Localização e Mapeamento On-line de Ambiente Interno para Pessoas com Deficiência

Edgar Joel Justavino Arauz
Universidade do Vale do Itajaí
Itajaí, Santa Catarina, Brasil
edgarja_26@hotmail.com

Anita Maria da Rocha Fernandes
Universidade do Vale do Itajaí
Itajaí, Santa Catarina, Brasil
anita.fernandes@univali.br

Wemerson Delcio Parreira
Universidade do Vale do Itajaí
Itajaí, Santa Catarina, Brasil
parreira@univali.br

ABSTRACT

In Brazil, there are more than 6.5 million people with some visual impairment. The total or partial loss of vision can cause a significant impact on the lives of these people. The cane is now the most widely available resource accessible to assist the locomotion of these individuals. However, it is a limited resource. Once it cannot classify the objects found, it only detects their existence. This work contributes to new discussions on location and online mapping on the theme of Assistive Technologies. We developed a system that enables users to classify environments, promoting localization through context science. Our solution uses images captured by a webcam, computer vision techniques, and probabilistic analysis. We evaluated on objective metrics showing satisfactory results.

KEYWORDS

Neural Networks, Context Science, Localization, Scene Detection

1 INTRODUÇÃO

No Brasil existem mais de 6,5 milhões de pessoas com alguma deficiência visual, segundo o Instituto Brasileiro de Geografia e Estatística (IBGE) em 2010. Além disso, de acordo com dados da OMS (Organização Mundial da Saúde), cerca de 36 milhões de pessoas no mundo são cegas enquanto outras 217 milhões têm baixa visão [1]. Para pessoas com deficiência visual (PDV) com baixa visão ou cegueira completa, a mobilidade em ambientes internos não é uma tarefa fácil, principalmente, quando o ambiente não é familiar. Geralmente, são necessárias adaptações para que a PDV possa se costumar com todos os detalhes do espaço, evitando acidentes ou ferimentos causados por obstáculos como mesas, cadeiras, camas, entre outros.

Assim, o uso de ferramentas para auxílio à navegação pode proporcionar à pessoa com deficiência visual uma maior segurança no deslocamento, independência e qualidade de vida. Atualmente, a maioria das pessoas com deficiência visual faz uso de bengalas para auxiliar na sua locomoção. As bengalas permitem o reconhecimento de obstáculos e evitam, ao máximo, as colisões indesejadas. Porém, são recursos limitados, não capazes de identificar as superfícies que tocam [2].

Um outro recurso usual são os cães-guia. De acordo com o Instituto Íris, uma das instituições especialistas em cães-guia no Brasil, o custo para preparar e doar é de aproximadamente 35 mil reais. Devido ao alto custo, o cão-guia ainda é um recurso pouco acessível. Em 2018, existiam cerca 150 deles no Brasil e o tempo de espera para receber pode chegar a 3 anos [3].

Zhang et al. [4] propuseram um novo sistema de navegação assistida baseado em localização e mapeamento simultâneos e planejamento de caminho semântico para ajudar pessoas com deficiência visual a navegar em ambientes internos. Esse sistema integra vários sensores vestíveis e dispositivos de feedback, incluindo um sensor RGB-D (Red-Green-Blue Depth Images) e uma Unidade de Medição Inercial (IMU – Inertial Measurement Unit) na cintura, uma câmera montada na cabeça, um microfone e um protetor auditivo/alto-falante. O sistema usou as imagens provenientes de uma câmera posicionada na cabeça para reconhecer os números das portas e o sensor RGB-D para detectar os principais pontos de referência, como os cantos do corredor. Ao combinar os pontos de referência detectados com as características correspondentes no mapa de piso digitalizado, o sistema localiza o usuário e fornece instruções verbais para guiar o usuário até o destino desejado. O protótipo do sistema de navegação assistida proposto foi avaliado por pessoas com visão vendada. Os testes de campo confirmaram a viabilidade dos algoritmos propostos e do protótipo do sistema.

Nguyen et al. [5] apresentaram um sistema Visual Simultaneous Localization And Mapping (SLAM) desenvolvido em um robô móvel para oferecer suporte a serviços de localização para pessoas com deficiência visual. O sistema proposto visa fornecer serviços em ambientes de pequena ou média escala, como dentro de um prédio ou campus da escola, onde dados de posicionamento convencionais, como o sistema de posicionamento global (GPS – Global Positioning System), sinais Wi-Fi, muitas vezes não estão disponíveis. Foi utilizado um método robusto de odometria visual ajustado para criar precisamente as rotas no ambiente e um algoritmo de mapeamento baseado em Fast-Apache, que pode ser o mais bem-sucedido para encontrar lugares em grandes cenários. Para estimar melhor a localização do robô, foi utilizado um Filtro de Kalman que combina os resultados correspondentes da observação atual e a estimativa dos estados do robô com base em seu modelo cinemático.

