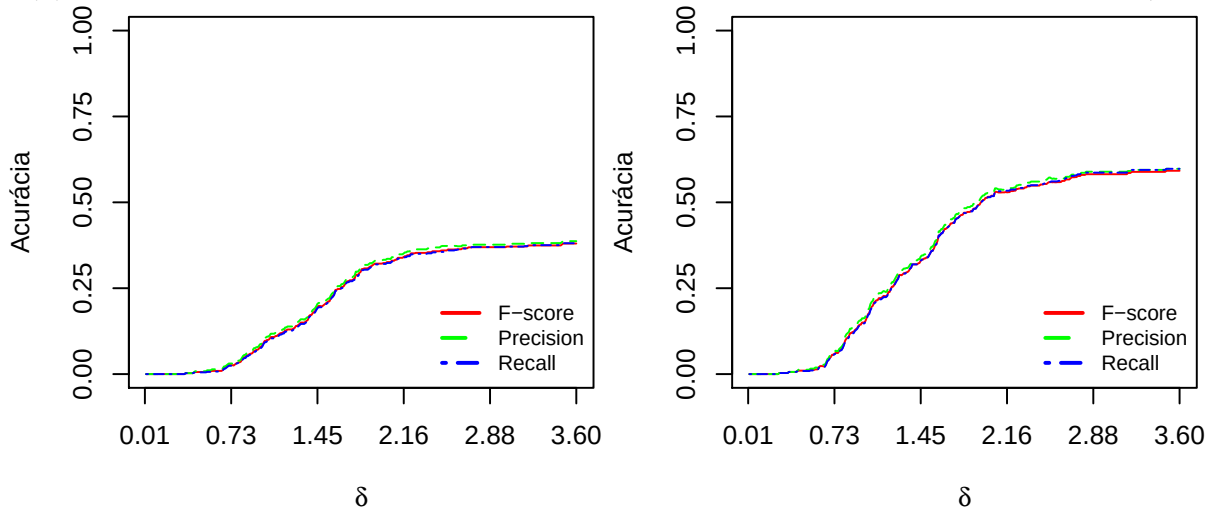
(b) Resultados para o *dataset* SALSA com os limiares de tolerância $T = 1$ e $T = 2/3$.



(c) Resultados para o *dataset* Cocktail Party com os limiares de tolerância $T = 1$ e $T = 2/3$.



(d) Resultados para o *dataset* Coffee Break com os limiares de tolerância $T = 1$ e $T = 2/3$.



(e) Resultados para o *dataset* GDet com os limiares de tolerância $T = 1$ e $T = 2/3$.

Figura 4.7: Avaliação do nível de acurácia para diferentes distâncias sociais.

desfeita de forma progressiva, realizamos uma filtragem temporal aplicando um filtro da mediana nos pesos correspondentes a determinação do LoI entre os pares de indivíduos, com a intenção de eliminar variações temporais bruscas no peso das arestas, visando amenizar os efeito da incerteza sobre os dados orientacional e posicional.

Os experimentos foram executados em todos os *datasets* que apresentam coerência temporal entre *frames*. O *dataset* Synthetic foi desprezado por se tratar de um conjunto de dados prioritariamente estático, possuindo apenas modificações situacionais a cada intervalo de 10 *frames*. Para os *datasets* que apresentam mais de uma sequência temporal (e.g. SALSA e Coffee Break), os testes foram realizados considerando cada parte individualmente. Foram escolhidos empiricamente os valores $\{1, 3, 5, 7$ e $9\}$ como parâmetros para variação da janela temporal (denotada como ws), onde o valor $\{1\}$ equivale a execução do método sem nenhum tipo de filtragem.

A Tabela 4.6 apresenta os resultados alcançados incluindo os melhores parâmetros

para a distância social e escolha da janela temporal. Uma melhor comparação visual é apresentada na Figura 4.8, onde são apresentados os resultados obtidos antes e depois da otimização dos parâmetros avaliados experimentalmente, considerando os limiares de tolerância $T = 1$ e $T = 2/3$. Os resultados para o filtro da mediana estão detalhados na Figura 4.9, sendo importante ressaltar que para cada um dos gráficos o eixo y foi reescalonado para uma melhor visualização. Novamente a maximização da métrica F_1-score foi utilizada como critério para definir o melhor tamanho de janela a ser utilizado como parâmetro de filtragem.

Tabela 4.6: Resultados incluindo os melhores parâmetros para cada *dataset* avaliado, onde a maximização da métrica F-score foi utilizada como critério para determinação dos melhores parâmetros de distância social e janela temporal.

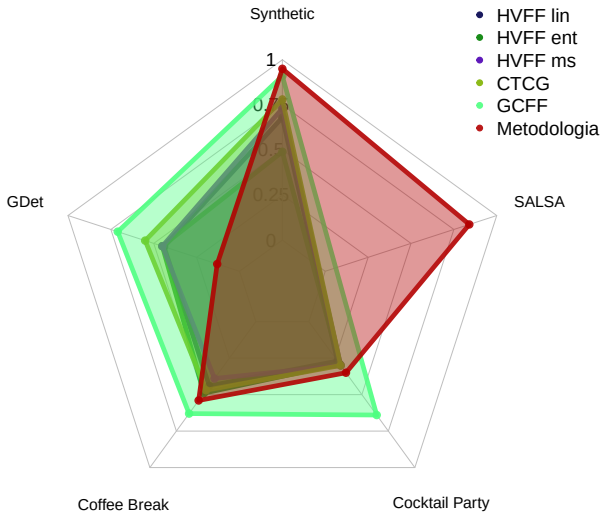
	T = 1					T = 2/3				
	Prec.	Rec.	F_1	δ	ws	Prec.	Rec.	F_1	δ	ws
Synthetic	**1.00**	**0.98**	**0.99**	1.25m	-	**1.00**	0.98	0.99	1.22m	-
SALSA	**0.85**	**0.94**	**0.88**	1.43m	5	0.88	**0.98**	**0.92**	1.33m	5
Cocktail Party	**0.81**	**0.81**	**0.80**	2.37m	5	**0.92**	**0.94**	**0.92**	2.51m	5
Coffee Break	0.63	**1.00**	**0.77**	3.27m	9	**0.85**	0.84	0.82	2.04m	7
GDet	0.39	0.38	0.38	3.49m	-	0.60	0.60	0.59	3.49m	-

Em todos os casos analisados, foi possível observar uma melhora da acurácia após à aplicação do método de filtragem. Nestes experimentos destacam-se os resultados obtidos para o nível de tolerância $T = 1$. Por ser um critério mais rigoroso de avaliação, melhorias no processo de identificação das interações impactam diretamente no acréscimo da acurácia. O maior nível de acréscimo da acurácia ocorreu para o *dataset* Cocktail Party (Figura 4.9b). Este resultado é condizente com o gráfico apresentado na Figura 4.2c, pois este *dataset* foi o que apresentou uma maior intermitência temporal, comportamento indicativo da presença de incerteza nos dados analisados.

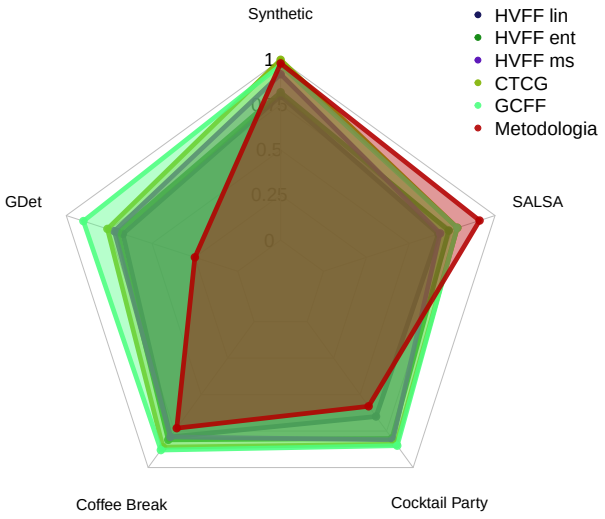
4.3 Espaço Social Atômico/Coletivo

A identificação e representação eficiente de grupos de pessoas é útil em diferentes tarefas. Esta seção mostra a capacidade de nossa abordagem para agrupar pessoas de acordo com o nível de interação entre indivíduos e simultaneamente definir uma zona de conforto considerando o espaço interpessoal de cada membro (ou grupo).

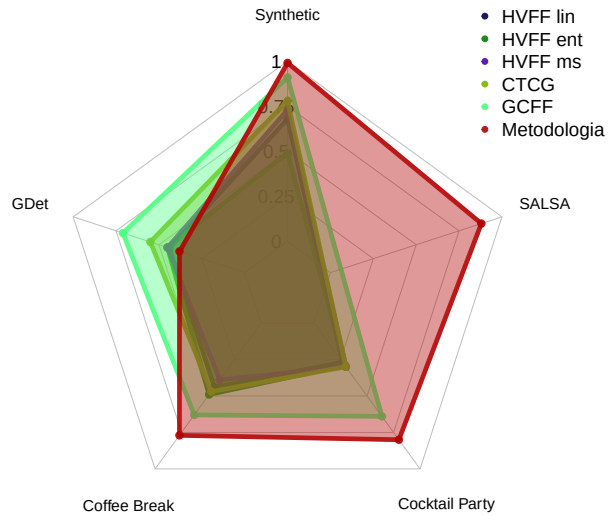
Inicialmente, um cenário sintético foi elaborado para demonstrar os diferentes níveis de espaços sociais coletivos (Figura 4.10), onde a cardinalidade dos grupos formados está diretamente relacionada com o tipo de espaço social definido para o limiar



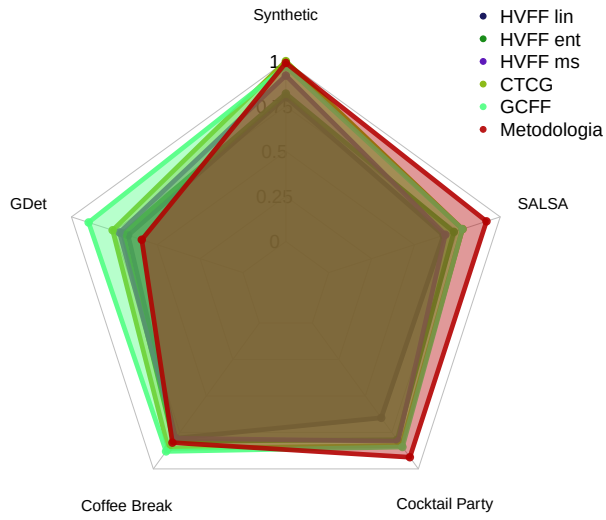
(a) Limiar de tolerância $T = 1$.



(b) Limiar de tolerância $T = 2/3$.



(c) Limiar $T = 1$ após otimizar parâmetros.

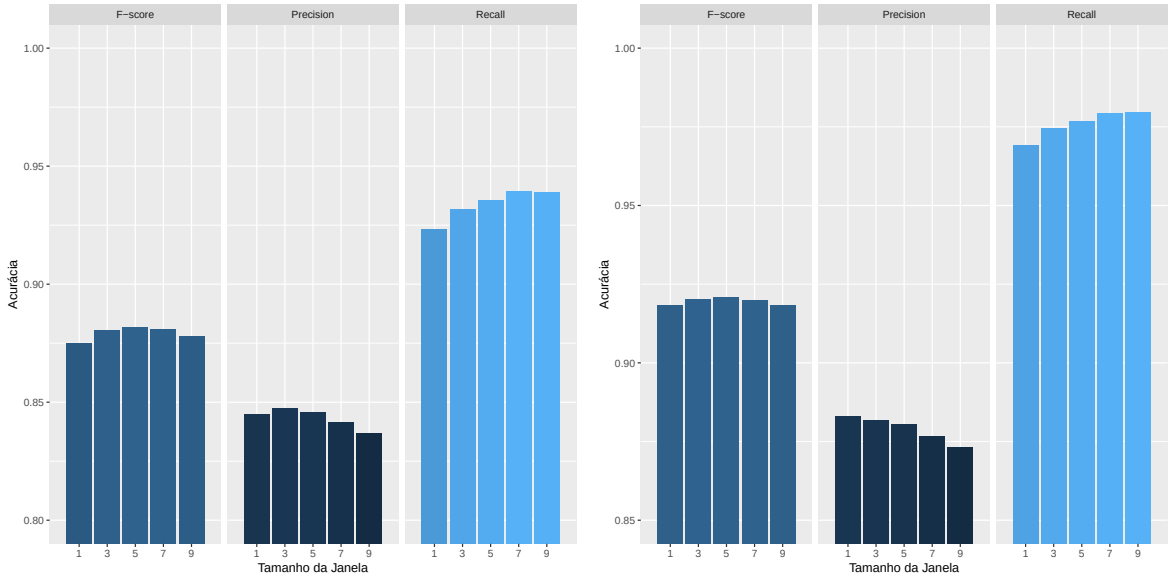


(d) Limiar $T = 2/3$ após otimizar parâmetros.

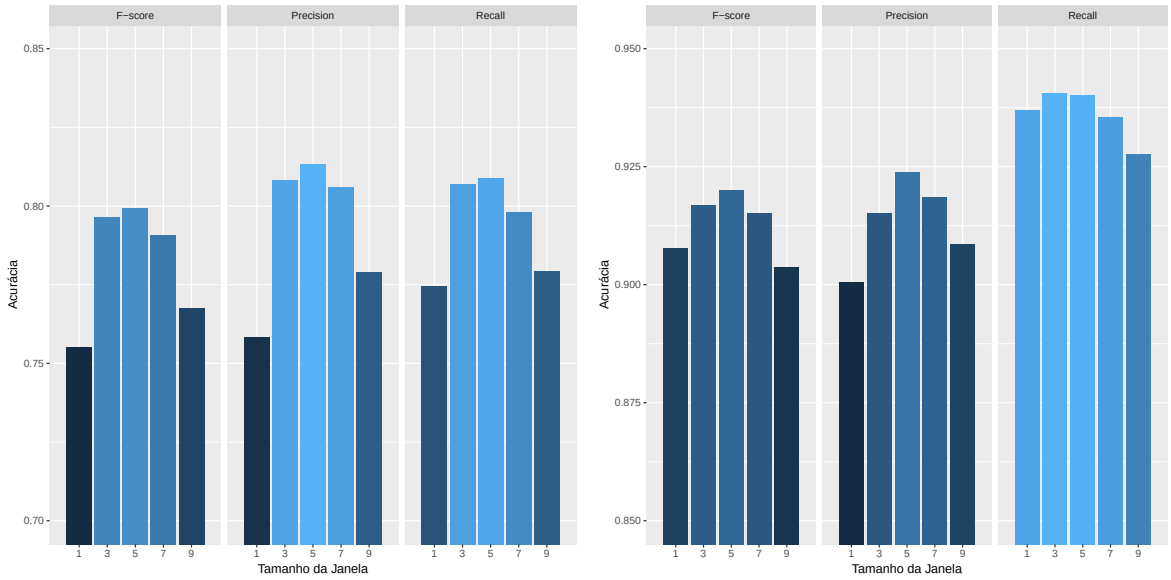
Figura 4.8: Gráfico de radar contendo as variáveis quantitativas referentes ao F_1-score incluindo os resultados obtidos após a otimização dos parâmetros para cada *dataset*.

δ. Um total de dezoito indivíduos estáticos foram espacialmente configurados de forma que houvesse diferentes cardinalidades de grupos sendo formados para cada uma das zona social especificadas.

