

# AVALIAÇÃO DE TÉCNICAS PARA DETECÇÃO DE PEDESTRES USANDO LASER E CÂMERA VISANDO A FUSÃO SENSORIAL

NATÁLIA COSSE BATISTA\*<sup>†</sup>, GUILHERME A. S. PEREIRA\*

\*Programa de Pós-Graduação em Engenharia Elétrica - Universidade Federal de Minas Gerais  
Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brasil

<sup>†</sup>Universidade Federal de Itajubá,  
Rua Irmã Ivone Drummond, 200, 35903-087, Itabira, MG, Brasil

Email: nataliabatista@unifei.edu.br, gpereira@ufmg.br

**Abstract**— One of the most essential tasks attributed to autonomous land vehicles is pedestrian detection, which requires advanced sensing and perception of the environment. In recent years, 2D laser range finder and monocular cameras are on focus of many researches aiming at fusing data from both sensors. However, these sensors are limited and can fail in detecting important information of the scene depending on the way they are combined. In this work, we study and make a preliminary evaluation of pedestrian detection techniques and analyse the behavior of laser sensor for future sensor fusion.

**Keywords**— Pedestrian detection, sensor fusion, autonomous land vehicles, computer vision.

**Resumo**— A detecção de pedestres é uma das tarefas mais essenciais para a autonomia de veículos autônomos e envolve uma avançada percepção do ambiente por meio de sensores. Atualmente, dentre os sensores mais utilizados na literatura para este fim estão o laser 2D e a câmera monocular, combinados por meio de técnicas de fusão sensorial. Entretanto, estes sensores possuem limitações e dependendo da forma como forem combinados, podem deixar de capturar importantes informações sobre a cena. Este trabalho tem o objetivo de realizar o estudo e avaliação preliminar de técnicas de detecção de pedestres utilizando laser e câmera e analisar o comportamento do sensor a laser visando a fusão sensorial.

**Keywords**— Detecção de pedestres, fusão sensorial, veículos autônomos, visão computacional.

## 1 Introdução

Para os veículos autônomos operarem de forma segura sem intervenção humana em ambientes urbanos, são necessários níveis cada vez maiores de autonomia e robustez. Uma das tarefas mais essenciais para autonomia de robôs que operam em ambientes urbanos é a detecção de obstáculos em rota de colisão, principalmente pessoas, de forma a evitar situações de risco durante sua navegação. Essa tarefa envolve uma avançada percepção do entorno por meio de sensores.

Vários tipos de sensores são utilizados para percepção por robôs autônomos, tais como laser, câmera para luz visível, radares e câmera para infravermelho. Esses sensores podem ser divididos em duas categorias: passivos e ativos. Conforme (Bertozzi et al., 2002), os sensores passivos, como aqueles baseados em visão, apresentam vantagens em relação aos sensores ativos (laser, sonar e radar). Os sensores baseados em visão possibilitam obter informações visuais ricas em detalhes para aplicações no reconhecimento de objetos e a classificação de texturas. Em tarefas como a detecção de pedestres, conforme (Broggi et al., 2009), a visão é a melhor tecnologia para detectar a presença de pedestres parcialmente oclusos, além do preço de aquisição mais baixo. A desvantagem dos sensores baseados em visão é que são menos robustos à neblina, falta de luminosidade ou condições de luz solar direta.

Por outro lado, os sensores ativos têm a vantagem de poderem medir movimento de uma

forma mais direta e com menor quantidade de dados que os sensores baseados em visão, resultando em um processamento mais rápido. Este tipo de sensor é robusto às condições de iluminação e pode operar a longas distâncias, além da possibilidade de medir distância de objetos com maior precisão que os sensores passivos. Entretanto, podem sofrer interferência de outros sensores do mesmo tipo, o que prejudica sua confiabilidade no uso em massa. Outra desvantagem é que o laser não provê cor nem textura dos objetos, dificultando a distinção entre eles.

A fusão de múltiplos sensores permite combinar as vantagens tecnológicas de diversos sensores para compensar suas limitações. Além disso, a redundância de informação que pode ser obtida com a fusão deve ser explorada para maximizar a confiabilidade da percepção ou para reduzir o espaço de busca em outros sensores, consequentemente diminuindo o tempo de processamento. A fusão sensorial é realizada em diversas aplicações em robótica, tais como navegação autônoma, detecção de pedestres, classificação de objetos e localização (Antunes et al., 2012).