Endo et al. [6] desenvolveram um sistema de navegação para pessoas com deficiência visual, para se adaptar a mudanças ambientais dinâmicas ou transitórias, como obstáculos temporários, o sistema faz uso de uma pequena câmera vestível para estimar a posição do usuário e construir um mapa ambiental 3D. Aplicaram o conceito de localização e mapeamento simultâneo monocular direto em grande escala, do inglês *Large Scale Direct Monocular* SLAM (LSD-SLAM) para desenvolver o sistema. Esse método gera a estimativa em tempo real do auto posicionamento e geração de um mapa ambiental, com alta precisão para espaços de grande escala como ao ar livre, usando a paralaxe temporal de uma câmera monocular. O LSD-SLAM é baseado na odometria visual direta, enquanto o mapa ambiental é representado por uma estrutura gráfica, de modo que

um algoritmo de otimização gráfica possa ser adaptado para esta finalidade.

Zhang and Ye [7] apresentaram um método de estimativa de posição (EP) de 6 graus de liberdade (6-GDL) e um sistema interno de orientação de percurso para pessoas com deficiência visual. O método de EP envolve processos simultâneos de localização e mapeamento de dois gráficos para reduzir o erro acumulativo de posicionamento do dispositivo. O sistema usa a localização estimada e a planta baixa para localizar o usuário do dispositivo em um prédio e orienta o usuário anunciando os pontos de interesse e os comandos de navegação por meio de uma interface de fala. Resultados experimentais validam a eficácia do método EP e demonstram que o sistema pode facilitar substancialmente uma tarefa de navegação interna.

Ramesh et al. [8] projetaram um sistema inteligente que aborda o problema de localização em tempo real e navegação de pessoas com deficiência visual em um ambiente interno usando uma câmera monocular. O sistema integrado computacionalmente barato foi proposto para combinar geometria de imagem, Odometria Visual, Detecção de Objeto e algoritmos de Estimação de Distância e Profundidade para navegação precisa e localização utilizando uma única câmera monocular como o único sensor. O algoritmo desenvolvido foi testado para conjuntos de dados padrão Karlsruhe e ambiente interno. Testes foram realizados em tempo real usando uma câmera de smartphone que captura dados de imagem do ambiente à medida que a pessoa se move e é enviada via Wi-Fi para processamento adicional ao modelo de software MATLAB executado em um processador Intel i7. O algoritmo fornece resultados precisos na navegação em tempo real no ambiente com um feedback de áudio sobre a localização da pessoa. A trajetória da navegação é expressa em uma escala arbitrária. A localização baseada em detecção de objetos é precisa. A estimativa fornece medições de distância e profundidade com uma precisão de 94–98%.

Nesse contexto, a proposta deste trabalho é implementar um sistema que possa permitir a localização, posicionamento, de uma pessoa em ambientes fechados. Essas informações são úteis, por exemplo, na localização de pessoas com deficiência visual. As imagens que são capturadas por uma câmera representam a entrada do sistema proposto, e um aviso sonoro é a saída, para informar à pessoa sua localização em relação ao ambiente interno. Sistemas que já foram desenvolvidos para pessoas com deficiência visual (PDV) usaram câmeras 3D a outros sensores ou algoritmos muito complexos que são computacionalmente caros. Neste trabalho, propõe-se um programa que atinja essas necessidades, não englobando a projeção de um sistema integrado computacional. Serão usadas técnicas de Visão Computacional (VC) para classificação de objetos em uma cena baseado no sistema You Only Look Once (YOLO). Uma metodologia é proposta baseada em teoria probabilística para promover a ciência de contexto e consequentemente a localização de uma pessoa em um ambiente fechado.

O restante do artigo está dividido como se segue. A Seção 2 apresenta a fundamentação teórica relacionada à técnica de visão computacional que foi usada no desenvolvimento deste trabalho. A Seção 3 descreve como informações provenientes de classificadores de objetos que podem ser usados para contextualizar a cena. Na Seção 4 são apresentadas as características do projeto desenvolvimento. A Seção 5 apresenta os resultados e a discussão. Finalmente, a Seção 6 fornece as conclusões e os trabalhos futuros.

2 SISTEMA YOU ONLY LOOK ONCE – YOLO

Sistemas atuais de detecção redirecionam os classificadores para realizar a detecção, ou seja, para detectar um objeto esses sistemas usam um classificador para esse objeto e o avaliam em vários locais e escalas em uma imagem de teste. Modelos baseados em partes deformáveis, do inglês *deformable parts models* (DPM) usam uma abordagem de janela deslizante, na qual o classificador é executado em locais uniformemente espaçados ao longo de todas as imagens [9].

Abordagens mais recentes, como Region Based Convolutional Neural Networks (R-CNN) usam métodos de proposta de região para primeiro gerar potenciais caixas delimitadoras em uma imagem e depois executar um classificador nessas caixas propostas. Após a classificação, o pós-processamento é usado para refinar as caixas delimitadoras, eliminar detecções duplicadas e pontuar novamente as caixas com base em outros objetos na cena [10]. Esses pipelines complexos são lentos e difíceis de otimizar, porque cada componente individual deve ser treinado separadamente.