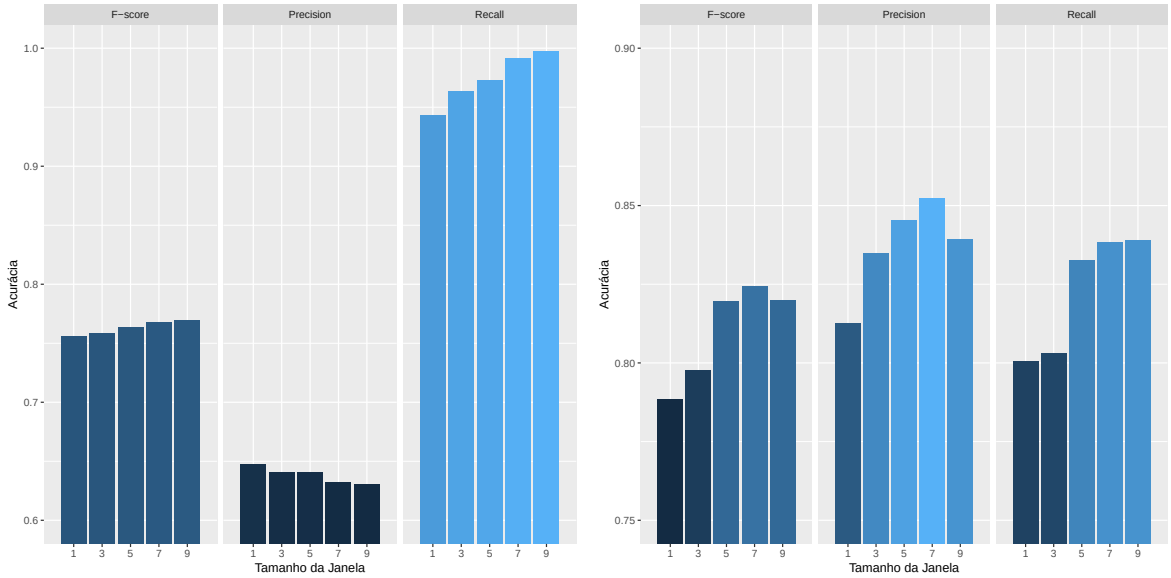
Os dois primeiros cenários (Figuras 4.10a e 4.10b) demonstram o método sendo aplicado às zonas Íntima e Pessoal, onde consideramos as premissas da interação do sistema de F-formation (Seção 3.1.1) juntamente com as regras da proxemics para determinar o conceito de espaço social coletivo. No primeiro caso, quase não existem grupos, configuração que é rapidamente modificada pelo aumento da distância social. No entanto, quando começamos a considerar as zonas sociais superiores a Pessoal, como Social e Pública (Figuras 4.10c e 4.10d), o agrupamento de indivíduos através da interação direta não faz mais sentido. Então, para casos subsequentes, os indivíduos são agrupados apenas



(a) Resultados para o *dataset* SALSA com os limiares de tolerância $T = 1$ e $T = 2/3$.

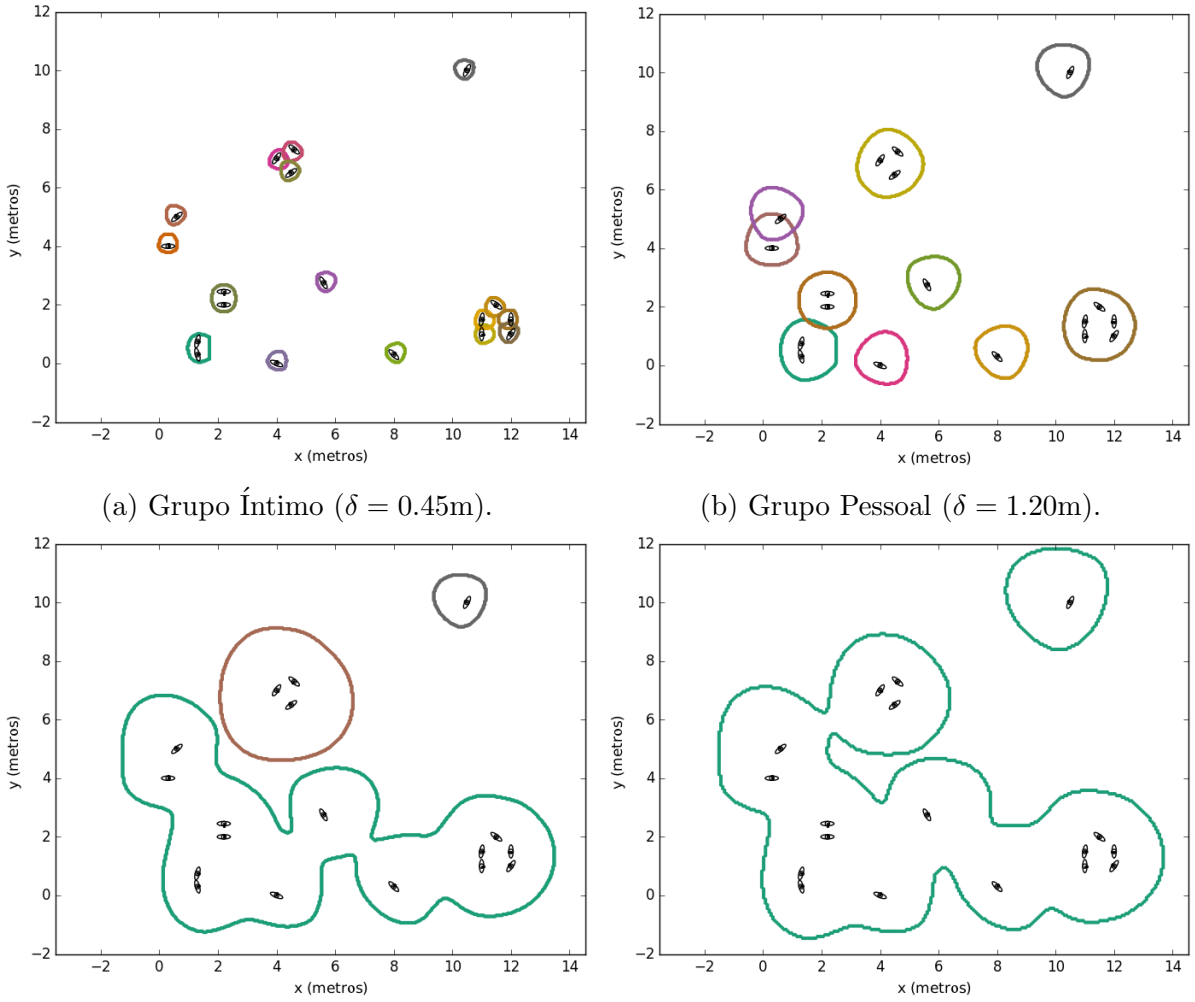


(b) Resultados para o *dataset* Cocktail Party com os limiares de tolerância $T = 1$ e $T = 2/3$.



(c) Resultados para o *dataset* Coffee Break com os limiares de tolerância $T = 1$ e $T = 2/3$.

Figura 4.9: Nível de acurácia aplicando filtro da mediana para diferentes janelas.



(a) Grupo Íntimo ($\delta = 0.45$m).

(b) Grupo Pessoal ($\delta = 1.20$m).

(c) Grupo Social ($\delta = 3.60$m).

(d) Grupo Público ($\delta > 3.60$m).

Figura 4.10: Representação do espaço social atômico/coletivo para diferentes zonas de proxemics.

de acordo com as regras da proxemics. É importante ressaltar que nestes últimos casos, o espaço alinhado em direção a frente do indivíduo unitário é estabelecido na distância pessoal, uma vez que manter distâncias muito grandes não tem muitas aplicações práticas em sistemas envolvendo HRI.

Considerando a zona pública, uma observação pode ser feita: devido à grande distância entre os indivíduos pertencentes ao mesmo grupo, mesmo com a mistura de Gaussianas, uma baixa densidade de probabilidade é gerada entre os indivíduos mais distantes. Esse tipo de vale formado dentro do grupo pode, eventualmente, deixar regiões específicas da função abaixo do limiar, criando divisões dentro de um único grupo, como observado na Figura 4.10d.

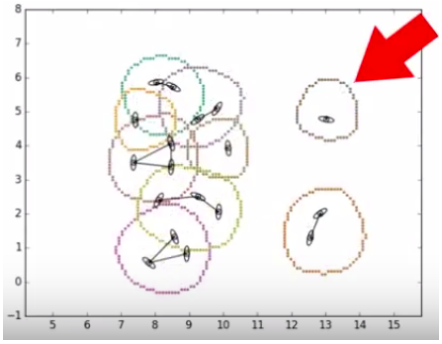
Visando demonstrar a execução do método proposto com dados do mundo real em ambiente dinâmico, apresentamos na Figura 4.11 uma sequência de *frames* correspondente ao textitdataset SALSA. Mais especificamente, sua sequência relacionada ao SALSA Coffee Break, devido a maior dinamicidade entre *frames* se comparado com a

sequência do SALSA Poster Session. No conjunto de *frames* selecionados, foi destacado o rastreamento de um indivíduo em específico enquanto o mesmo se deslocava de um grupo de interação para outro, de forma a demonstrar o surgimento e o desaparecimento de diferentes fronteiras relativas ao espaço interpessoal. Nas Figuras 4.12 e 4.13, destacamos a formação do espaço social considerando indivíduos *singleton* e em grupo.

Como forma de validar a aplicabilidade do método proposto em aplicações envolvendo HRI e simultaneamente avaliar a representação do espaço social coletivo, foi analisado por intermédio de simulações a capacidade de um robô móvel gerar caminhos socialmente aceitáveis para diferentes configurações sociais e distâncias interpessoais. As configurações utilizadas para os conjuntos de experimentos, assim como as métricas adotadas na realização dos procedimentos experimentais, foram empregues com base no trabalho apresentado em Pérez-Higueras et al. [47], visando assim obter uma forma de comparativo entre técnicas distintas. Diferente do que foi apresentado em Pérez-Higueras et al. [47], não consideraremos neste trabalho testes realizados em cenários reais e dinâmicos para avaliação da navegação social.



(a) Perspectiva câmera 1. (b) Perspectiva câmera 2.



(c) Espaço Social Coletivo.

Figura 4.12: Exemplo de formação do espaço social coletivo para $\delta = 1.25$m, destacando a presença de um indivíduo *singleton* e a delimitação do seu espaço pessoal no ambiente.

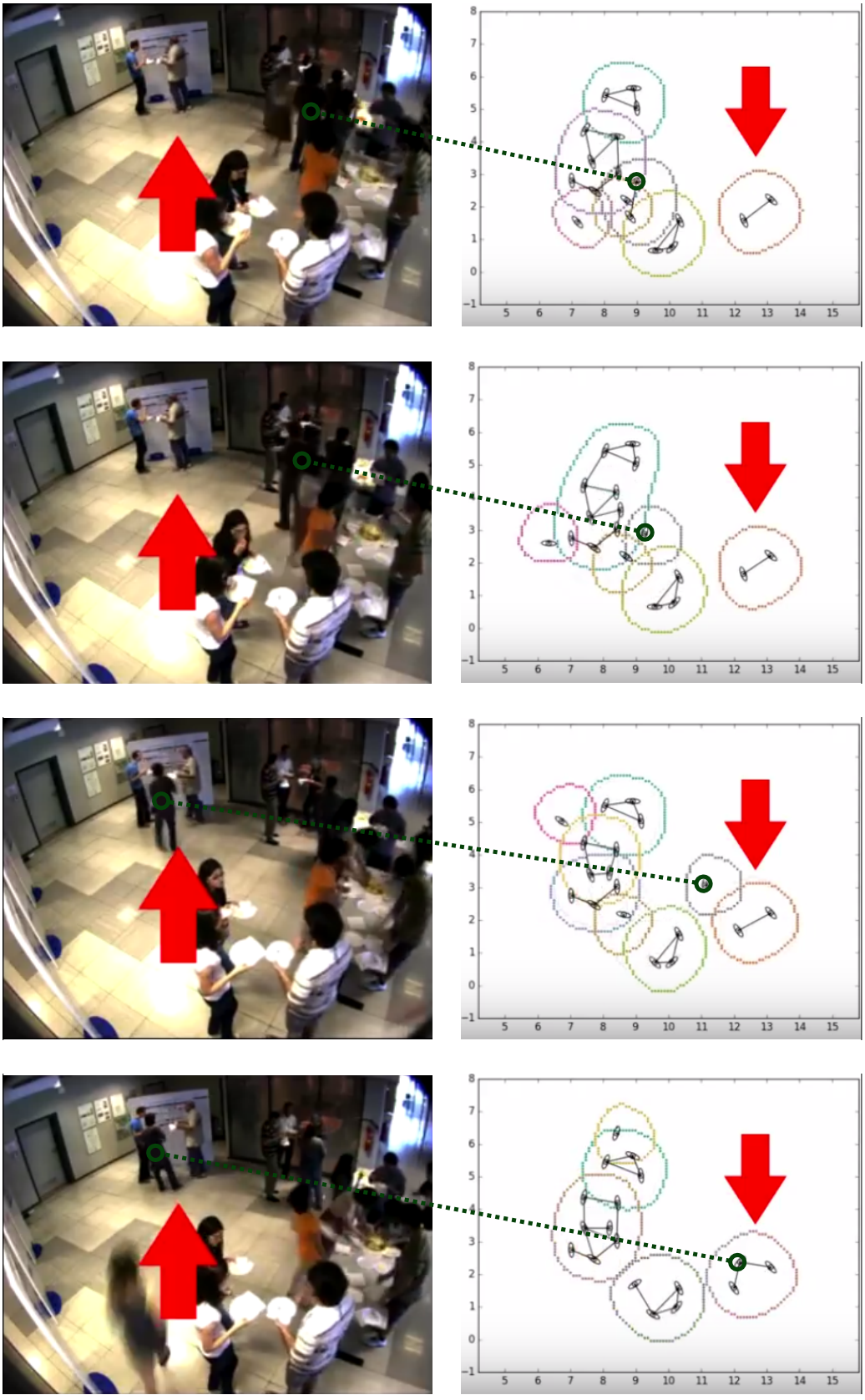
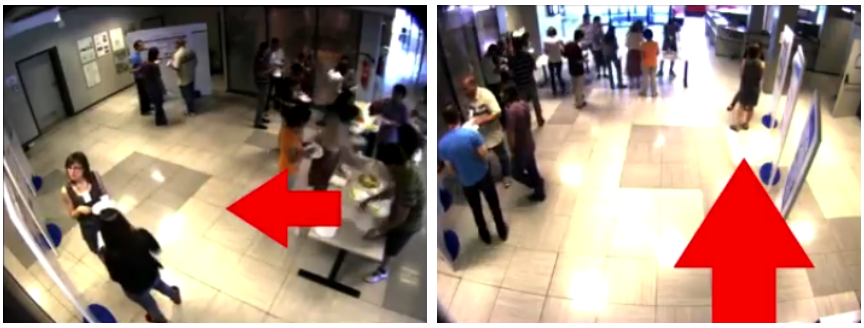
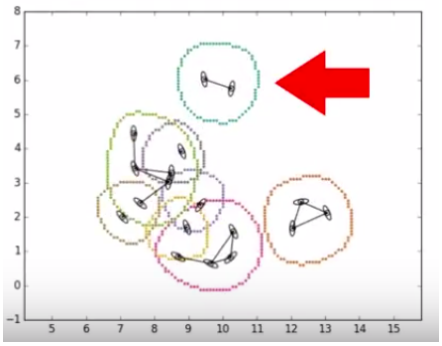


Figura 4.11: Sequência de dados temporal demonstrando a inserção de um novo indivíduo em um grupo de interação e a readaptação do espaço social coletivo para incorporar o espaço interpessoal de todos os indivíduos do grupo.