O objetivo deste artigo é avaliar de forma preliminar técnicas de detecção de pedestres utilizando um sensor passivo (câmera) e um sensor ativo (laser), visando combinar suas informações para realizar a fusão sensorial. A próxima seção apresenta trabalhos relacionados à detecção de pessoas. A Seção 3 explica como foram avaliadas as técnicas escolhidas. Os experimentos e seus resultados são apresentados na Seção 4. Finalmente, a

Seção 5 relata as conclusões e trabalhos futuros.

## 2 Sistemas e algoritmos para detecção de pedestres

A detecção de pessoas tem sido o foco de diversos trabalhos recentes em robótica devido à sua aplicação em robôs que interagem com pessoas ou veículos que devem trafegar em ambientes com pedestres. Os sensores mais comuns nesses trabalhos são o laser e a câmera, que são detalhados nas subseções a seguir.

### 2.1 Laser

O laser é um sensor de proximidade que opera medindo o tempo de ida e volta de pulsos de luz, que é diretamente proporcional à distância entre o scanner e o objeto atingido. Como resultado, obtém-se pontos no plano de varredura do laser, tipicamente 2D, para cada ângulo (geralmente de 0 a 180 graus). Os pontos gerados, em coordenadas polares, são representados por sua distância e ângulo em relação ao referencial do laser. O laser apresenta campo de visão amplo e frequências altas de leitura e é utilizado em vários trabalhos recentes da literatura para detectar pernas humanas.

Na detecção de pessoas com laser há abordagens que realizam a extração de características geométricas tais como tamanho da borda, convexidade, número de pontos, linhas, quinas, etc. As características são utilizadas para treinar classificadores ou para determinar limiares, como em (Spinello and Siegwart, 2008), (Premebida et al., 2009), (Varvadoukas et al., 2012) e (Mozos et al., 2010). Por outro lado, as abordagens que não são baseadas em características utilizam busca por mínimos locais, detecção baseada em movimentos ou subtração do fundo (Cui et al., 2005) e casamento de padrões (Oliveira et al., 2010), (Pereira et al., 2013), (Bellotto and Hu, 2009).

Neste trabalho serão avaliados os algoritmos de (Bellotto and Hu, 2009) e (Spinello and Siegwart, 2008), por estarem livremente disponíveis na internet e também por apresentarem resultados satisfatórios comprovados em outros trabalhos como (Pereira et al., 2013) e (Varvadoukas et al., 2012).

O trabalho de (Bellotto and Hu, 2009) realiza a detecção de pernas em ambiente interno, identificando padrões típicos relativos à posição das pernas, tais como pernas juntas, pernas abertas ou pernas em posição de caminhar. Já no trabalho de (Spinello and Siegwart, 2008) são computadas características geométricas da leitura do laser que são utilizadas na etapa de treinamento com uma técnica de aprendizado supervisionado baseado em cascatas de SVMs (*Support Vector Machines*). Ambos os métodos extraem as características necessárias para a detecção de ape-

nas uma leitura do laser, independentemente do movimento dos objetos da cena.

### 2.2 Câmera

A câmera, que possui um sensor de imagem digital, capta a luminosidade da cena que é projetada sobre ele, gerando uma imagem bidimensional formada por uma matriz de pontos, onde cada ponto é um pixel.

A detecção de pessoas em imagens é uma tarefa bastante desafiadora devido à imensa variedade de aparências e poses em que as pessoas podem ser registradas. Há dois algoritmos bastante comuns: detecção de faces utilizando métodos baseados no trabalho de Viola and Jones (2001) e detecção de pessoas utilizando características baseadas em histogramas de gradientes orientados (HOG) (Dalal and Triggs, 2005). Como a detecção de faces é mais adequada para ser utilizada em aplicações nas quais as pessoas, além de estarem próximas aos sensores, estão voltadas para a frente e geralmente paradas para interagir com um robô, neste trabalho será avaliada a detecção de pessoas baseadas em características HOG, pois o objetivo é que nos experimentos as pessoas caminhem livremente, tais como pedestres na rua.