No sistema You Only Look Once (YOLO) utiliza-se apenas uma vez para cada imagem para prever quais objetos estão presentes e qual a sua localização na imagem. A detecção de objetos é remodelada como um único problema de regressão, diretamente dos pixels da imagem para as coordenadas das caixas delimitadoras e probabilidades de classe [11].

A Figura 1 apresenta uma representação da detecção feita pelo YOLO a partir de uma única rede convolucional que prevê simultaneamente várias caixas delimitadoras e as probabilidades de classe para essas caixas. O YOLO treina com imagens completas e otimiza

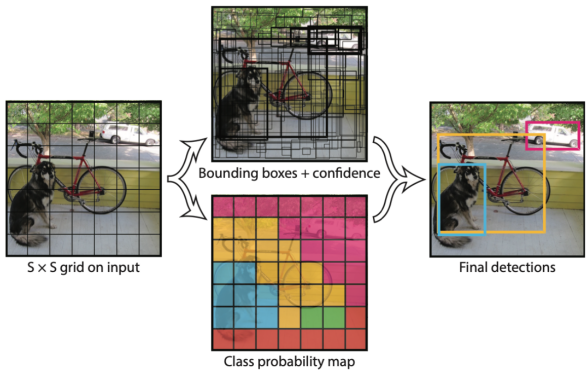

Figure 1: Sistema de Detecção YOLO [11].

diretamente o desempenho da detecção. Esse modelo unificado possui benefícios sobre os métodos tradicionais de detecção de objetos. O primeiro benefício é a velocidade. Como a detecção é tratada de modo análogo ao problema de regressão, não é preciso de um pipeline complexo. Assim, é executada uma rede neural em uma

nova imagem no momento do teste para prever as detecções. Segundo, o YOLO avalia globalmente a imagem ao fazer as previsões. Diferentemente das técnicas baseadas em janelas deslizantes e propostas por região, o YOLO reconhece a imagem inteira durante o treinamento e o tempo de teste, por isso codifica implicitamente informações contextuais sobre as classes, bem como seu aspecto [11]. No YOLO os componentes separados em uma detecção de objetos são unificados em uma única rede neural, usando recursos de toda a imagem para prever cada caixa delimitadora. Além disso, prevê todas as caixas delimitadoras em todas as classes para uma imagem simultaneamente. Isso significa que a rede atua globalmente, isto é, sobre a imagem completa e todos os objetos na imagem. O design do YOLO permite treinamento de ponta a ponta e velocidades em tempo real, mantendo alta precisão média [11].

3 LOCALIZAÇÃO E MAPEAMENTO USANDO CIÊNCIA DE CONTEXTO

Ciência de contexto significa que é possível usar as informações de contexto em tomadas de decisão, por exemplo. Um sistema reconhece o contexto se puder extrair, interpretar e usar informações de contexto e adaptar sua funcionalidade ao contexto atual de uso [12]. O termo *contextaware computing* é comumente entendido por aqueles que trabalham com reconhecimento de contexto. Um objetivo comum dos sistemas sensíveis ao contexto é adquirir e utilizar informações sobre o contexto de um dispositivo para, por exemplo, fornecer serviços que são apropriados para pessoas, lugares, horas, eventos, etc. [13].

Alguns exemplos de computação ciente de contexto são os sistemas de navegação que se utilizam de parâmetros como a localização atual do dispositivo para automaticamente ajustar a visualização do mapa. Além dessas, as indicações das setas e as instruções dadas, parâmetros como a hora do dia, condições de luz, de clima e de tráfego, usados para ajustar a visualização da luz de fundo do mapa (fundo claro para o dia e escuro para noite) e a própria rota são exemplos relevantes.

Os elementos naturais da cena não têm especificações de padrões predeterminados, como por exemplo códigos de barras ou outros tipos de marcadores. Essa variedade de visuais de características distintas requer o uso de técnicas mais robustas em termos de requisitos de potência de processamento. Devido a esse fato e a capacidade de processamento em tempo real, sua detecção e o seu reconhecimento usando pequenos dispositivos portáteis têm sido, há algum tempo, bastante limitado.

A literatura sobre reconhecimento de objetos é muito rica e está em crescimento. O objetivo de permitir ao robô um conhecimento contextual pode ser alcançado a partir do reconhecimento de objetos e da tomada de decisão com base nessas ideias. Qualquer um dos objetos classificados ou segmentados na cena pode ser usado para fins de seu entendimento. A segmentação de uma cena, enquanto cada vez mais intensivo em termos de computação, fornece informações com resultados mais precisos [14].

Esses avanços permitem a construção de mapas de ambientes desconhecidos usando Localização e Mapeamento Simultâneos (SLAM, VSLAM) e fornecem conhecimento contextual extraindo informações semânticas sobre os objetos presentes nos arredores, a fim de tomar decisões específicas da situação (por exemplo, estabelecer uma rota).

Nesse sentido, este trabalho contribui com problema de SLAM, VSLAM para mapeamento e movimentação em ambientes internos com ciência de contexto. Não será considerado no escopo deste trabalho a produção ou desenvolvimento de mapas para movimentação. O sistema apenas, produz a identificação e a ciência de contexto para o mapeamento e a localização não baseada em coordenadas do usuário considerando apenas a classificação da cena (ambiente).