(a) Perspectiva câmera 1. (b) Perspectiva câmera 2.



(c) Espaço Social Coletivo.

Figura 4.13: Exemplo de formação do espaço social coletivo para $\delta = 1.25$m, destacando a interação entre dois indivíduos e a delimitação do espaço pessoal abrangendo o grupo.

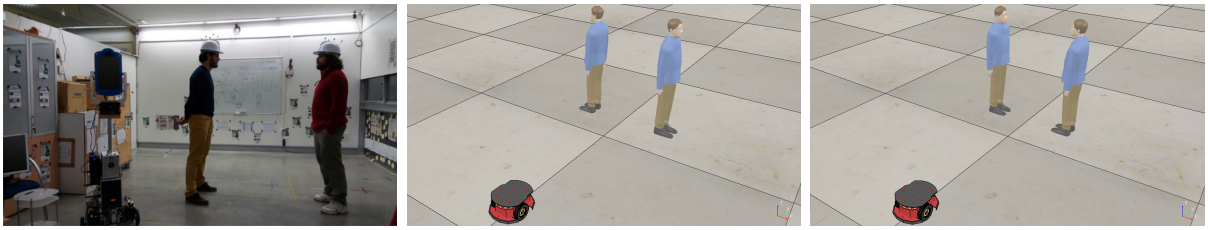
Para avaliar a representação das fronteiras de conforto social em grupos de interação, utilizamos como exemplo de aplicação o problema da navegação de um robô móvel na presença de pedestres, onde avaliamos por meio de simulações a representação do espaço social coletivo em ambiente estático e livre de obstáculos, utilizando os seguintes conjuntos de métricas retirados da literatura [47, 44]:

- Tempo necessário para atingir o objetivo (T_p).
- Comprimento do percurso (L_p).
- Alterações cumulativas de orientação (CHC).
- Distância média para a pessoa mais próxima (CP_{avg}).
- Distância mínima e máxima para pessoas (CP_{min} e CP_{max}).
- Intrusões de espaço pessoal (CP_{prox}).

A última métrica mencionada (CP_{prox}), refere-se à classificar a distância entre o robô e a pessoa mais próxima em cada etapa de tempo em um dos espaços de proxemics, e obter a porcentagem do tempo gasto em cada uma das zonas sociais durante toda a trajetória (apenas três distâncias foram consideradas na métrica CP_{prox}: Íntima, Pessoal e Social+Pública). O comportamento desejável dessa métrica para nosso método é indicar que o robô passou a maior parte do tempo na zona social superior à especificada.

Na Figura 4.14a pode ser observado um exemplo da configuração utilizada como referência para execução dos teste de navegação, onde duas situações diferentes em ambiente estático foram consideradas: na primeira, temos a formação de um grupo contendo um par de indivíduos interagindo mutuamente (Figura 4.14b); a segunda, apresenta um par de indivíduos opostamente orientados representando entidades singulares (Figura 4.14c).

Para cada conjunto de teste foram realizadas quatro repetições, onde o robô com uma configuração inicial ($x = 0.0$, $y = -2.6$) deverá alcançar uma configuração final ($x = 0.0$, $y = 2.6$). Os dois indivíduos foram posicionados a uma distância de 1.20 metros entre eles, possuindo as configurações posicionais de ($x = -0.6$, $y = 0.0$) e ($x = 0.6, y = 0.0$). Todos os experimentos foram realizados utilizando o simulador *Virtual Robot Experimentation Platform* (V-REP) [52], juntamente com a *framework Robot Operating System* (ROS) [48].



(a) Cenário real. (b) Indivíduos singulares. (c) Formação de grupo.

Figura 4.14: Configuração utilizada para avaliar a utilização do espaço social coletivo de forma socialmente aceitável (Figura (a) extraída de Pérez-Higueras et al. [47]).

Considerando as características de construção do espaço social coletivo aplicado neste trabalho, foi adotada a utilização da técnica artificial de Campos Potenciais [33] para a tarefa de navegação, por se adaptar melhor ao processo de navegação utilizando os pontos de borda criados no entorno dos grupos sociais. A função potencial artificial do algoritmo é definida segundo a fórmula

$$U(q) = U_{att}(q) + U_{rep}(q), \tag{4.3}$$

onde q denota a posição do robô, e as funções $U_{att}(q)$ e $U_{rep}(q)$ denotam o potencial de atração e repulsão respectivamente como

$$-\nabla U_{att(q)} = -k_{att} \cdot \|q - q_{goal}\| \tag{4.4}$$

e

$$-\nabla U_{rep(q)} = \begin{cases} k_{rep}\left(\frac{1}{\rho(q)} - \frac{1}{\rho_0}\right)\frac{1}{\rho^2(q)}\frac{q-q_{obs}}{\rho(q)} & , \text{se } \rho(q) \leq \rho_0 \\ 0 & , \text{se } \rho(q) > \rho_0 \end{cases} \tag{4.5}$$

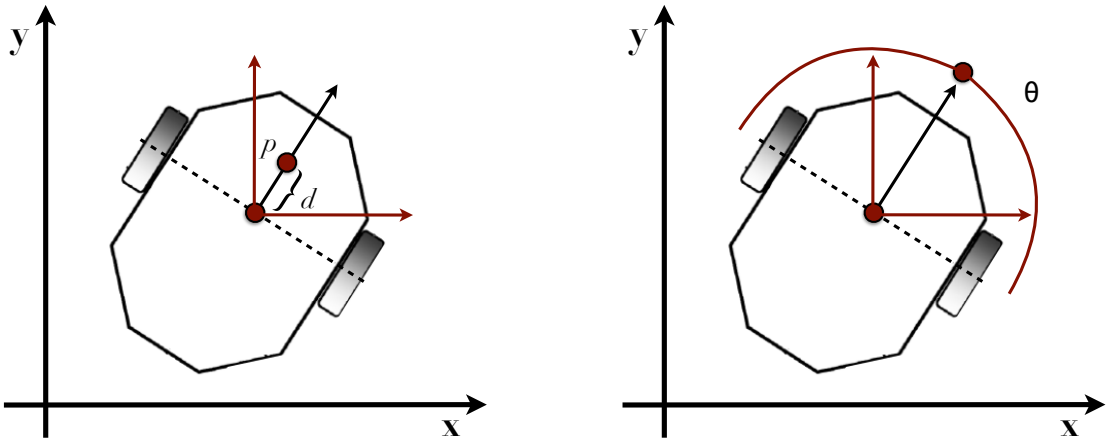
onde k_{att} e k_{rep} representam o fator escalar (ganho), q_{goal} a posição objetivo, $\rho(q)$ a distância para um obstáculo e ρ_0 a distância mínima de influência.

Como configuração do modelo de repulsão do algoritmo de Campos Potenciais, utilizou-se o potencial de repulsão tangencial para realizar o desvio dos obstáculos, onde adotamos os seguintes valores para os parâmetros: $k_{att} = 0.10$; $k_{rep} = 1.00$; $\rho_0 = 1.00m$ e $q_{goal}(x, y) = \{0.0, 2.7\}$. Posteriormente, definimos um controle $u = -\nabla_q U(q)$, onde ∇_q é o gradiente em relação a coordenadas do robô, e calculamos a sua integral em relação ao tempo, para encontrar um vetor de velocidade correspondente

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \int u \, dt, \tag{4.6}$$

no qual, $\dot{x}$ e $\dot{y}$ denotam as coordenadas de velocidade desejadas. Em seguida, este vetor é utilizado como uma entrada em um esquema de linearização de *feedback* estático [19], fornecendo as informações de velocidade linear v_i e angular w_i para o modelo diferencial do robô Pioneer p3dx utilizado nos experimentos simulados, dado pela equação:

$$\begin{bmatrix} v_i \\ w_i \end{bmatrix} = \begin{bmatrix} cos(\theta_i) & sin(\theta_i) \\ \frac{-sin(\theta_i)}{d} & \frac{cos(\theta_i)}{d} \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix}. \tag{4.7}$$



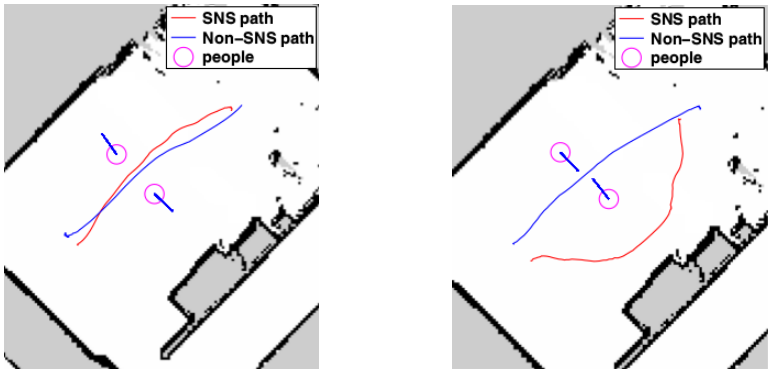
(a) (b)

Figura 4.15: Configurações aplicadas ao modelo de navegação diferencial do Pioneer p3dx, sendo o parâmetro d constante correspondente ao deslocamento do ponto p referente ao centroide do robô projetado e o parâmetro θ correspondente ao ângulo de projeção permitido para o deslocamento do robô.

O modelo apresentado na Equação 4.7 requer a especificação do parâmetros $d > 0$, constante correspondente ao deslocamento do ponto referente ao centroide do robô (Figura 4.15a), sendo utilizado neste caso o valor de $d = 2.5m$. Para o modelo de navegação previamente explicado, foram impostos como restrições uma velocidade linear máxima fixada em 0.25m/s e a impossibilidade de inversão linear dos movimentos (Figura 4.15b), de forma que o robô sempre realize continuamente o desvio de obstáculos indo para frente.

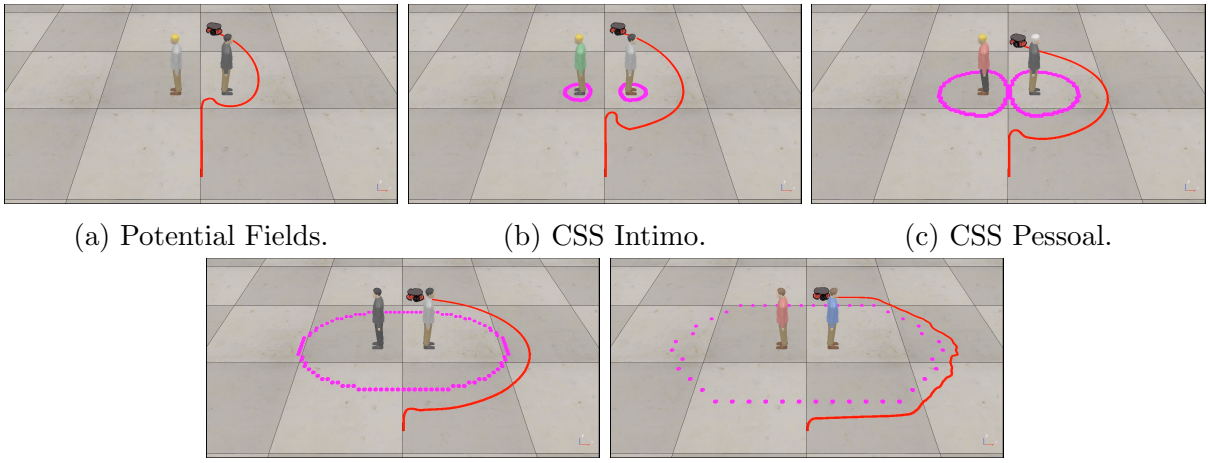
A execução dos experimentos foi dividida em cinco fases distintas, tanto para a configuração de ambiente incluindo os indivíduos singulares quanto para os grupos

de interação. Na primeira fase, o algoritmo Potential Fields foi aplicado como método *baseline* de comparação, considerando apenas a posição dos indivíduos como obstáculos a serem contornados. Nas quatro fases seguintes, o algoritmo foi aplicado juntamente com o método de detecção do espaço social coletivo (denotado como *Collective Social Space* (*Espaço Social Coletivo*, CSS)), dessa vez considerando os pontos de borda do espaço social como obstáculos a serem contornados. Para cada execução, foi requerida a identificação de uma zona pessoal distinta. As trajetórias do robô em ambas as etapas dos experimentos são apresentadas nas Figuras 4.16, 4.17 e 4.18.



(a) Pessoas singulares (b) Formação de grupo

Figura 4.16: Trajetória executadas pelo robô nos experimentos realizados por Pérez-Higueras et al. [47].

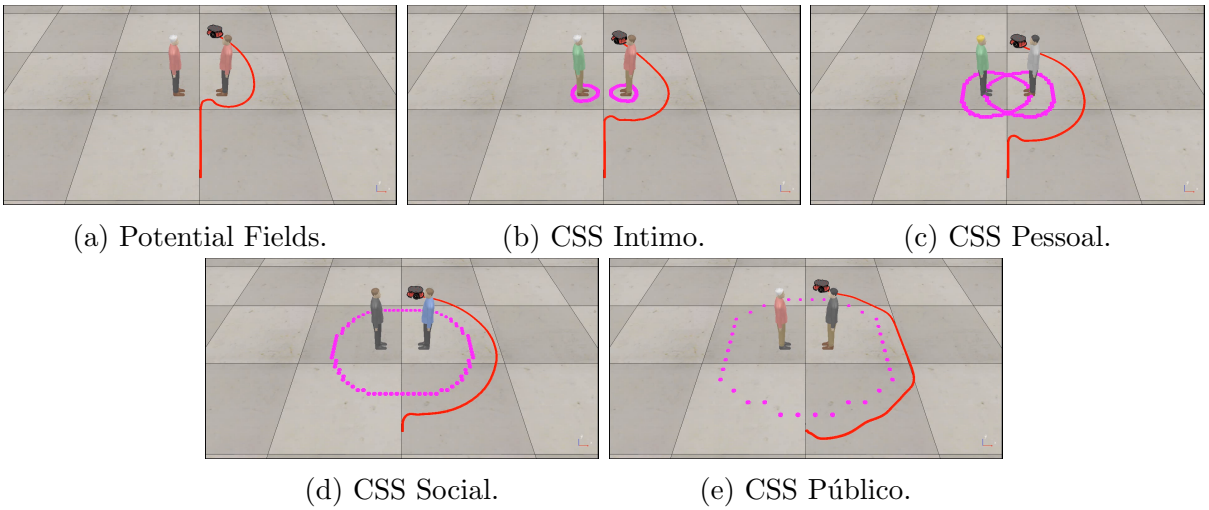


(a) Potential Fields. (b) CSS Intimo. (c) CSS Pessoal.

(d) CSS Social. (e) CSS Público.

Figura 4.17: Trajetórias executadas pelo robô para indivíduos singulares considerando as diferentes etapas dos experimentos.

Os resultados relativos à avaliação da navegação social apresentado em Pérez-Higueras et al. [47] foram adicionados como extensão dos métodos de comparação, onde foram realizadas comparações entre as capacidades sociais de um sistema de navegação empregando a função de custo aprendida de um conjunto de demonstrações (denotado como *Social Navigation System* (SNS)) e a *framework* pertencente ao ROS move_base



(a) Potential Fields. (b) CSS Intimo. (c) CSS Pessoal.