Os descritores HOG, introduzidos no trabalho (Dalal and Triggs, 2005), têm como idéia básica que a aparência local e a forma dos objetos podem ser caracterizados pela distribuição local de gradientes de intensidade ou direção das bordas. Os HOG são similares aos descritores SIFT, mas são computados em um *grid* denso com células uniformemente espaçadas. A imagem é analisada em múltiplas escalas por meio de janelas, que são divididas em regiões pequenas (as células), cada uma tendo um histograma local de direções dos gradientes. Os gradientes são computados utilizando suavização gaussiana seguida de várias máscaras de derivações discretas. Para cada janela é obtido um descritor que é classificado por uma SVM linear para determinar se existe ou não a presença de pedestres. Ao final da varredura da imagem em múltiplas escalas, cada objeto obtém um grupo denso de janelas de detecção, sendo necessário unilas em uma janela de identificação final. Esses descritores são utilizados relacionados à técnicas de aprendizado supervisionado em (Spinello and Siegwart, 2008), (Varvadoukas et al., 2012), (Premebida et al., 2009) e (Oliveira et al., 2010).

### 2.3 Fusão dos sensores

Existem vários trabalhos de detecção de pedestres que utilizam simultaneamente dados visuais e lasers. O objetivo deste trabalho é avaliar algoritmos para detecção de pedestres e analisar o comportamento do sensor de laser, estudando suas limitações e contribuições para a fusão sensorial. De acordo com (Oliveira et al., 2010) as duas abor-

dagens mais comuns para a fusão de laser e câmera são (Figura 1):

1. Integração de características dos sensores (fusão a nível de características) ou dos resultados de classificadores (fusão a nível de classificadores).
2. Uma região de interesse ou ROI (*region of interest*) é determinada por segmentação dos dados do laser e um classificador baseado em visão é utilizado para rotular a ROI projetada.

Dentro da primeira abordagem de fusão de laser e câmera, pode-se citar os trabalhos (Cui et al., 2005), (Oliveira et al., 2010), (Bellotto and Hu, 2009), (Araújo et al., 2011) e (Pereira et al., 2013). Nesses trabalhos, as informações complementares e redundantes dos dois sensores são exploradas para maximizar os níveis de confiança e inferência da detecção de pessoas. Na segunda abordagem de fusão, os dados provenientes da leitura do laser são determinantes para que a imagem seja avaliada, ou seja, em situações nas quais o laser falhar na detecção de uma pessoa não haverá detecção na imagem pois não haverá ROI. Pode-se citar os trabalhos (Premebida et al., 2009), (Broggi et al., 2009), (Wu et al., 2011) e (Spinello and Siegwart, 2008) que baseiam-se na segunda abordagem.

O laser pode falhar por diversas razões que incluem baixa refletividade de roupas pretas a grandes distâncias (Oliveira et al., 2010), reflexões e efeitos da luz direta do sol (Spinello and Siegwart, 2008). Oclusões também podem causar falhas no laser, por exemplo quando uma pessoa está atrás de uma caixa de tamanho médio, a câmera consegue visualizar a parte superior do seu corpo, mas um laser posicionado na altura das pernas somente detectará a caixa.

Dessa forma, as técnicas de fusão baseadas na segunda abordagem tendem a ter maior número de falsos negativos quando avaliadas em relação à quantidade de pessoas realmente existentes na cena e não apenas aquelas computadas por meio do laser. Além disso, como as ROIs são obtidas pela projeção dos segmentos do laser no plano da imagem, são passíveis de erros de imprecisão da calibração dos sensores, irregularidades da via e vibrações do veículo (Premebida et al., 2009).

### 3 Metodologia experimental

Esta seção apresenta a metodologia utilizada na avaliação preliminar de técnicas para detecção de pedestres. O ambiente experimental para avaliação consiste em um laser e uma câmera fixos, posicionados a aproximadamente 40 cm e 150 cm do chão, respectivamente. O laser estava aproximadamente alinhado à câmera em relação ao eixo  $x$  e deslocado de 70 cm à frente. A rotação da

câmera em relação ao laser fornecido pela calibração (explicada a seguir) é dada pelo vetor:  $R = \{18^\circ; -22^\circ; -4^\circ\}$ .