4 PROJETO

Para o desenvolvimento deste trabalho um sistema com a integração da linguagem Python e da biblioteca OpenCV foi implementado. O propósito é investigar as potencialidades das técnicas de visão computacional para auxílio de pessoas na localização on-line dentro de ambientes internos utilizando uma câmera como único sensor. Uma visão geral do sistema proposto pode ser observada na Figura 2.

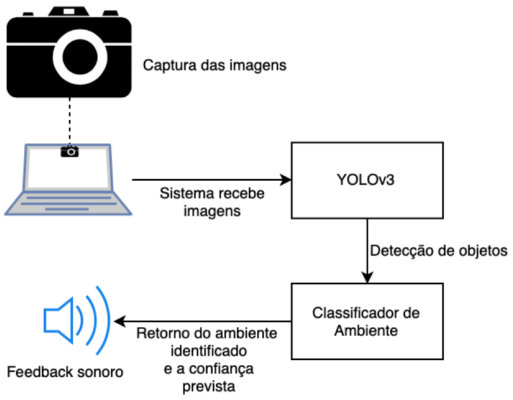

Figure 2: Visão geral do sistema proposto.

A Figura 3 representa o diagrama de blocos do sistema desenvolvido. O sistema se inicia com a obtenção das imagens do ambiente a partir da câmera. Essas imagens são enviadas ao bloco de detecção de objetos, YOLOv3, para identificar os objetos de interesse para localização no ambiente conforme a pessoa está se movendo em um processo on-line. Em seguida, inicializa o processo que é proposto neste trabalho. Dos objetos identificados são extraídas as classes que são comuns de encontrar dentro de ambientes internos, baseados em cenários residenciais, como, por exemplo, casa e apartamento. Assim, é informado ao usuário o ambiente que foi reconhecido e o nível de confiança. O retorno das informações — o feedback informando o ambiente em que se encontra — é feito no formato de áudio, o que permite o acesso pela PDV.

Para o desenvolvimento deste trabalho foi utilizado o dataset COCO (*Common Object in Context*) dado que esse possui uma grande quantidade de objetos que se encontram dentro dos ambientes alvo. O COCO possui 80 categorias de classes diferentes.

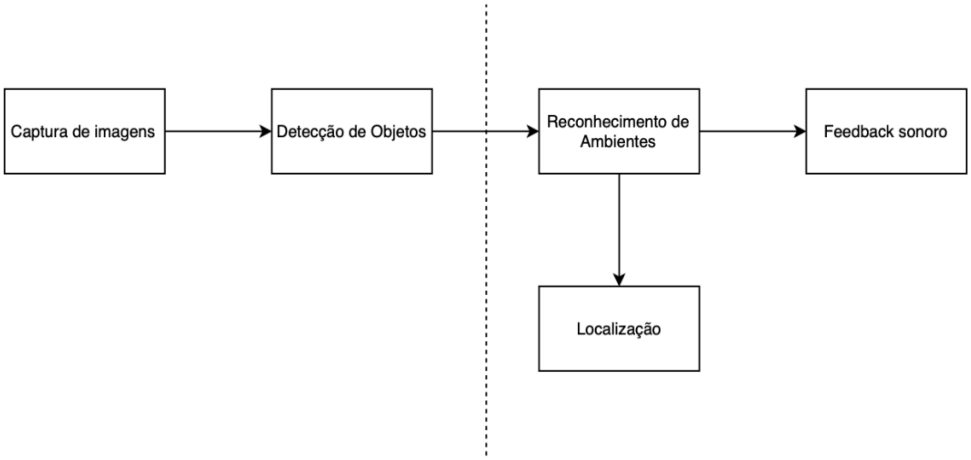

Figure 3: Diagrama de blocos do sistema proposto.

Foram utilizadas as classes as quais os objetos de interesse pertencem. Além disso, são considerados a qual ambiente o objeto pertence e seus respectivos pesos.

Os pesos atribuídos permitem uma separação dos objetos do dataset em que se encontra, está relacionado com sua capacidade de serem trocados de lugar. Assim, estabelece-se a seguinte classificação quanto a mobilidade dos objetos: fixos (peso = 3), pouca mobilidade (peso = 2) e alta mobilidade com (peso = 1). Por exemplo, baseado no ambiente de uma sala de estar, um sofá seria um objeto com peso 3 por ser fixo, uma caixa de som seria um objeto de peso 2 por poder trocar de lugar em algumas ocasiões, ou até um aparelho de TV, pois atualmente é um objeto encontrado em mais ambientes além da sala, e um livro ou controle remoto seriam objetos com peso 1 por poder se encontrar com mais facilidade em outros ambientes por sua alta mobilidade.