(d) CSS Social. (e) CSS Público.

Figura 4.18: Trajetórias executadas pelo robô para grupo de indivíduos considerando as diferentes etapas dos experimentos.

empregue como sistema de navegação padrão (denotado como *Non-Social Navigation System* (Non-SNS)), sendo todas as entidades pertencentes a cena consideradas como simples obstáculos.

Todos os dados utilizados para avaliação do sistema de navegação social estão reunidos nas Tabelas 4.7 e 4.8. Para melhor visualização e discussão dos resultados obtidos em nossos testes, cada uma das métricas pode ser separadamente observada nas figuras apresentadas no Apêndice A, contendo as cinco fases de execução incluindo o algoritmo *baseline* Potential fields e o CSS.

No caso do cenário incluindo duas pessoas que não formam um grupo, ao observarmos a Tabela 4.7 conseguimos verificar inicialmente uma disparidade de valores entre os métodos aplicados neste trabalho e os resultados extraídos de Pérez-Higueras et al. [47]. Essa divergência advém principalmente da diferença de arquiteturas entre ambas as abordagens. Tanto o SNS quanto o Non-SNS utilizam uma arquitetura de planejamento global, enquanto o Potential Fields tem como base um sistema de navegação local reativo, já sendo esperado a obtenção de valores elevados para as métricas T_p, L_p e CHC em nossos resultados. No entanto, os comportamentos obtidos em algumas métricas podem ser destacados: a distância mínima para as pessoas (CP_{min}) se amplifica em virtude da zona social requerida para o CSS. Esse comportamento esta melhor evidenciado pela métrica CP_{prox}, onde diferentemente dos métodos *baseline* de comparação, a navegação do robô pode ser delimitada para diferentes zonas de proxemics.

Para os cenários que consideram o par de indivíduos formando um grupo de interação (Tabela 4.8), à abordagem de CSS novamente foi capaz de prover rotas socialmente aceitáveis conforme as diferentes zonas interpessoais. É importante ressaltar que nesta configuração em particular, a trajetória do robô perpassa pelas costas dos indivíduos, onde a distância interpessoal (representada por uma AGF) sofre uma diminuição, justifi-

cando o decrescimento dos valores de CP_{min} em comparação aos resultados obtidos com os indivíduos singulares. Consecutivamente, pelo mesmo motivo anteriormente descrito, o tempo de permanência do robô em cada uma das zonas sociais é acrescido, como indicado pela métrica CP_{prox}.

Tabela 4.7: Resultados para a avaliação da navegação social em um cenário estático com duas pessoas.

	Tp (s)	Lp (m)	CHC (rad)	CPavg (m)	CPmin (m)	CPmax (m)	CPprox (%)		
							Íntima	Pessoal	Social+
Campos Potenciais	44.03 ± 2.22	8.77 ± 0.31	18.75 ± 0.20	2.00 ± 0.05	0.42 ± 0.00	3.03 ± 0.05	1.28 ± 0.09	23.12 ± 1.12	75.60 ± 1.21
Non-SNS	22.70 ± 1.52	4.86 ± 0.40	4.54 ± 0.79	1.74 ± 0.09	1.08 ± 0.16	2.47 ± 0.13	0.00 ± 0.00	12.77 ± 1.02	87.23 ± 0.87
SNS	24.18 ± 1.73	4.92 ± 0.23	4.37 ± 0.40	1.90 ± 0.14	0.63 ± 0.07	2.82 ± 0.01	0.00 ± 0.00	17.24 ± 0.98	82.76 ± 0.98
CSS Intimo	52.54 ± 5.67	11.79 ± 0.50	22.60 ± 6.02	2.01 ± 0.11	0.87 ± 0.01	2.94 ± 0.12	0.00 ± 0.00	9.44 ± 7.05	90.56 ± 7.05
CSS Pessoal	43.24 ± 1.80	11.99 ± 0.10	18.50 ± 0.02	1.92 ± 0.01	1.42 ± 0.00	2.67 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	100 ± 0.00
CSS Social	35.76 ± 0.31	12.61 ± 0.11	18.23 ± 0.17	2.34 ± 0.01	2.07 ± 0.01	2.67 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	100 ± 0.00
CSS Público	40.94 ± 0.35	14.45 ± 0.14	15.62 ± 5.74	2.71 ± 0.02	2.29 ± 0.00	3.22 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	100 ± 0.00

Tabela 4.8: Resultados para a avaliação da navegação social em um cenário estático com um grupo formado por duas pessoas.

	Tp (s)	Lp (m)	CHC (rad)	CPavg (m)	CPmin (m)	CPmax (m)	CPprox (%)		
							Íntima	Pessoal	Social+
Campos Potenciais	37.32 ± 3.89	8.20 ± 0.28	18.21 ± 0.26	1.80 ± 0.13	0.42 ± 0.00	2.84 ± 0.13	1.71 ± 0.18	27.81 ± 3.30	70.48 ± 3.47
Non-SNS	23.15 ± 1.34	5.02 ± 0.08	4.41 ± 0.08	1.88 ± 0.12	0.42 ± 0.15	2.74 ± 0.17	2.60 ± 3.67	14.50 ± 3.06	82.90 ± 0.61
SNS	23.65 ± 0.64	5.45 ± 0.27	4.18 ± 2.55	1.87 ± 0.08	0.79 ± 0.17	2.71 ± 0.11	0.00 ± 0.00	17.83 ± 1.52	82.17 ± 1.52
CSS Intimo	37.21 ± 1.76	10.03 ± 0.32	30.38 ± 0.10	1.67 ± 0.03	0.86 ± 0.17	2.67 ± 0.00	0.00 ± 0.00	34.20 ± 1.48	65.80 ± 1.48
CSS Pessoal	35.84 ± 1.05	10.39 ± 0.32	17.87 ± 0.08	1.79 ± 0.01	1.13 ± 0.36	2.67 ± 0.00	0.00 ± 0.00	0.03 ± 0.07	99.97 ± 0.07
CSS Social	30.16 ± 0.63	10.64 ± 0.22	17.42 ± 1.17	2.07 ± 0.01	1.77 ± 0.01	2.67 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	100 ± 0.00
CSS Público	33.16 ± 0.49	11.70 ± 0.17	19.38 ± 0.05	2.35 ± 0.01	1.88 ± 0.00	2.80 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	100 ± 0.00

Capítulo 5

Conclusão e Trabalhos Futuros

5.1 Conclusão

Nesse trabalho foi apresentada uma nova abordagem baseada em grafo com o objetivo de identificar possíveis associações de indivíduos no ambiente de acordo com sua posição e orientação corporal. Posteriormente, com base nessa detecção dos grupos de interação, propomos uma representação do espaço social coletivo baseada em uma função de densidade, por intermédio da junção de funções Gaussianas assimétricas para delimitação do espaço socialmente aceitável, de forma que um robô móvel consiga realizar tarefas na presença de pedestres, minimizando qualquer impacto social para os indivíduos presentes no ambiente.

A maioria das abordagens relacionadas a HRI se restringem em abordar apenas situações contendo uma pessoa ou pequenos grupos de indivíduos. A definição de um espaço social coletivo é um problema complexo, pois requer a determinação dos grupos de interação e do espaço interpessoal, abrangendo tanto os indivíduos agrupados quanto os isolados, considerando as diferentes zonas sociais referentes às regras de proxemics. Nesse sentido, as principais contribuições geradas neste trabalho são: 1) uma metodologia baseada em grafos para identificar possíveis associações de indivíduos no ambiente, onde definimos matematicamente um modelo de interação de acordo com a posição e orientação corporal entre os pares de indivíduos presentes no ambiente, sendo possível segregar diferentes grupos de interação de acordo com as zonas sociais especificadas nas regras de proxemics. 2) a representação do espaço social coletivo a partir da detecção dos grupos de interação, onde a determinação de tais limites sociais oferece a possibilidade de tratar os grupos como entidades únicas, delimitando o acesso indevido a sua zona de conforto e garantindo flexibilidade no uso de abordagens previamente desenvolvidas no contexto de HRI.

Em cada etapa de desenvolvimento do projeto, foram realizadas uma série de análises experimentais. Inicialmente, foram utilizados *datasets* sintéticos e do mundo real para avaliar a determinação da associação dos indivíduos, sendo possível demonstrar que

a metodologia proposta apresenta performance superior aos trabalhos do estado-da-arte na maioria dos casos. Posteriormente, foi realizada a análise da representação do espaço social coletivo por intermédio de simulações, onde a metodologia proposta foi capaz de gerar comportamentos mais socialmente favoráveis do que métodos que não consideram tais restrições. Como principais limitações, temos: a dependência de uma boa qualidade de detecção das informações posicionais e orientacionais dos indivíduos na cena; e possibilidade de subdivisões dentro de um único grupo, ocorrência gerada devido a concentração esparsa dos indivíduos, criando vales nas regiões de baixa densidade de probabilidade durante o processo de mistura das Gaussianas.

5.2 Trabalhos Futuros

Uma das principais dificuldades na abordagem proposta reside em encontrar o limiar de segmentação para determinar o espaço social coletivo em grupos de indivíduos. Visando melhorar a metodologia proposta, uma solução alternativa seria o estudo de possíveis técnicas que já permitam a geração direta de bordas a partir da utilização do espaço interpessoal de cada indivíduo, por exemplo, Convex Hull [9], entretanto com uma maior suavização da linha de contorno. Do mesmo modo, é primordial a proposição de técnicas de filtragem para lidar com as inconsistências temporais.

Outros pontos interessantes para proposição de trabalhos futuros incluem a extensão da metodologia para lidar com a possível associação entre humanos e objetos na cena, por exemplo, uma vitrine ou televisão. Também pretendemos usar a metodologia em tarefas como navegação socialmente aceitável, abrangendo à abordagem de grupos e escolta de indivíduos. Além disso, a determinação de limites fronteiriços tem grande aplicabilidade na navegação de sistemas multi-agentes. Dessa forma, é interessante a extensão do trabalho para o utilização simultânea de múltiplos robôs. Finalmente, uma evolução da aplicabilidade para o método seria explorar as informações pertinentes ao espaço social coletivo, como por exemplo a direção e velocidade de deslocamento dos grupos, concentração espacial dos focos de atenção, área ocupada pelo grupo, ocorrência de união e dispersão de interações, entre outras características que possam ser utilizadas para viabilizar o aprendizado de informações com alto nível semântico, como detecção de eventos suspeitos, receptividade dos grupos, dentre outros.

Referências

[1] M. Al-Hamdani. Preferred interpersonal distances: A global comparison. *Journal of Business Communication*, 47(2):189–227, 2010.

[2] X. Alameda-Pineda, J. Staiano, R. Subramanian, L. Batrinca, E. Ricci, B. Lepri, O. Lanz, and N. Sebe. Salsa: A novel dataset for multimodal group behavior analysis. *IEEE transactions on pattern analysis and machine intelligence*, 38(8):1707–1720, 2016.

[3] T. Amaoka, H. Laga, and M. Nakajima. Modeling the personal space of virtual agents for behavior simulation. In *CyberWorlds, 2009. CW'09. International Conference on*, pages 364–370. IEEE, 2009.

[4] J. C. Baxter. Interpersonal spacing in natural settings. *Sociometry*, pages 444–456, 1970.

[5] L. Bazzani, M. Cristani, D. Tosato, M. Farenzena, G. Paggetti, G. Menegaz, and V. Murino. Social interactions by visual focus of attention in a three-dimensional environment. *Expert Systems*, 30(2):115–127, 2013.

[6] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. The interactive museum tour-guide robot. In *Aaai/iaai*, pages 11–18, 1998.

[7] D. Carton, A. Turnwald, K. Kühnlenz, D. Wollherr, and M. Buss. Proactive human approach in dynamic environments. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3320–3321, Oct 2012.

[8] K. Charalampous, I. Kostavelis, and A. Gasteratos. Recent trends in social aware robot navigation: A survey. *Robotics and Autonomous Systems*, 2017.

[9] I. Chatterjee and A. Steinfeld. Performance of a low-cost, human-inspired perception approach for dense moving crowd navigation. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pages 578–585. IEEE, 2016.

[10] C. Chen and J.-M. Odobez. We are not contortionists: Coupled adaptive learning for head and body orientation estimation in surveillance video. In *Computer Vision and*

Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 1544–1551. IEEE, 2012.

[11] T. M. Ciolek and A. Kendon. Environment and the spatial arrangement of conversational encounters. *Sociological Inquiry*, 50(3-4):237–271, 1980.

[12] A. Cosgun, D. A. Florencio, and H. I. Christensen. Autonomous person following for telepresence robots. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, pages 4335–4342, May 2013. doi: 10.1109/ICRA.2013.6631191.

[13] A. Cosgun, E. A. Sisbot, and H. I. Christensen. Anticipatory robot path planning in human environments. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pages 562–569. IEEE, 2016.

[14] M. Cristani, L. Bazzani, G. Paggetti, A. Fossati, D. Tosato, A. Del Bue, G. Menegaz, and V. Murino. Social interaction discovery by statistical analysis of f-formations. In *BMVC*, volume 2, page 4, 2011.

[15] M. Cristani, R. Raghavendra, A. Del Bue, and V. Murino. Human behavior analysis in video surveillance: A social signal processing perspective. *Neurocomputing*, 100: 86–97, 2013.

[16] I. Damian, B. Endrass, P. Huber, N. Bee, and E. André. Individualized agent interactions. *Motion in games*, pages 15–26, 2011.

[17] K. Dautenhahn. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 362 (1480):679–704, 2007.

[18] N. Davies, K. Mitchell, K. Cheverst, and G. Blair. Developing a context sensitive tourist guide. In *1st Workshop on Human Computer Interaction with Mobile Devices, GIST Technical Report G98-1*, volume 1, 1998.

[19] J. P. Desai, J. Ostrowski, and V. Kumar. Controlling formations of multiple mobile robots. In *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, volume 4, pages 2864–2869. IEEE, 1998.

[20] D. Feil-Seifer and M. Matarić. People-aware navigation for goal-oriented behavior involving a human partner. In *Development and Learning (ICDL), 2011 IEEE International Conference on*, volume 2, pages 1–6. IEEE, 2011.

[21] A. Garrell and A. Sanfeliu. Local optimization of cooperative robot movements for guiding and regrouping people in a guiding mission. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 3294–3299. IEEE, 2010.

[22] W. Ge, R. T. Collins, and R. B. Ruback. Vision-based analysis of small groups in pedestrian crowds. *IEEE transactions on pattern analysis and machine intelligence*, 34(5):1003–1016, 2012.

[23] R. Gockley, J. Forlizzi, and R. Simmons. Natural person-following behavior for social robots. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, pages 17–24. IEEE, 2007.

[24] J. V. Gómez, N. Mavridis, and S. Garrido. Fast marching solution for the social path planning problem. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 2243–2248. IEEE, 2014.

[25] M. A. Goodrich and A. C. Schultz. Human-robot interaction: a survey. *Foundations and trends in human-computer interaction*, 1(3):203–275, 2007.

[26] E. T. Hall. A system for the notation of proxemic behavior. *American anthropologist*, 65(5):1003–1026, 1963.

[27] E. T. Hall. *The Hidden Dimension: Man's Use of Space in Public and Private.* The Bodley Head Ltd, 1966.

[28] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995.