O laser LMS291 - S05, fabricado pela SICK, possui ângulo de visão de  $180^\circ$ , alcance configurado para aproximadamente 8 metros e resolução angular de  $0,5^\circ$ . A câmera utilizada foi a Bumblebee2, fabricada pela Point Grey Research. Embora seja uma câmera estéreo, neste trabalho foram utilizadas apenas as imagens geradas pela câmera esquerda. Cada imagem gerada tem tamanho  $640 \times 480$  pixels. Para a realização dos experimentos, utilizou-se o ROS (*Robot Operating System*), um *software* que provê bibliotecas e ferramentas para aplicações em robótica e por meio do qual é possível acessar os dados da câmera e do laser.

A metodologia experimental pode ser dividida em três etapas: calibração dos sensores, obtenção e anotação da base de imagens e da base do laser e realização dos experimentos. As próximas subseções explicam cada etapa.

#### 3.1 Calibração dos sensores

A calibração é uma etapa essencial do trabalho pois sem ela não é possível relacionar as informações de sensores diferentes, já que é necessário conhecer a posição do laser em relação à câmera e vice-versa. Assim, o problema da calibração consiste em estimar os valores dos parâmetros intrínsecos da câmera e também os parâmetros extrínsecos em relação ao referencial do laser.

Para a calibração câmera-laser foi escolhido o pacote Matlab disponibilizado por (Kassir and Peynot, 2010). Inicialmente, é realizada a calibração da câmera, em que os parâmetros intrínsecos da câmera são calculados, bem como a transformação rígida entre a câmera e um padrão de calibração. A transformação é necessária para a calibração câmera-laser, que usa os pontos originados do padrão de calibração que aparecem na leitura do laser para encontrar a transformação rígida do laser para a câmera. O padrão de calibração utilizado é o tradicional tabuleiro de xadrez usado para calibrar câmeras. Como resultado da calibração, obtém-se o vetor de translação e a matriz de rotação que representam a orientação e posição da câmera em relação ao laser.

#### 3.2 Anotação da base de dados

Para os experimentos foram coletados dados de câmera e laser dentro de um ambiente interno (prédio da Engenharia/UFMG), gerando a base do laser com 2719 leituras e a base de imagens com 1693 quadros.

O *ground-truth* da base de dados indica, tanto para os dados do laser quanto da câmera, a localização dos pedestres. É utilizado para avaliação de algoritmos aplicados à base de dados e a maioria dos trabalhos revisados na Seção 2 realiza a anotação manual das imagens ou das leituras

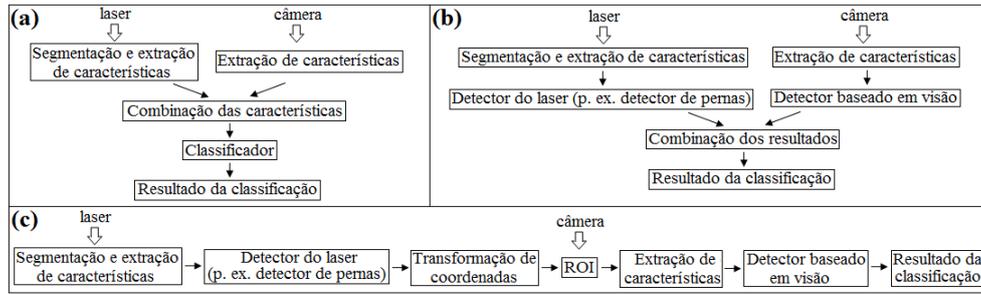


Figura 1: Abordagem 1: (a) Fusão a nível de características e (b) Fusão a nível de classificadores. Abordagem 2: (c) Algoritmos de visão e do laser executados de forma sequencial.

do laser (como Bellotto and Hu (2009)). Em (Premebida et al., 2009), os dados do laser são manualmente validados e a base de imagens foi automaticamente extraída pelo mapeamento das ROI (regiões de interesse) detectadas pelo laser, estando sujeito aos erros do sistema do laser. Em (Oliveira et al., 2010) ocorre o processo inverso.

Neste trabalho realizou-se a anotação do laser automaticamente, de forma similar a Mozos et al. (2010). Para rotular os pontos do laser como positivos ou negativos, detectou-se inicialmente fundo estático do ambiente. Os pontos foram rotulados por meio da diferença entre o fundo e a leitura, sendo que foram considerados apenas a região dentro do campo de visão da câmera. As imagens da câmera foram anotadas manualmente, gerando os vértices dos retângulos que delimitam as pessoas. A anotação das imagens inclui todas as pessoas visíveis, exceto as mais distantes (aproximadamente 20 metros), que são representadas por poucos pixels na imagem e cuja visualização demanda uma maior atenção.