Finalmente, considerando o índice de confiança de cada objeto e suas características é possível estabelecer uma regra para classificar (reconhecer) um ambiente. Para isso, propõe-se o uso de uma metodologia baseada em probabilidade. A probabilidade do indivíduo estar no j-ésimo ambiente, $Pr(L_j)$, é definida por:

$$Pr(L_j) = \frac{\sum_{n=1}^{N} \text{conf}_n \, \text{peso}_n}{\sum_{n=1}^{N} \text{peso}_n} \tag{1}$$

em que N é o número de objetos que foram classificados com confiança maior de $0,65$ na cena e o n-ésimo peso associado ao n-ésimo objeto detectado. Assim, usando (1), o ambiente L, em que o usuário está, é obtido por:

$$L = \arg\max_{j} \{Pr(L_j)\}. \tag{2}$$

A estratégia proposta permite que os objetos classificados possam dar um contexto a cena sem a necessidade de procedimento extras de análise baseando-se por exemplo em outros bancos de dados ou outros algoritmos de classificação ou rotulação de ambiente, tais como [4–8]. Além disso, a classificação do ambiente a partir da classificação de vários objetos permite ao usuário uma realimentação da localização e do mapeamento interpretável da cena.

5 RESULTADOS

Nesta seção são apresentados os resultados obtidos na fase de validação do projeto desenvolvido. A partir dos experimentos realizados com o sistema responsável por realizar a localização e mapeamento on-line para ambientes internos considerando a possibilidade de ser usado para integrar um dispositivo automático para PDV.

Os testes foram realizados em uma máquina com processador Dual-Core Intel Core i5 2,7GHz, memória RAM de 8GB 1867MHz, SSD Apple SM0256G de 251GB e placa de vídeo Intel Iris Graphics 6100 1536MB. No total, foram realizados 10 testes em cada ambiente, de modo que os ambientes foram aleatoriamente modificados para a avaliação do sistema proposto.

A Figura 4 apresenta um exemplo de cena de como é feito o reconhecimento de ambiente a partir dos objetos detectados na imagem. Usando a classificação dos objetos da cena é possível estabelecer a localização do usuário. Na imagem apresentada foram detectados os objetos sofá, cadeira, teclado, livro, mouse, controle remoto, cada um com a sua confiança dada pelo YOLO.

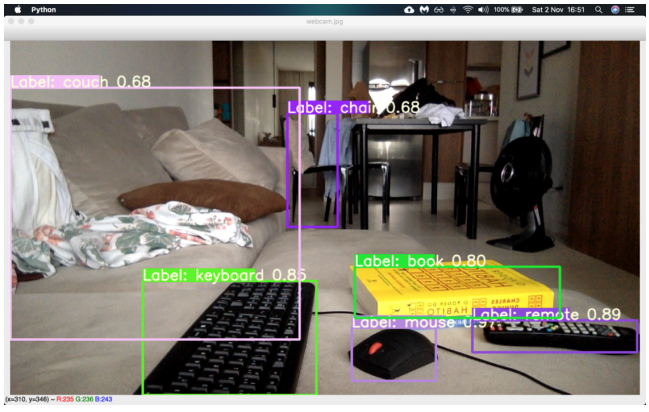

Figure 4: Exemplo de cena para classificar o ambiente.

A Figura 5 representa um exemplo para o funcionamento do sistema proposto. Inicialmente, foram detectados 6 objetos, a saber,

sofá, cadeira, teclado, livro, mouse, controle remoto, todos com os valores de confiança acima de 65%, superior ao limiar estabelecido, para compor os valores usados na verificação de classificação de ambiente. Posteriormente, é feito a soma ponderada da confiança (%) dos objetos identificados, usando os respectivos pesos nos índices dos ambientes a qual pertencem. Os pesos são somados também ao vetor de somatórios de pesos, funcionando igual à posição do vetor de ambientes. Verificado todos os objetos, é conferido no vetor de ambientes qual foi a posição que obteve maior confiança em relação àquela imagem obtida pela câmera. É utilizado o índice do vetor de ambientes que obteve maior confiança no vetor de somatórios de pesos e feito o cálculo de média ponderada, assim, obtendo uma confiança final em relação àquele ambiente.

A Tabela 1 apresenta a Matriz de Confusão com a quantidade de acertos e erros que o sistema apresentou como resultado quando testado em uma área residencial que possuía todos os ambientes classificados nesse trabalho. Foram realizadas 10 observações independentes para cada ambiente de uma residência composta por 1 sala de estar (SE), 1 sala de jantar (SJ), 1 cozinha (CZ), 1 quarto (QT), 1 escritório (ES) e 2 banheiros (BN). Totalizando 70 testes independentes. Os dados estão sumarizados na Figura 6.

Table 1: Matriz de Confusão

	SE	SJ	CZ	BN	QT	ES
SE	9	0	0	0	1	0
SJ	0	9	1	0	0	0
CZ	0	0	10	0	0	0
BN	0	0	3	17	0	0
QT	1	0	0	0	7	2
ES	0	0	0	0	0	10

Pode ser visto na Figura 6 que o usuário se encontrando na Sala de Estar o sistema reconheceu aquele ambiente 9 vezes como Sala de Estar e 1 vez como Quarto. Isso ocorreu porque no momento da captura da imagem o sistema reconheceu objetos que pertencem a mais de uma classe como dos objetos TV, controle remoto e objetos que frequentemente podem ser levados de um ambiente para outro como é o caso do objeto de ursinho de pelúcia. Assim, houve um reforço a classificação para o ambiente Sala de Estar como ambiente Quarto.