[29] D. Helbing and P. Molnar. Self-organization phenomena in pedestrian crowds. *arXiv preprint cond-mat/9806152*, 1998.

[30] P. Henry, C. Vollmer, B. Ferris, and D. Fox. Learning to navigate through crowded environments. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 981–986. IEEE, 2010.

[31] H. Hung and B. Kröse. Detecting f-formations as dominant sets. In *Proceedings of the 13th international conference on multimodal interfaces*, pages 231–238. ACM, 2011.

[32] A. Kendon. *Conducting interaction: Patterns of behavior in focused encounters*, volume 7. CUP Archive, 1990.

[33] O. Khatib. Real-time obstacle avoidance for manipulators and mobile robots. In *Autonomous robot vehicles*, pages 396–404. Springer, 1986.

[34] R. Kirby. *Social robot navigation.* Carnegie Mellon University, 2010.

[35] T. Kruse, A. Kirsch, E. A. Sisbot, and R. Alami. Exploiting human cooperation in human-centered robot navigation. In *19th International Symposium in Robot and Human Interactive Communication*, pages 192–197. IEEE, 2010.

[36] T. Kruse, A. K. Pandey, R. Alami, and A. Kirsch. Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 61(12):1726–1743, 2013.

[37] H. Laga and T. Amaoka. Modeling the Spatial Behavior of Virtual Agents in Groups for Non-verbal Communication in Virtual Worlds. In *Proceedings of the 3rd International Universal Communication Symposium*, IUCS'09, pages 154–159, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-641-0. doi: 10.1145/1667780.1667811. URL http://doi.acm.org/10.1145/1667780.1667811.

[38] R. Mead and M. J. Matarić. Autonomous human–robot proxemics: socially aware navigation based on interaction potential. *Autonomous Robots*, 41(5):1189–1201, 2017.

[39] J. Müller, C. Stachniss, K. Arras, and W. Burgard. Socially inspired motion planning for mobile robots in populated environments. In *Proc. of International Conference on Cognitive Systems*, 2008.

[40] Y. Nakauchi and R. Simmons. A social robot that stands in line. *Autonomous Robots*, 12(3):313–324, 2002.

[41] V. K. Narayanan, A. Spalanzani, F. Pasteau, and M. Babel. On equitably approaching and joining a group of interacting humans. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 4071–4077. IEEE, 2015.

[42] V. K. Narayanan, A. Spalanzani, R. C. Luo, and M. Babel. Analysis of an adaptive strategy for equitably approaching and joining human interactions. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pages 341–346. IEEE, 2016.

[43] T. Ohki, K. Nagatani, and K. Yoshida. Collision avoidance method for mobile robot considering motion and personal spaces of evacuees. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 1819–1824. IEEE, 2010.

[44] B. Okal and K. O. Arras. Formalizing normative robot behavior. In *International Conference on Social Robotics*, pages 62–71. Springer, 2016.

[45] A. K. Pandey and R. Alami. A step towards a sociable robot guide which monitors and adapts to the person's activities. In *Advanced Robotics, 2009. ICAR 2009. International Conference on*, pages 1–8. IEEE, 2009.

[46] P. Papadakis, P. Rives, and A. Spalanzani. Adaptive spacing in human-robot interactions. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2627–2632. IEEE, 2014.

[47] N. Pérez-Higueras, F. Caballero, and L. Merino. Teaching robot navigation behaviors to optimal rrt planners. *International Journal of Social Robotics*, pages 1–15, 2017.

[48] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, page 5, 2009.

[49] J. Rios-Martinez, A. Spalanzani, and C. Laugier. Understanding human interaction for probabilistic autonomous navigation using risk-rrt approach. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2014–2019, Sept 2011. doi: 10.1109/IROS.2011.6094496.

[50] J. Rios-Martinez, A. Renzaglia, A. Spalanzani, A. Martinelli, and C. Laugier. Navigating between people: A stochastic optimization approach. In *Robotics and automation (ICRA), 2012 IEEE international conference on*, pages 2880–2885. IEEE, 2012.

[51] J. Rios-Martinez, A. Spalanzani, and C. Laugier. From Proxemics Theory to Socially-Aware Navigation: A Survey. *International Journal of Social Robotics*, 7(2):137–153, 2015. ISSN 1875-4791. doi: 10.1007/s12369-014-0251-1.

[52] E. Rohmer, S. P. N. Singh, and M. Freese. V-rep: a versatile and scalable robot simulation framework. In *Proc. of The International Conference on Intelligent Robots and Systems (IROS)*, 2013.

[53] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita. How to Approach Humans? – Strategies for Social Robots to Initiate Interaction. In *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 109–116, March 2009. doi: 10.1145/1514095.1514117.

[54] F. Setti, H. Hung, and M. Cristani. Group detection in still images by f-formation modeling: A comparative study. In *Image Analysis for Multimedia Interactive Services (WIAMIS), 2013 14th International Workshop on*, pages 1–4. IEEE, 2013.

[55] F. Setti, O. Lanz, R. Ferrario, V. Murino, and M. Cristani. Multi-scale f-formation discovery for group detection. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 3547–3551. IEEE, 2013.

[56] F. Setti, C. Russell, C. Bassetti, and M. Cristani. F-formation detection: Individuating free-standing conversational groups in images. *PloS one*, 10(5):e0123783, 2015.

[57] J. Shao, C. Change Loy, and X. Wang. Scene-independent group profiling in crowd. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2219–2226, 2014.

[58] R. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza. *Introduction to autonomous mobile robots*. MIT press, 2011.

[59] E. Sisbot, L. Marin-Urias, R. Alami, and T. Simeon. A Human Aware Mobile Robot Motion Planner. *IEEE Transactions on Robotics*, 23(5):874–883, oct. 2007. ISSN 1552-3098. doi: 10.1109/TRO.2007.904911.

[60] F. Solera, S. Calderara, and R. Cucchiara. Socially constrained structural learning for groups detection in crowd. *IEEE transactions on pattern analysis and machine intelligence*, 38(5):995–1008, 2016.

[61] A. Taylor and L. D. Riek. Robot perception of human groups in the real world: State of the art. In *AAAI Fall Symposium Series: Artificial Intelligence for Human-Robot Interaction Technical Report FS-16-01. Retrieved January*, volume 4, page 2017, 2016.

[62] K. N. Tran, A. Bedagkar-Gala, I. A. Kakadiaris, and S. K. Shah. Social cues in group formation and local interactions for collective activity analysis. In *VISAPP (1)*, pages 539–548, 2013.

[63] P. Trautman and A. Krause. Unfreezing the robot: Navigation in dense, interacting crowds. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 797–803. IEEE, 2010.

[64] X.-T. Truong and T.-D. Ngo. Dynamic social zone based mobile robot navigation for human comfortable safety in social environments. *International Journal of Social Robotics*, 8(5):663–684, 2016.

[65] J. Varadarajan, R. Subramanian, S. R. Bulò, N. Ahuja, O. Lanz, and E. Ricci. Joint estimation of human pose and conversational groups from social scenes. *International Journal of Computer Vision*, pages 1–20, 2017.

[66] S. Vascon, E. Z. Mequanint, M. Cristani, H. Hung, M. Pelillo, and V. Murino. A game-theoretic probabilistic approach for detecting conversational groups. In *Asian Conference on Computer Vision*, pages 658–675. Springer, 2014.

[67] D. Vasquez, P. Stein, J. Rios-Martinez, A. Escobedo, A. Spalanzani, and C. Laugier. Human aware navigation for assistive robotics. In *Experimental Robotics*, pages 449–462. Springer, 2013.

[68] R. Vaughan, N. Sumpter, J. Henderson, A. Frost, and S. Cameron. Experiments in automatic flock control. *Robotics and Autonomous Systems*, 31(1-2):109–117, 2000.

[69] M. Vázquez, A. Steinfeld, and S. E. Hudson. Maintaining awareness of the focus of attention of a conversation: A robot-centric reinforcement learning approach. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pages 36–43. IEEE, 2016.

[70] A. W. Vieira, P. L. J. Drews, and M. F. M. Campos. Spatial Density Patterns for Efficient Change Detection in 3D Environment for Autonomous Surveillance Robots. *IEEE Transactions on Automation Science and Engineering*, 11(3):766–774, July 2014. ISSN 1545-5955. doi: 10.1109/TASE.2013.2294851.

[71] A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing: Survey of an emerging domain. *Image and vision computing*, 27(12):1743–1759, 2009.

[72] K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg. Who are you with and where are you going? In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1345–1352. IEEE, 2011.

[73] J. E. Young, Y. Kamiyama, J. Reichenbach, T. Igarashi, and E. Sharlin. How to walk a robot: A dog-leash human-robot interface. In *RO-MAN, 2011 IEEE*, pages 376–382. IEEE, 2011.

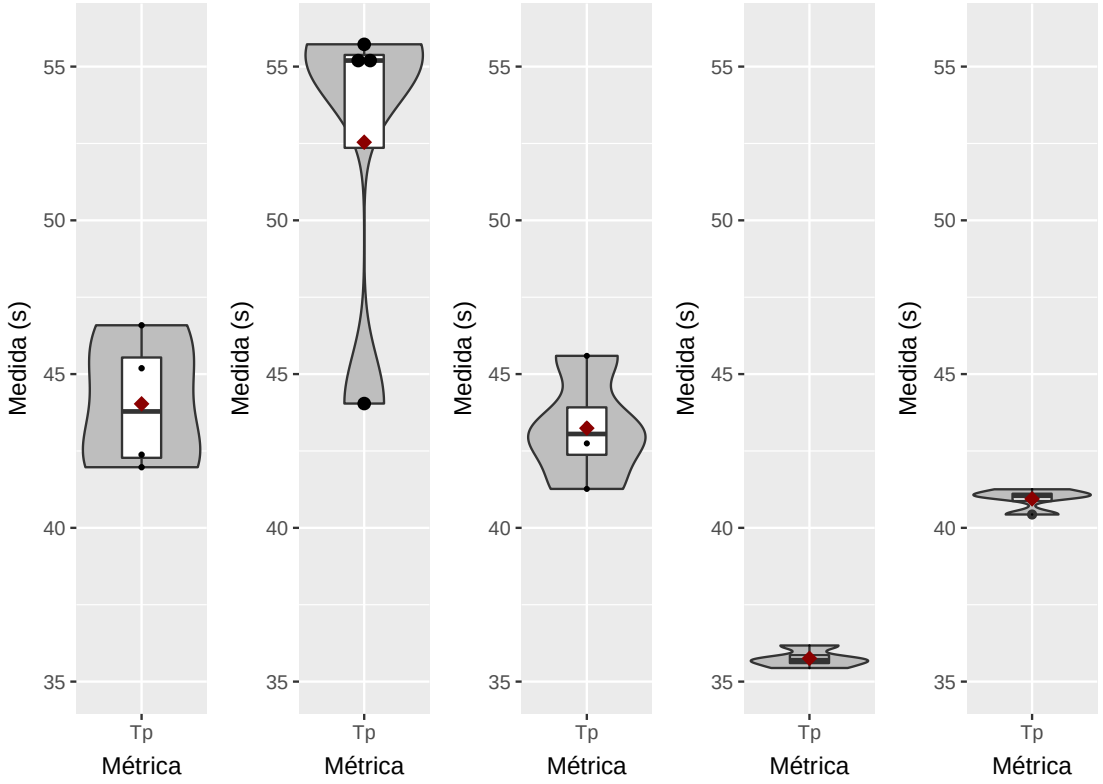
[74] M. Zanotto, L. Bazzani, M. Cristani, and V. Murino. Online bayesian nonparametrics for group detection. In *Proc. of BMVC*, 2012.

[75] G. Zen, B. Lepri, E. Ricci, and O. Lanz. Space speaks: towards socially and personality aware visual surveillance. In *Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis*, pages 37–42. ACM, 2010.

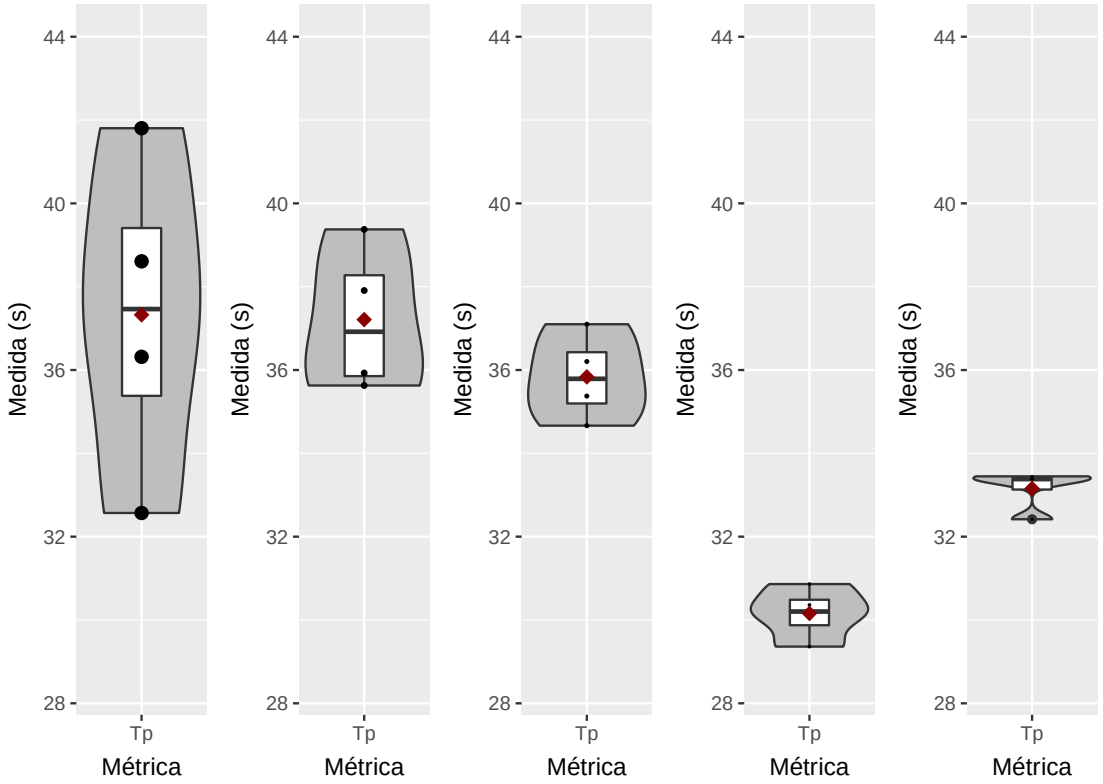
Apêndice A

Métricas de avaliação do sistema de navegação social

Este apêndice contém a representação gráfica dos resultados apresentados nas Tabelas 4.7 e 4.8, referentes ao experimento para avaliação da representação das fronteiras de conforto social. A execução dos experimentos foi dividida em cinco fases distintas, tanto para a configuração de ambiente incluindo os indivíduos singulares quanto para os grupos de interação. Cada coluna apresentada nos gráficos, refere-se à uma das cinco fases de execução do experimento, onde a primeira coluna equivale aos resultados do algoritmo *baseline* Potential fields e as demais colunas aos resultados do método de detecção do espaço social coletivo (CSS) considerando as zonas Íntima, Pessoal, Social e Pública respectivamente.