### 3.3 Experimentos

Os experimentos realizados para estudo dos sensores são listados a seguir:

1. Comparação do *ground-truth* das leituras do laser e das imagens, simulando a detecção com base na anotação do laser (cem por cento de acerto nos dados laser) e avaliando o resultado em relação à anotação das imagens, por meio da projeção dos pontos detectados nas imagens utilizando as informações da calibração. O objetivo é mostrar que o número de pessoas anotadas nas leituras do laser, mesmo considerando apenas o campo de visão da câmera, difere do número de pessoas anotadas nas imagens. Neste experimento e nos seguintes que foram avaliados em relação à anotação das imagens, foi adicionada uma tolerância de 15 pixels antes e depois do retângulo demarcado na anotação da imagem devido a erros de sincronização, movimento das pessoas e calibração.

2. Detecção de pernas no laser utilizando o algoritmo de Spinello and Siegwart (2008) avaliada com base na anotação do laser e também em relação à anotação das imagens. Como o algoritmo

requer treinamento, este foi realizado utilizando-se a base fornecida pelos mesmos autores, porém subamostrada de  $1^\circ$  para  $0,5^\circ$ . A base do laser foi dividida em duas partes: a primeira, com 719 leituras foi usada para determinar o limiar da classificação  $(-0, 12)$  e a segunda parte, com 2000 leituras foi utilizada para teste.

3. Detecção de pernas no laser utilizando o algoritmo de Bellotto and Hu (2009) avaliada em relação à anotação do laser e também em relação à anotação das imagens, de forma similar ao experimento anterior. Toda a base do laser foi utilizada para teste, já que o algoritmo não requer treinamento.

4. Detecção de pessoas nas imagens utilizando aprendizado supervisionado com características HOG avaliada em relação à anotação das imagens. Para este experimento foi utilizada a implementação do trabalho de Dalal and Triggs (2005) disponível livremente na biblioteca OpenCV, com o treinamento e parâmetros padrões. Como resultado da detecção, são obtidas as coordenadas do retângulo englobando a região da imagem considerada como pessoa pelo classificador. Para comparar com a anotação das imagens, foi calculada a área de intersecção entre o retângulo da detecção e o retângulo da anotação, sendo que a razão entre a área de intersecção e a área do retângulo da anotação precisa ser maior que um limiar  $(0, 5)$  para a detecção ser considerada como verdadeiro positivo.

## 4 Resultados experimentais

As seguintes métricas foram utilizadas para avaliar os resultados dos experimentos:

- Porcentagem de verdadeiros positivos (PVP): porcentagem de detecções corretas em relação ao total de pessoas anotadas.
- Porcentagem de falsos negativos (PFN): porcentagem de pessoas anotadas que não foram detectadas ou  $100 - \text{PVP}$ . O total de pessoas anotadas (100%) é dado então por  $\text{PVP} + \text{PFN}$ .
- Precisão: razão entre o número de detecções corretas e o total de detecções ou  $\text{VP}/(\text{VP} + \text{VN})$ .

FP). A precisão é uma medida do número de acertos em relação ao total de objetos identificados pelo classificador como positivos.

O tempo médio de execução também foi calculado para os detectores de pessoas do laser e da câmera, sendo considerado como tempo inicial a partir do momento em que o dado está disponível para processamento e como tempo final o momento em que o detector fornece a resposta da classificação. Os resultados dos experimentos são mostrados na Tabela 1.

Os resultados mostraram que há diferenças entre as anotações e as detecções do laser e da câmera. Quando a anotação do laser é comparada com a anotação da câmera, nota-se que as marcações de pessoas em uma não condizem completamente com as marcações da outra, ou seja, cerca de 67,6% das pessoas anotadas nas imagens são as mesmas anotadas no laser. Cabe observar que, além das falhas do laser já citadas, essa diferença deve-se também a pessoas fora do alcance do laser que estavam visíveis na imagem, anotação minuciosa das imagens que incluiu não somente pessoas inteiras mas também partes do corpo quando havia oclusão parcial da pessoa, problemas com as bordas da imagem (uma pessoa parcialmente fora do campo de visão da câmera pode gerar um ponto médio do laser fora dos limites da imagem, não sendo computada) e segmentação imperfeita da anotação do laser para projeção na imagem. Nos resultados de detecção no laser comparados com a anotação das imagens há um número considerável de falsos negativos, que podem ser causados também por pessoas visíveis na câmera mas oclusas no laser, como mostrado na Figura 2. Como esperado, os resultados das detecções do laser, quando avaliados com a anotação das imagens também tiveram seus resultados prejudicados.