Nos testes realizados com o usuário na Sala de Jantar o sistema o reconheceu como se o usuário estivesse na Sala de Jantar 9 vezes e na cozinha 1 vez. Pois, no momento da captura da imagem foi detectada a geladeira e a pia que pertencem ao ambiente Cozinha, que fica próximo, e sem barreiras físicas ou paredes a Sala de Jantar. Dado que a geladeira e a pia possuem um peso 3, enfatizou em detrimento dos pesos dados a mesa de jantar e as duas cadeiras exclusivas à Sala de Jantar.

No ambiente da Cozinha o sistema reconheceu como o ambiente Cozinha as dez vezes. Esse resultado é esperado, dado que, a Cozinha possui vários objetos do tipo fixo (peso 3) e outros de pouca mobilidade.

Em relação ao ambiente Banheiro, havia dois banheiros na área residencial testada, por isso a maior quantidade de testes. O ambiente foi reconhecido como Banheiro 17 vezes e como Cozinha 3 vezes pelo sistema. Isso ocorreu porque a cena estava contaminada com objetos pertencentes a outros ambientes com objetos da Cozinha que possuíam peso 2, classificando-o assim como ambiente Cozinha.

Para o ambiente Quarto o sistema reconheceu o ambiente Quarto 7 vezes, 2 vezes como o ambiente Escritório e 1 vez como o ambiente Sala de Estar. Isso foi devido ao fato de que o ambiente foi contaminado com objetos do ambiente Escritório, o mouse, o teclado, o notebook e a TV por ser do ambiente Escritório acabou dando maior confiança em ser reconhecido com ambiente Escritório. Foi classificado como Sala de Estar 1 vez quando havia pouca iluminação no Quarto e a cama não foi detectada, sendo detectado só a TV e o controle remoto, dando como resultado final a Sala de Estar.

No ambiente Escritório o sistema reconheceu o ambiente como Escritório nas dez vezes, sendo que só o notebook foi removido do ambiente quando testado com contaminação, para teste de robustez do sistema.

Dado os valores de acertos e erros foi construída a partir da Matriz de Confusão um resumo dos indicadores apresentado no Quadro 2, que mostra as frequências de classificação para cada classe do modelo. Essa classificação é dividida em 4 tipos indicadores:

- Verdadeiro positivo (VP): ocorre quando no conjunto real, a classe que estamos buscando foi prevista corretamente. Por exemplo: quando o ambiente é a sala e o sistema previu corretamente que o usuário está na sala;
- Falso Positivo (FP): ocorre quando no conjunto real, a classe que estamos buscando prever foi prevista incorretamente. Por exemplo: o usuário não está na sala e o sistema prevê que ele está;
- Falso Negativo (FN): ocorre quando no conjunto real, a classe que não estamos buscando prever foi prevista incorretamente. Por exemplo: quando o usuário está na sala e o sistema previu incorretamente que ele não está na sala;
- Verdadeiro Negativo (VN): ocorre quando no conjunto real, a classe que não estamos buscando prever foi prevista corretamente. Por exemplo: o usuário não está na sala e o sistema previu corretamente que ele não está na sala.

A partir dos valores apresentados no Quadro 2 efetua-se a análise de desempenho. Foram aplicadas 6 métricas de avaliação: Acurácia (ACC – *Accuracy*), Razão de erro (EER – *Error rate*), Precisão (PREC – *Precision*), Sensibilidade (SN – *Sensitivity*), Valor-F (F_1) e Especificidade (SP – *Specificity*). Essas métricas são descritas, a seguir:

i. A acurácia é um indicador que mede a quantidade de acertos sobre o todo. Pode-se afirmar que é o percentual de instâncias classificadas corretamente. Essa medida é calculada usando:

$$\mathrm{ACC} = \frac{\mathrm{VP} + \mathrm{VN}}{\mathrm{VP} + \mathrm{VN} + \mathrm{FP} + \mathrm{FN}} \tag{3}$$

ii. Razão de erro verifica se o erro está alto ou baixo, podendo descartar toda a amostragem avaliada, ou informar se há algum erro na coleta dos dados, ou seja:

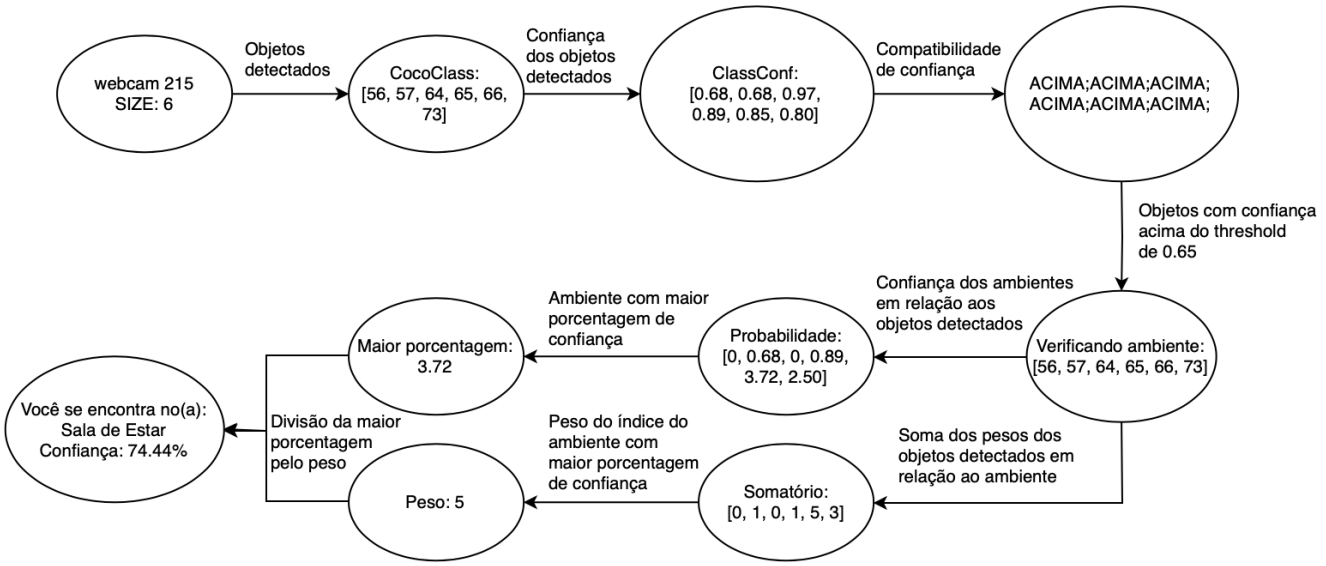

Figure 5: Funcionamento do sistema implementado.

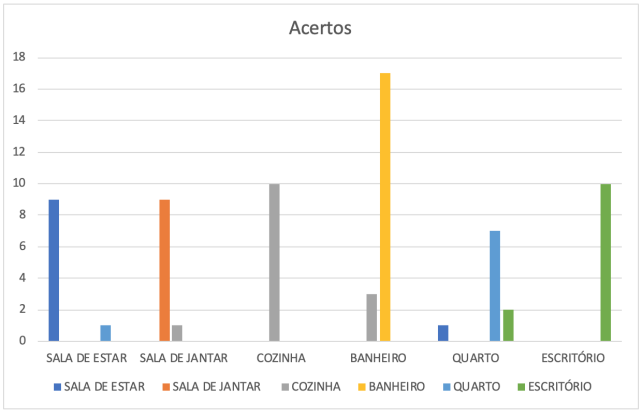

Figure 6: Acertos e erros no processo de detecção de cada ambiente.

$$\text{EER} = \frac{\text{FP} + \text{FN}}{\text{VP} + \text{VN} + \text{FN} + \text{FP}} \tag{4}$$

iii. Precisão verifica o número de vezes que uma classe foi predita corretamente sobre o falso positivo, dada por:

$$\text{PREC} = \frac{\text{VP}}{\text{VP+FP}} \tag{5}$$

iv. Sensibilidade verifica o número de vezes que uma classe foi predita corretamente sobre o falso negativo, que é:

$$\text{SN} = \frac{\text{VF}}{\text{VP} + \text{FN}} \tag{6}$$

Table 2: Resumo dos indicadores a partir da Matriz de Confusão

AMBIENTES	VP	FP	FN	VN
SE	9	1	1	59
SJ	9	0	1	60
CZ	10	4	0	56
BN	17	0	3	50
QT	7	1	3	43
ES	10	2	10	58
TOTAL	62	8	8	326

v. Valor-F informa o desempenho das amostras a partir da média harmônica entre a precisão e a sensibilidade, frequentemente usada para medir o desempenho da pesquisa e é dada por:

$$\text{F}_1 = 2\,\frac{\text{PREC} \times \text{SN}}{\text{PREC} + \text{SN}} \tag{7}$$

vi. Especificidade é o número de vezes que uma classe foi predita verdadeiramente positiva, dada por:

$$\text{SP} = \frac{\text{VN}}{\text{VN} + \text{FP}} \tag{8}$$

Os valores encontrados com a aplicação das métricas de desempenho, Equações (3) – (8), são exibidas na Tabela 3. Ao analisar os resultados obtidos para o reconhecimento de ambiente on-line, recebendo como entrada a imagem capturada pela webcam do

notebook, observou-se que o sistema proposto obteve um bom desempenho na classificação de ambiente. Porém, em alguns casos em que o ambiente foi classificado como erroneamente por motivos de ambientes muito próximos havendo a detecção de objetos de outros ambientes, como a Sala de Jantar próxima à Cozinha. Outro motivo foram os testes de robustez do sistema que envolvia contaminação do ambiente com outros objetos, havendo falha na identificação da Sala de Estar e Quarto. E por último, a iluminação reduzida, em que as imagens do Banheiro foram penalizadas, gerando uma dificuldade extra ao sistema.

Table 3: Métricas de desempenho, Equações (3) – (8).