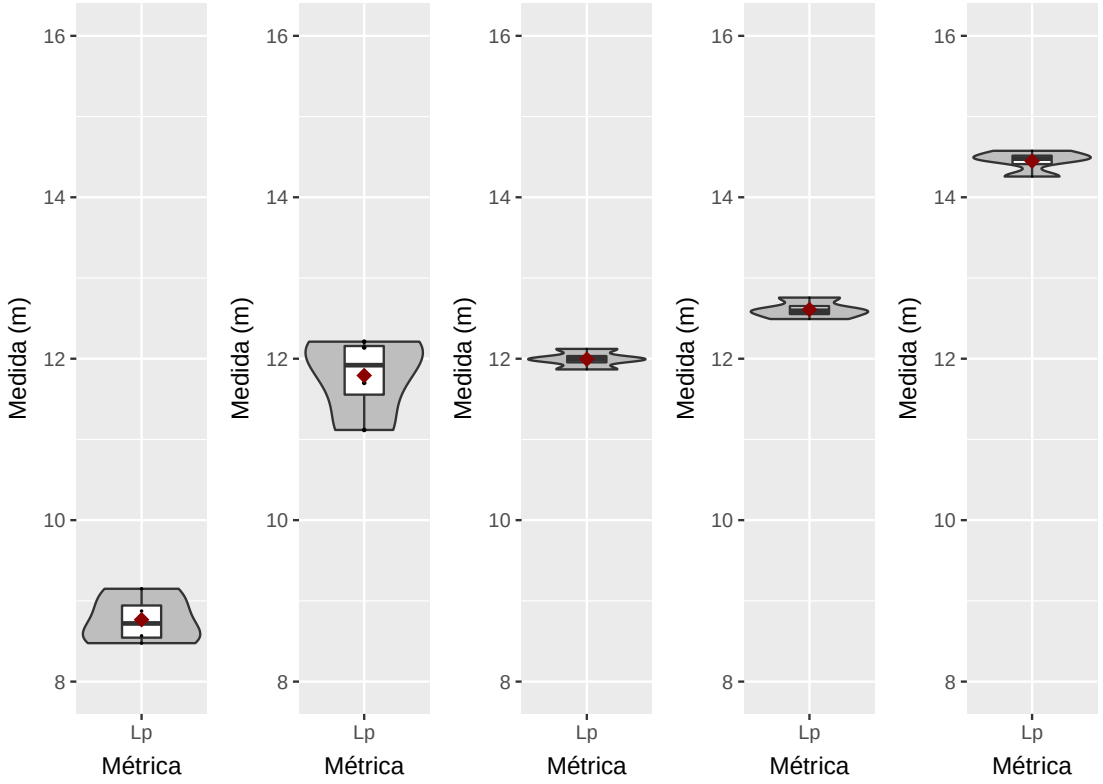


(a)

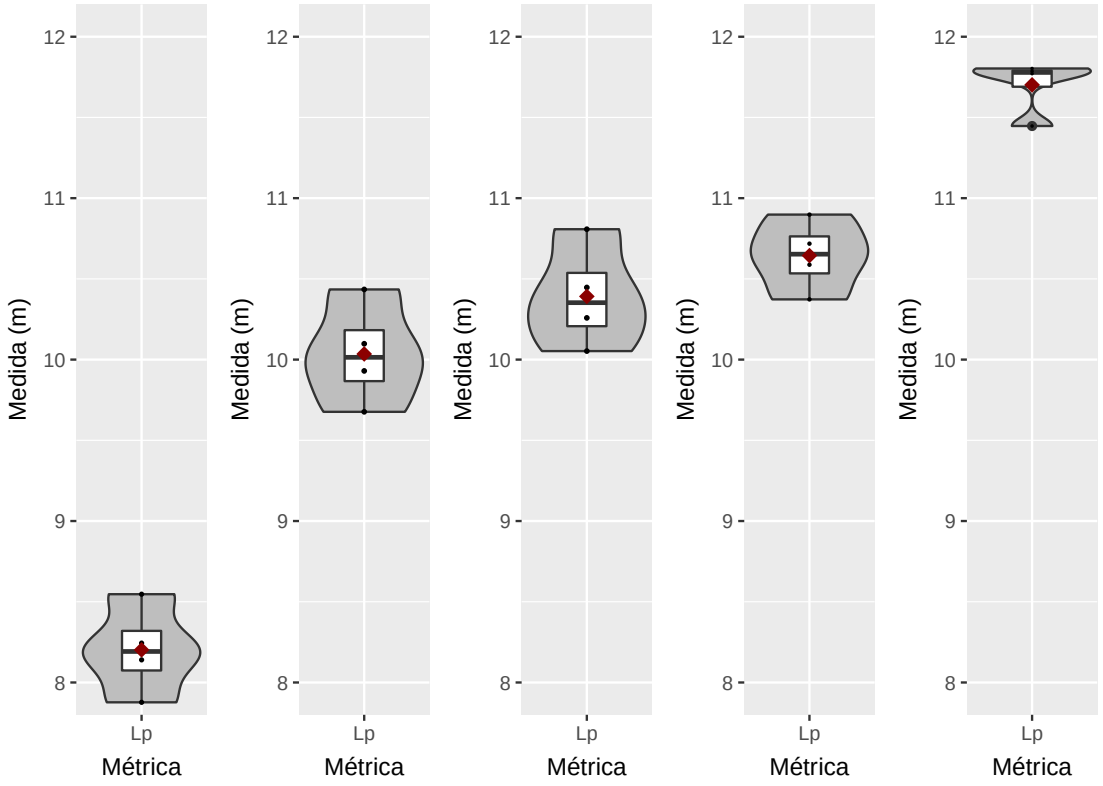


(b)

Figura A.1: Resultados relativos ao tempo necessário para atingir o objetivo (T_p), para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação.

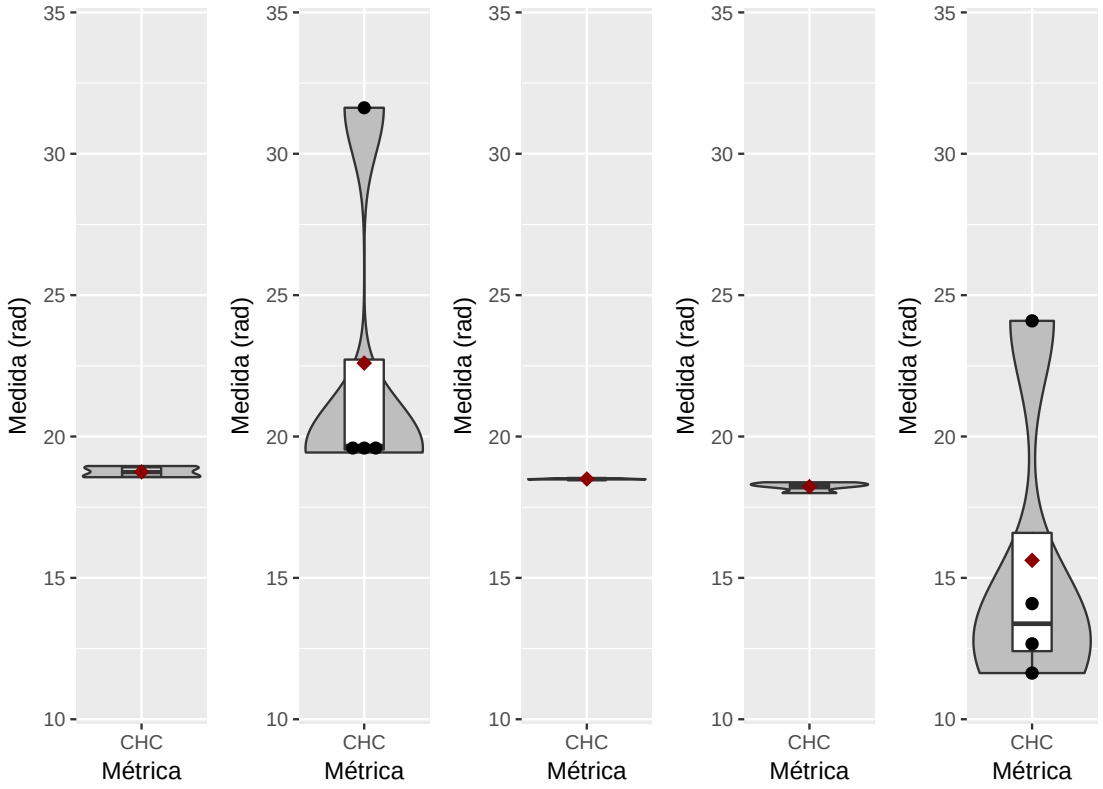


(a)

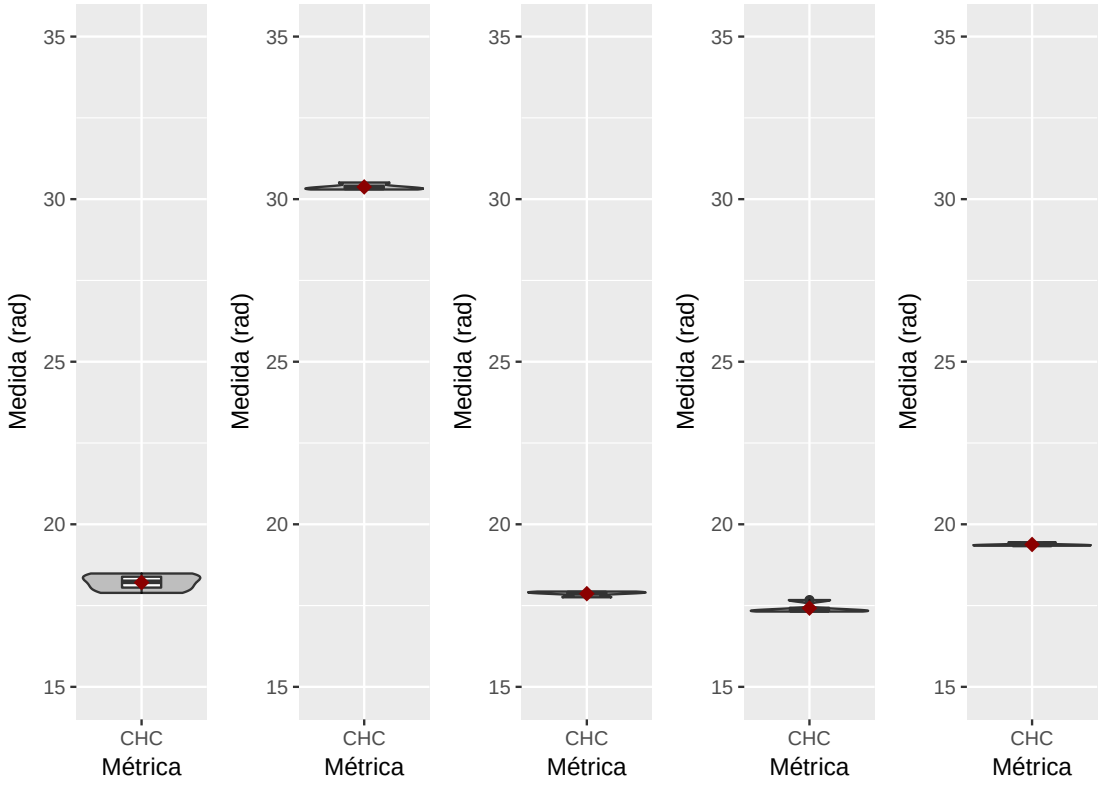


(b)

Figura A.2: Resultados relativos ao comprimento do percurso (L_p), para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação.

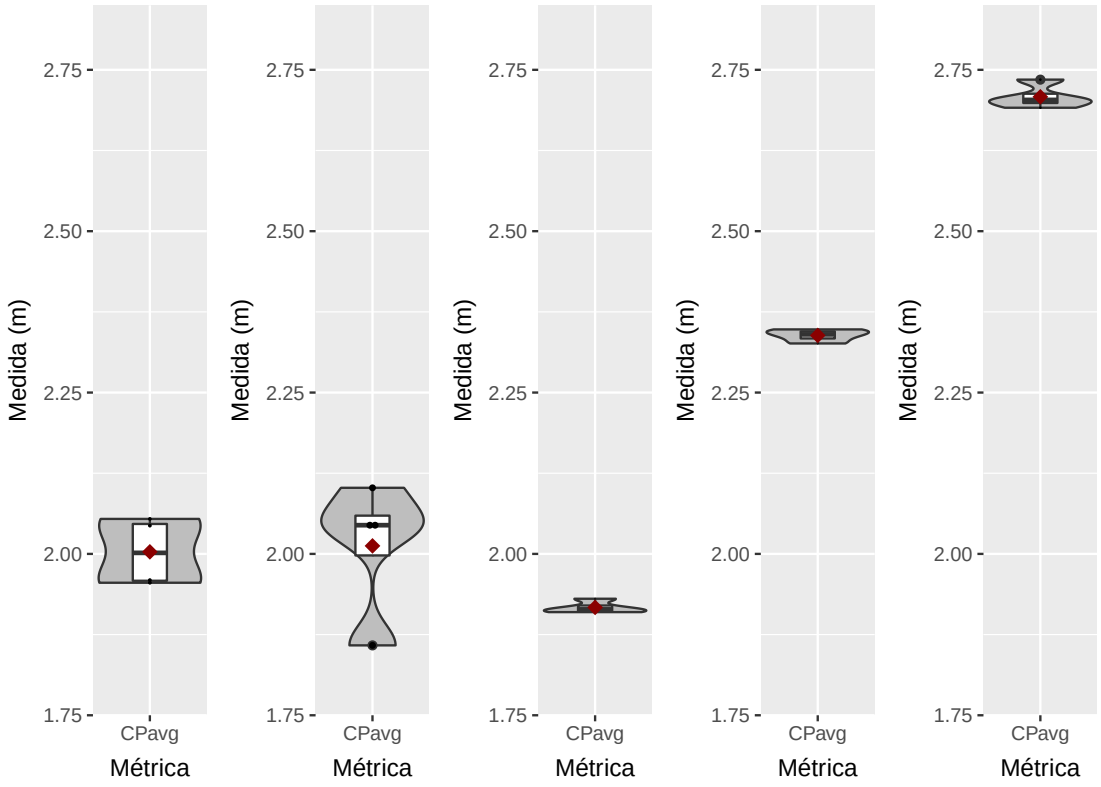


(a)

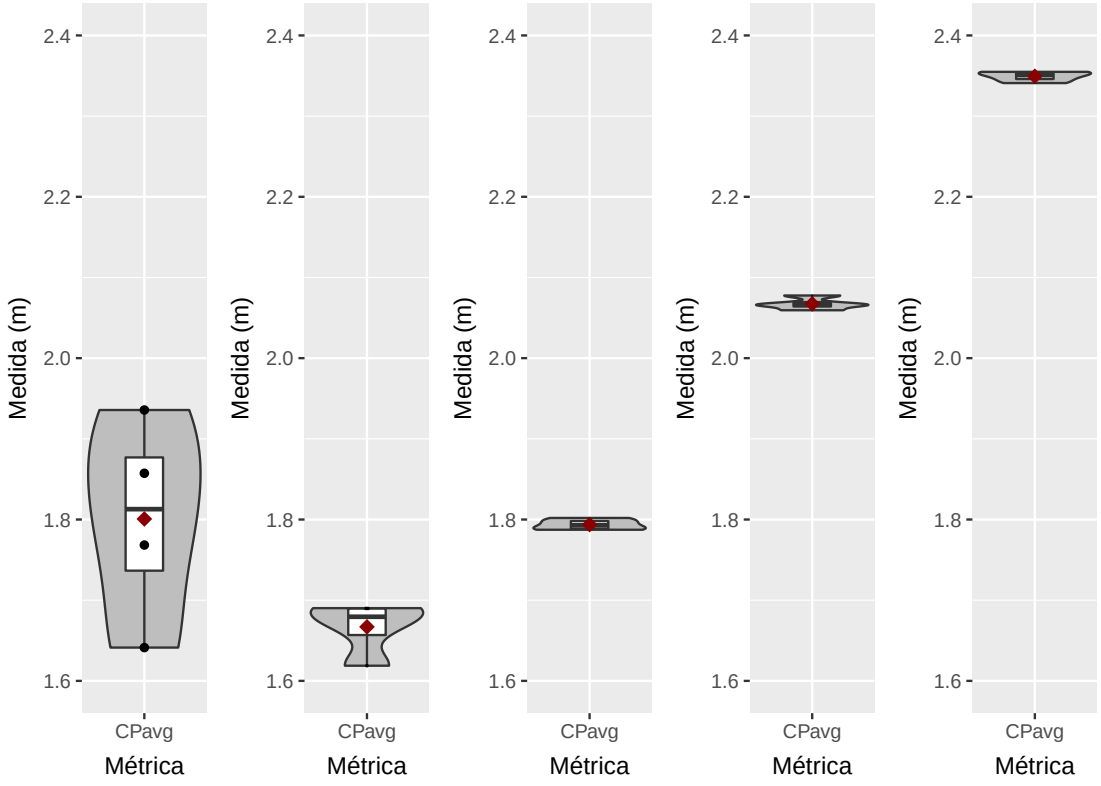


(b)

Figura A.3: Resultados relativos a alterações cumulativas de orientação (CHC), para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação.