Os detectores do laser, quando comparados com a anotação do laser, mostraram bons resultados com baixo tempo de execução. Já o detector baseado em visão, apesar de apresentar precisão acima de 0,80, teve muitos falsos negativos. Neste caso, o treinamento padrão utilizado pode ter influenciado os resultados pois a base de imagens (INRIA) não possui as mesmas configurações do experimento, tais como tamanho das pessoas nas imagens. Entretanto, mesmo com um melhor treinamento, o tempo de execução da detecção em imagem é alto em comparação à detecção no laser.

Esses experimentos preliminares mostraram que os resultados de detectores de laser e câmera sozinhos podem não refletir a situação real. Apesar dos bons resultados da detecção do laser em relação à anotação do laser, este detector deixa de computar pessoas existentes na cena e que a câmera teria condições de detectar (e vice-versa). Como este trabalho está em andamento, os resultados não são definitivos mas indicam direções de pesquisa para combinação dos detectores de am-

bos os sensores.

## 5 Conclusões

Um dos principais objetivos das abordagens de fusão que detectam candidatos a pessoas com o laser e realizam a classificação das ROIs mapeadas com algoritmos de visão é a redução no tempo de processamento, pois o processamento da imagem é custoso em relação ao tempo e o espaço de busca do algoritmo de visão é reduzido com a fusão. Entretanto, os experimentos mostraram que a detecção com o laser e com a câmera estão sujeitas a falhas, comprometendo o resultado final. Dessa forma, conclui-se que a solução mais promissora para aproveitar as informações de ambos os sensores é combinar o resultado das detecções em paralelo (Figura 1(a) e (b)) e não da forma sequencial (Figura 1(c)). Isso motiva o estudo de técnicas de detecção de pessoas em imagens mais rápidas e eficientes para serem executadas em tempo real.

Em geral, o tempo de processamento tem se mostrado um dos maiores desafios dos trabalhos pois uma restrição observada é que o funcionamento dos sistemas está condicionado a baixas velocidades do robô, o que está distante da realidade de um veículo trafegando em rodovias, por exemplo. Além disso, a maior parte dos trabalhos pode detectar e rastrear apenas uma ou poucas pessoas (Cui et al., 2005).

Como trabalhos futuros pode-se listar a avaliação de algoritmos baseados em visão mais recentes como (Benenson et al., 2012) e a utilização de um método que combine as probabilidades dos detectores de cada um dos sensores.

## Agradecimentos

Os autores agradecem Abdallah Kassir pelo suporte na utilização do RADLOCC e aos colegas do laboratório CORO/UFMG pela participação na coleta das bases de dados. O presente trabalho é financiado pela FAPEMIG e foi realizado com o apoio financeiro da CAPES-Brasil. Guilherme Pereira é bolsista do CNPq.

## Referências

- Antunes, M., Barreto, J., Premebida, C. and Nunes, U. (2012). Can stereo vision replace a laser rangefinder?, *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, pp. 5183–5190.
- Araújo, R. L., Lacerda, V., Hernandez, A., Mendonca, A. and Becker, M. (2011). Classificação de pedestres usando câmera e sensor lidar, *Anais do Simpósio Bras. de Automação Inteligente*, pp. 416–420.
- Bellotto, N. and Hu, H. (2009). Multisensor-based human detection and tracking for mobile ser-

Tabela 1: Resultados dos experimentos.