AMB.	ACC	EER	PREC	SN	F_1	SP
SE	0,9714	0,0286	0,9000	0,9000	0,9000	0,9833
SJ	0,9857	0,0143	1,0000	0,9000	0,9474	1,0000
CZ	0,9429	0,0571	0,7143	1,0000	0,8333	1,0000
BN	0,9571	0,0429	1,0000	0,8500	0,9189	0,9434
QT	0,9429	0,0571	0,8750	0,7000	0,7778	0,9516
ES	0,9714	0,0286	0,8333	1,0000	0,9091	1,0000
MÉDIA	0,9619	0,0381	0,8871	0,8917	0,8811	0,9797

O tempo de resposta do sistema no processo de mapeamento e localização de ambientes está na faixa de tempo de 500 a 750 ms.

Suprime-se uma comparação de desempenho (como precisão, acurácia e tempo médio de processamento) do sistema proposto com os trabalhos relacionados por não ser possível realizar de forma justa. Pois, em tais sistemas o desempenho está condicionado também aos datasets e aos sensores empregados, por exemplo.

6 CONSIDERAÇÕES FINAIS

Neste trabalho relaciona os conceitos de reconhecimento de objetos, utilizando estruturas de inteligência artificial como redes neurais convolucionais e Visão Computacional. Com base nesse estudo implementou-se um sistema que realiza a detecção de objetos online com base em redes pré-treinadas utilizando YOLO e o COCO. Também buscou-se recursos computacionais capazes de extrair as informações precisas a partir de imagens de uma forma eficaz, de baixo custo financeiro e com um bom desempenho nos resultados finais. Os testes mostram que o projeto atendeu aos requisitos levantados para o desenvolvimento. Os valores médios das métricas de desempenho foram de 96% para ACC, 3% para EER, 88% para PREC, 89% para SN, 88% F_1 e 97% para SP. Como trabalhos futuros considera-se uma investigação do comportamento do sistema proposto usando o YOLO com outros datasets que possuam objetos que se encontrem dentro de ambientes residencias, e até focar em outros ambientes internos que possam ser um grande auxílio para pessoas com deficiência visual. Além disso, considera-se a proposta de um dispositivo portátil que possa ser testados com PDV.

REFERENCES

[1] Fundação Dorina. Estatísticas da deficiência visual. https://www.fundacaodorina.org.br/a-fundacao/deficiencia-visual/estatisticas-da-deficiencia-visual/, 2017.

[2] Victor Correia. Sistema permite que deficientes visuais andem sem usar bengala. https://www.correiobraziliense.com.br/app/noticia/tecnologia/2017/06/19/interna_tecnologia,603263/sistema-permite-que-deficientes-visuais-andem-sem-usar-bengala.shtml, 2017.

[3] Fernando Freitas. 8 curiosidades sobre o cão-guia. https://www.fundacaodorina.org.br/blog/8-curiosidades-sobre-o-cao-guia/, 2018.

[4] Xiaochen Zhang, Bing Li, Samleo L Joseph, Jizhong Xiao, Yi Sun, Yingli Tian, J Pablo Muñoz, and Chucai Yi. A slam based semantic indoor navigation system for visually impaired users. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1458–1463. IEEE, 2015.

[5] Quoc-Hung Nguyen, Hai Vu, Thanh-Hai Tran, David Van Hamme, Peter Veelaert, Wilfried Philips, and Quang-Hoan Nguyen. A visual slam system on mobile robot supporting localization services to visually impaired people. In *European Conference on Computer Vision*, pages 716–729. Springer, 2014.

[6] Yuki Endo, Kei Sato, Akihiro Yamashita, and Katsushi Matsubayashi. Indoor positioning and obstacle detection for visually impaired navigation system based on lsd-slam. In *2017 International Conference on Biometrics and Kansei Engineering (ICBAKE)*, pages 158–162. IEEE, 2017.

[7] He Zhang and Cang Ye. An indoor wayfinding system based on geometric features aided graph slam for the visually impaired. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(9):1592–1604, 2017.

[8] Kruthika Ramesh, SN Nagananda, Hariharan Ramasangu, and Rohini Deshpande. Real-time localization and navigation in an indoor environment using monocular camera for visually impaired. In *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*, pages 122–128. IEEE, 2018.

[9] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 32(9):1627–1645, 2009.

[10] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[11] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[12] Hee Eon Byun and Keith Cheverst. Utilizing context history to provide dynamic adaptations. *Applied Artificial Intelligence*, 18(6):533–548, 2004.

[13] Jong-yi Hong, Eui-ho Suh, and Sung-Jin Kim. Context-aware systems: A literature review and classification. *Expert Systems with applications*, 36(4):8509–8522, 2009.

[14] Khashayar Asadi, Hariharan Ramshankar, Harish Pullagurla, Aishwarya Bhandare, Suraj Shanbhag, Pooja Mehta, Spondon Kundu, Kevin Han, Edgar Lobaton, and Tianfu Wu. Vision-based integrated mobile robotic system for real-time applications in construction. *Automation in Construction*, 96:470–482, 2018.