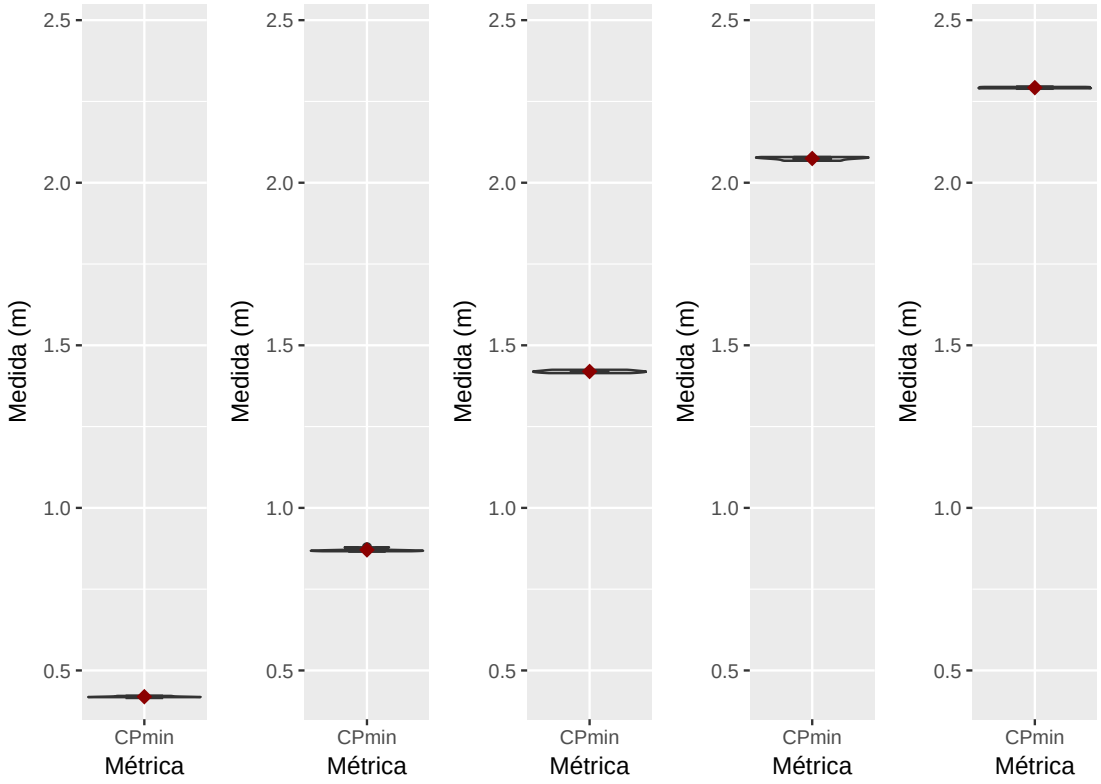


(a)

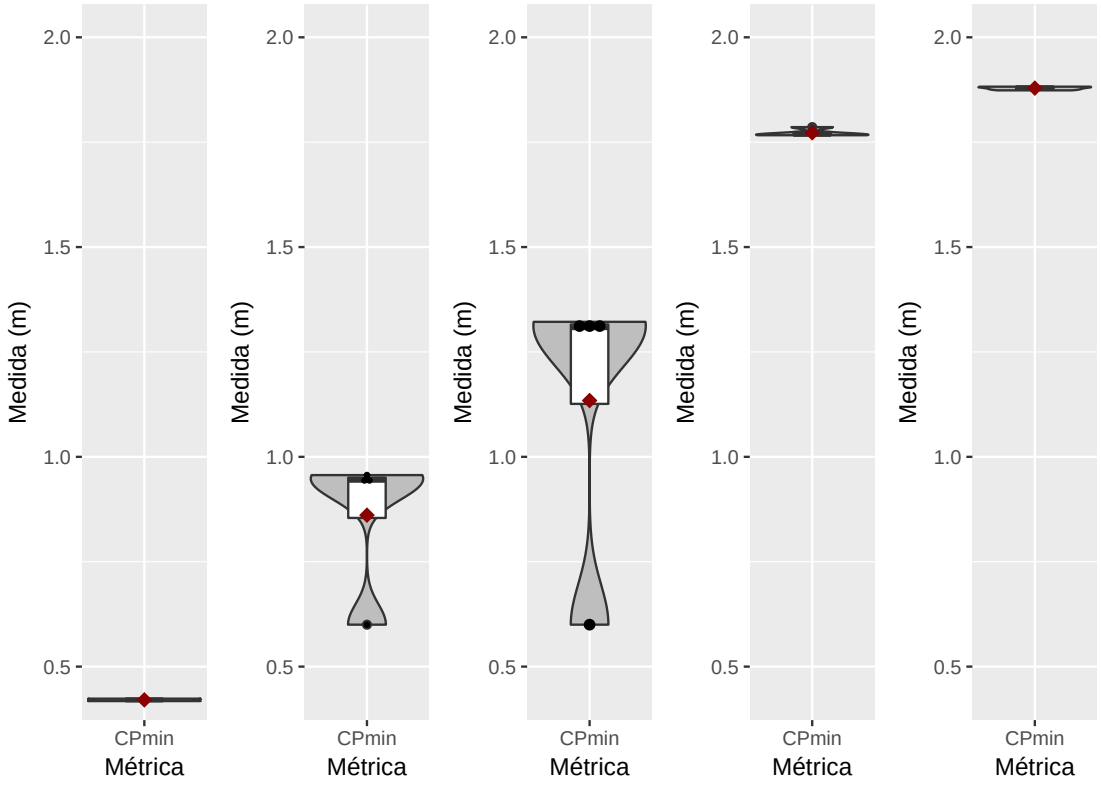


(b)

Figura A.4: Resultados relativos a distância média para a pessoa mais próxima (CP_{avg}), para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação.

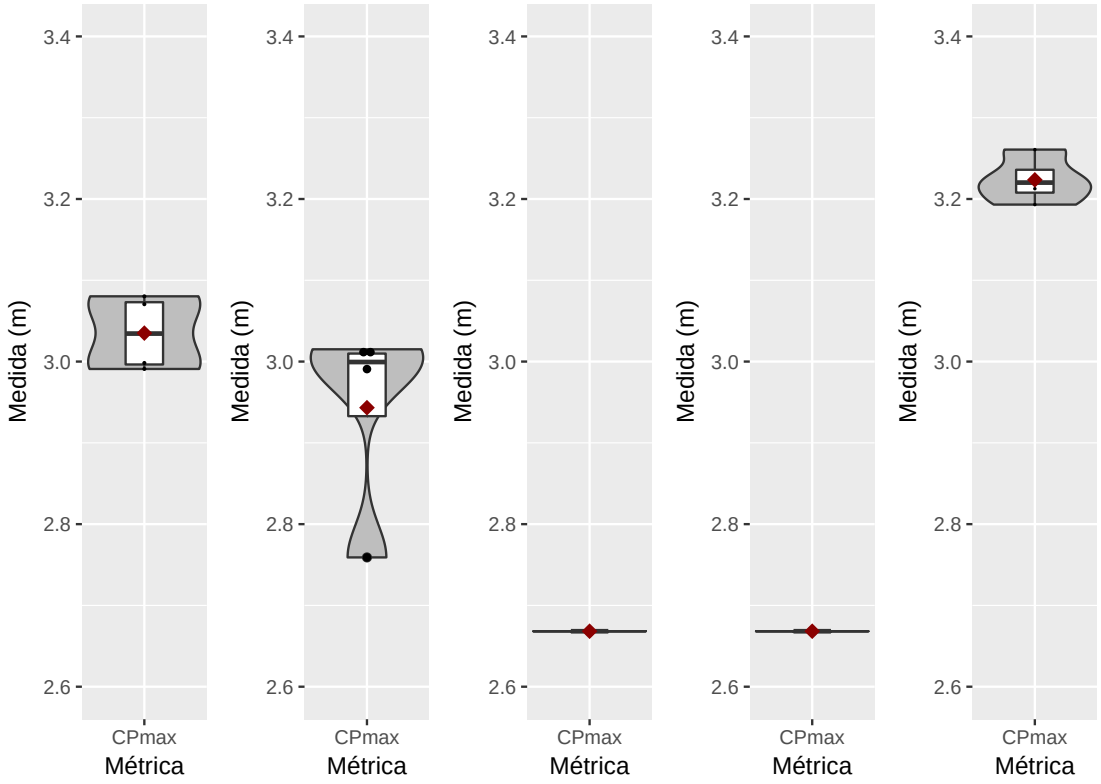


(a)

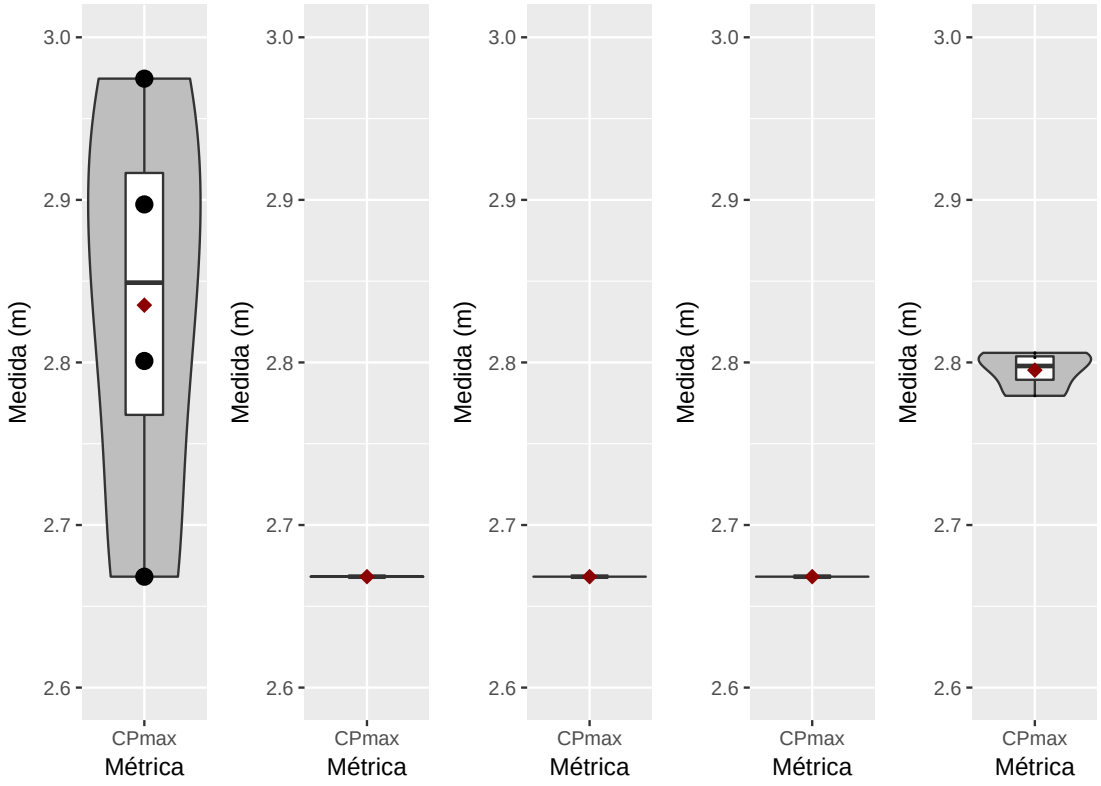


(b)

Figura A.5: Resultados relativos a distância mínima para a pessoa mais próxima (CP_{min}), para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação.

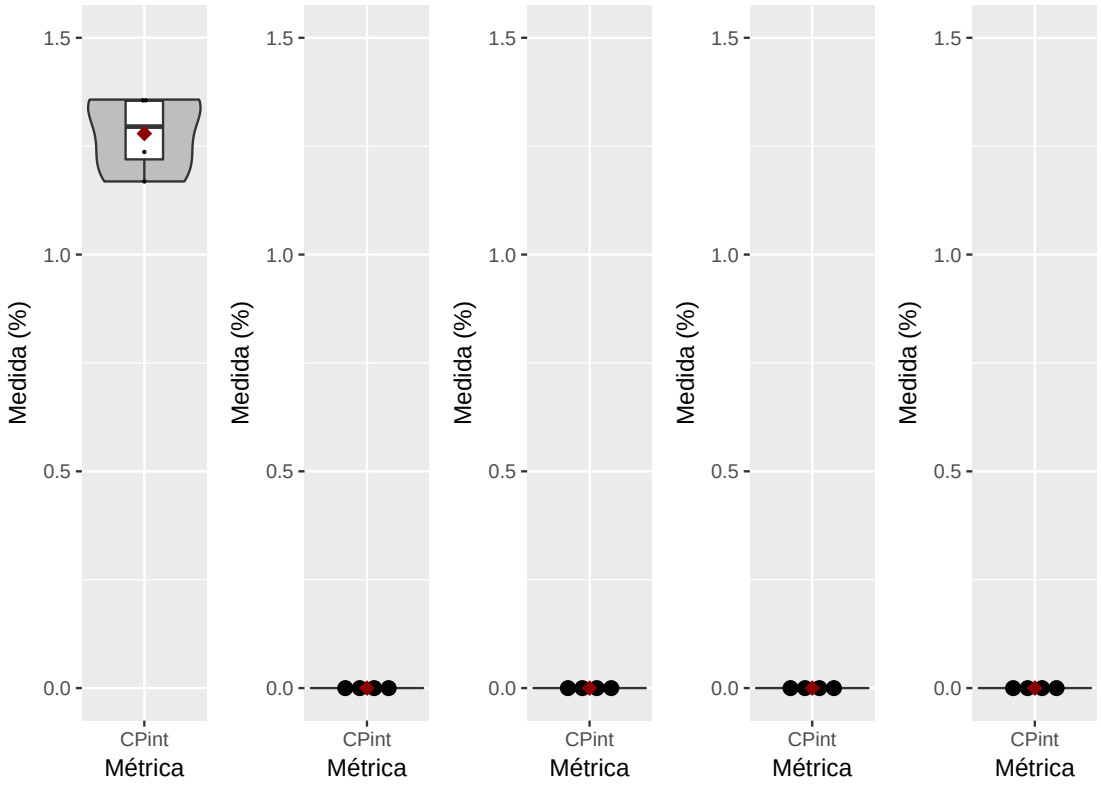


(a)