Métrica	PVP	PFN	Precisão	Tempo médio (s)
Anotação do laser × anotação das imagens	67,6%	32,4%	0,90	-
HOG × anotação das imagens	35,4%	64,6%	0,81	0,97
Laser (algoritmo Spinello × anotação do laser)	96,9%	3,1%	0,66	$2,4 \times 10^{-4}$
Laser (algoritmo Bellotto × anotação do laser)	38,6%	61,4%	0,73	$3,0 \times 10^{-4}$
Laser (algoritmo Spinello × anotação das imagens)	48,6%	51,4%	0,79	-
Laser (algoritmo Bellotto × anotação das imagens)	38,7%	61,3%	0,88	-

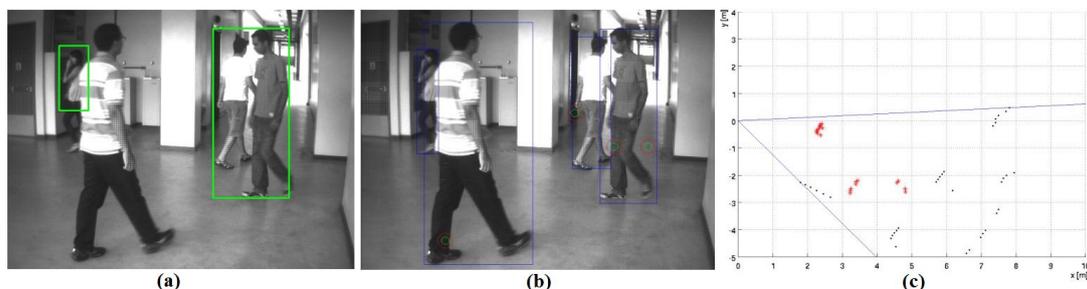


Figura 2: (a) Detecção baseada em visão. (b) Detecção do laser projetada na imagem (círculos) e anotação da imagem em azul. (c) Detecção do laser em vermelho. Nota-se que a mulher vista nas imagens não é percebida pelo laser.

- vice robots, *IEEE Trans. on Systems, Man, and Cybernetics* **39**(1): 167–181.
- Benenson, R., Mathias, M., Timofte, R. and Van Gool, L. (2012). Fast stixel computation for fast pedestrian detection, *Proc. of the Intl. Conf. on Computer Vision*, pp. 11–20.
- Bertozzi, M., Broggi, A., Cellario, M., Fascioli, A., Lombardi, P. and Porta, M. (2002). Artificial vision in road vehicles, *Proc. of the IEEE*, Vol. 90, pp. 1258–1271.
- Broggi, A., Cerri, P., Ghidoni, S., Grisleri, P. and Jung, H. G. (2009). A new approach to urban pedestrian detection for automatic braking, *IEEE Trans. on Intelligent Transportation Systems* **10**(4): 594–605.
- Cui, J., Zha, H., Zhao, H. and Shibasaki, R. (2005). Tracking multiple people using laser and vision, *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, pp. 2116–2121.
- Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection, *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, San Diego, USA, pp. 886–893.
- Kassir, A. and Peynot, T. (2010). Reliable automatic camera-laser calibration, *Proc. of the Austral. Conf. on Robotics & Automation*, p. 10.
- Mozos, O. M., Kurazume, R. and Hasegawa, T. (2010). Multi-part people detection using 2D range data, *Intl. J. of Social Robotics* **2**(1): 31–40.
- Oliveira, L., Nunes, U., Peixoto, P., Silva, M. and Moita, F. (2010). Semantic fusion of laser and vision in pedestrian detection, *Pattern Recogn.* **43**(10): 3648–3659.
- Pereira, F. G., Vassallo, R. F. and Salles, E. O. T. (2013). Human-robot interaction and cooperation through people detection and gesture recognition, *J. of Control, Automation and Electrical Systems* **24**(3): 187–198.
- Premebida, C., Ludwig, O. and Nunes, U. (2009). Lidar and vision-based pedestrian detection system, *J. of Field Robotics* **26**(9): 696–711.
- Spinello, L. and Siegwart, R. (2008). Human detection using multimodal and multidimensional features, *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, pp. 3264–3269.
- Varvadoukas, T., Giotis, I. and Konstantopoulos, S. (2012). Detecting human patterns in laser range data, *Proc. of the European Conf. on AI*, Vol. 242, pp. 804–809.
- Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 511–518.
- Wu, B., Liang, J., Ye, Q., Han, Z. and Jiao, J. (2011). Fast pedestrian detection with laser and image data fusion, *Proc. of the Intl. Conf. on Image and Graphics*, pp. 605–608.