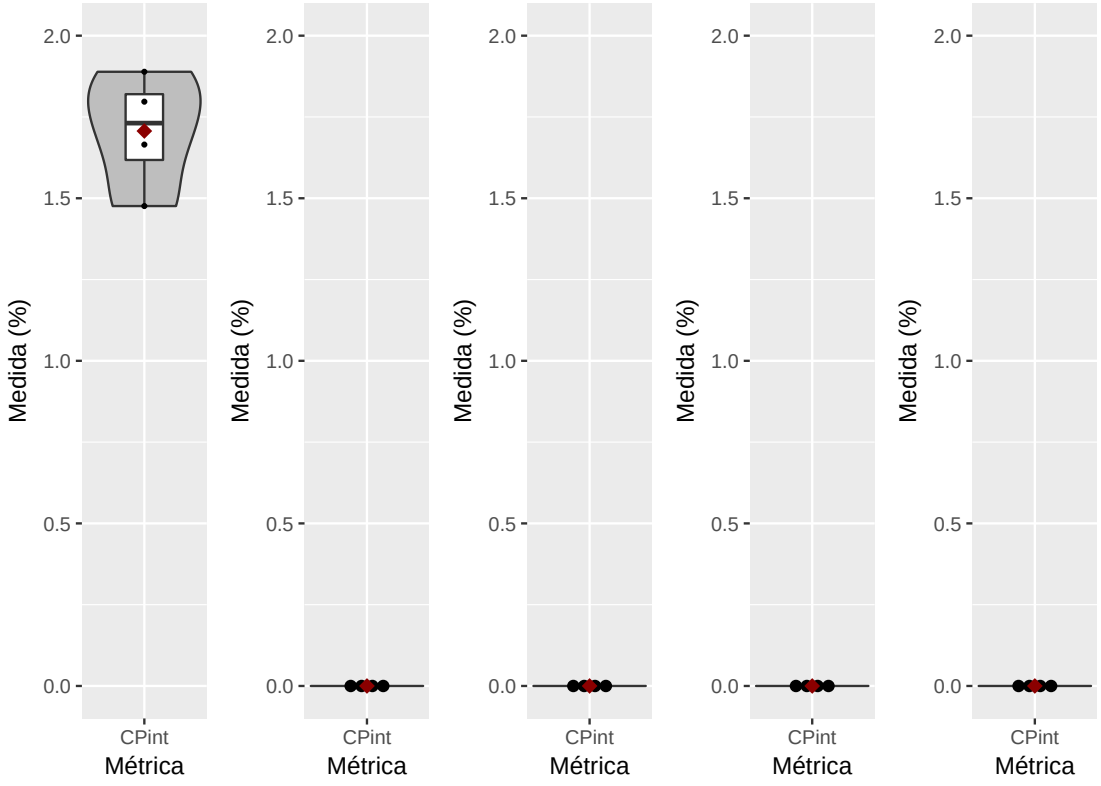


(b)

Figura A.6: Resultados relativos a distância máxima para a pessoa mais próxima (CP_{max}), para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação.

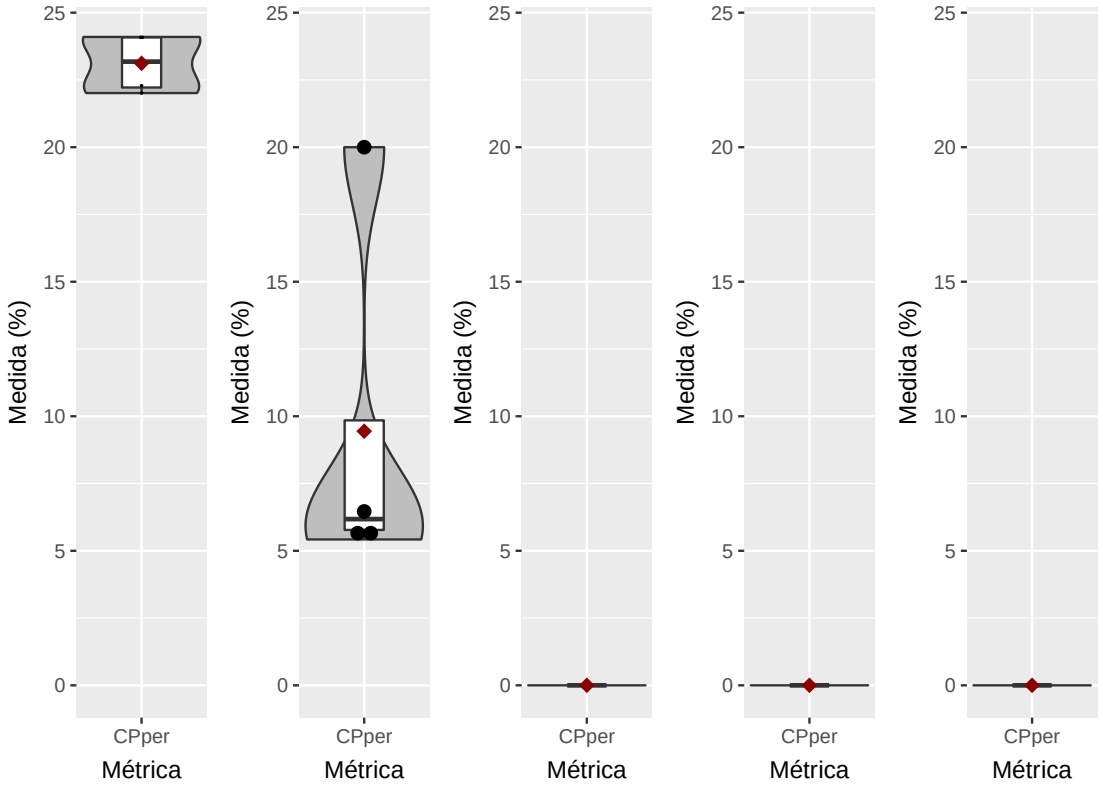


(a)

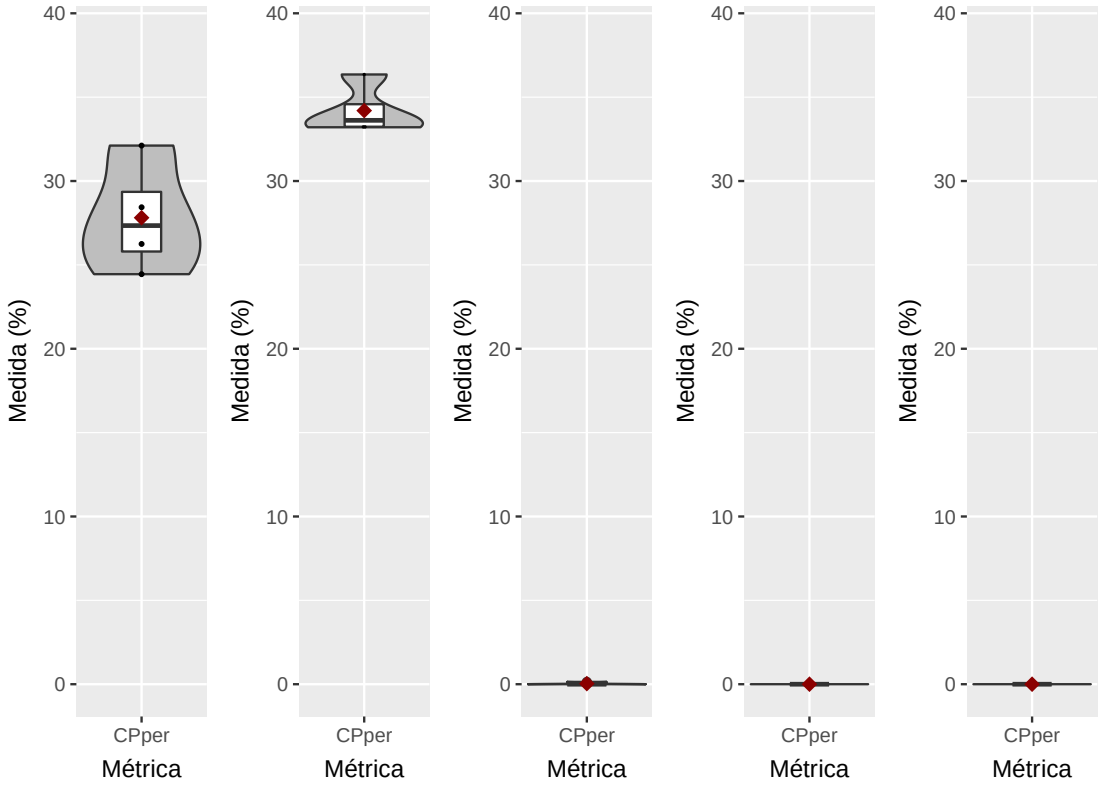


(b)

Figura A.7: Resultados relativos a intrusões de espaço pessoal (CP_{prox}) na distância íntima, para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação.

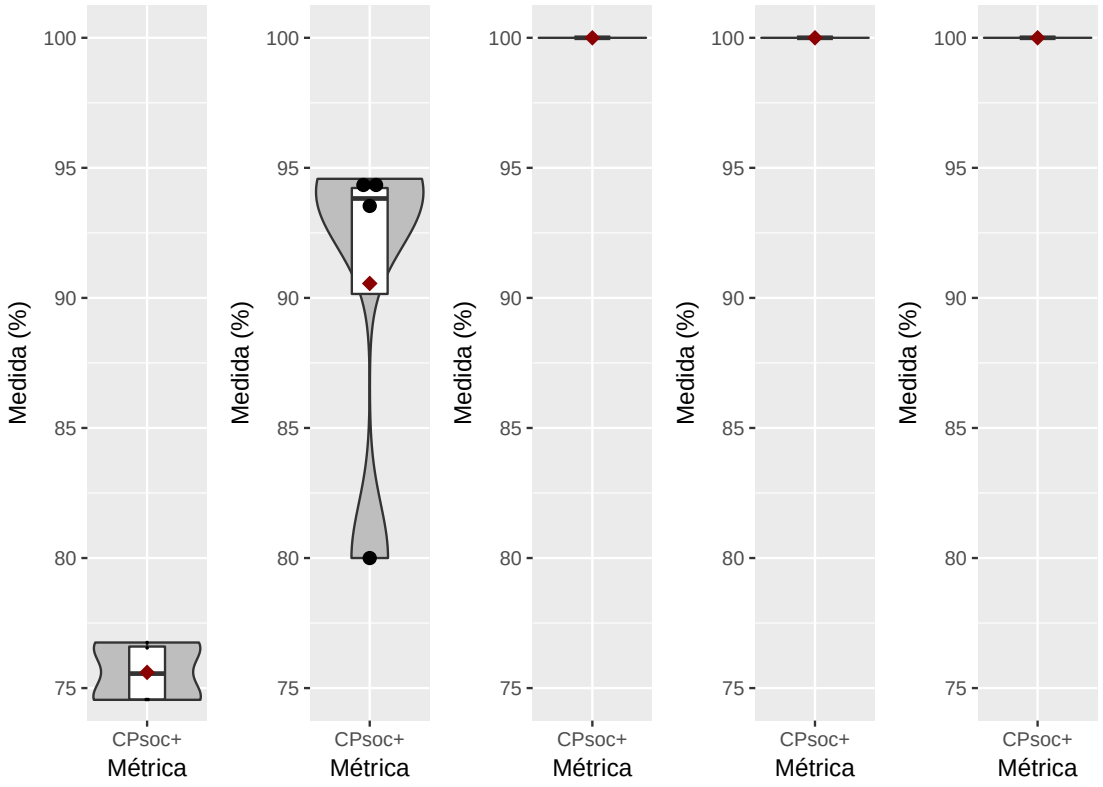


(a)

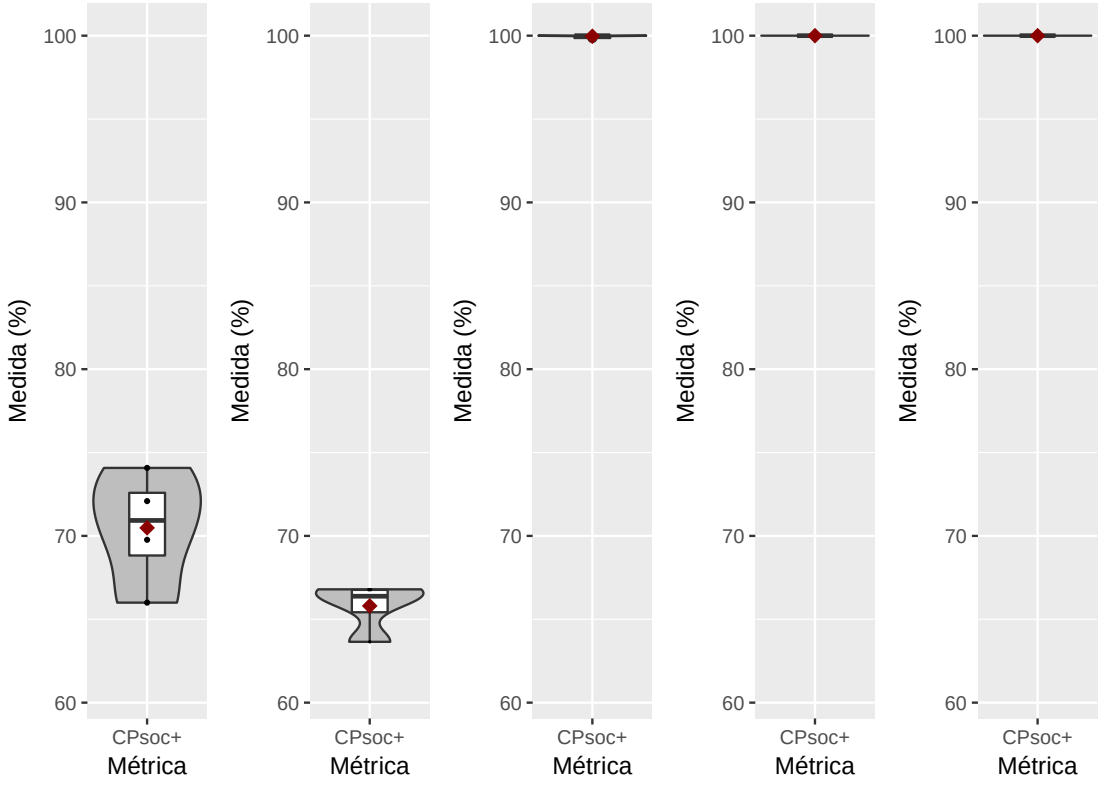


(b)

Figura A.8: Resultados relativos a intrusões de espaço pessoal (CP_{prox}) na distância pessoal, para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação.



(a)



(b)

Figura A.9: Resultados relativos a intrusões de espaço pessoal (CP_{prox}) na distância social+pública, para os algoritmos Potential Fields e CSS nas zonas Íntima, Pessoal, Social e Pública respectivamente. Considerando em (a) um par de indivíduos e em (b) um grupo de interação.