

MobiLysa - Sistema de localização e controle do cão-guia robô Lysa para ambientes internos baseado em visão computacional

Wagner Augusto Aranda Cotta
waacotta@gmail.com
Univers. Federal do Espírito Santo
Vitória, Espírito Santo

Alexandre Pereira do Carmo
alexandre.carmo@ifes.edu.br
Instituto Federal do Espírito Santo
Guarapari, Espírito Santo

Flávio Machado
k.flaviomachado@gmail.com
Univers. Federal do Espírito Santo
Vitória, Espírito Santo

Anilton Salles Garcia
aniltonsallesg@gmail.com
Univers. Federal do Espírito Santo
Vitória, Espírito Santo

Raquel Frizera Vassallo
raquel@ele.ufes.br
Univers. Federal do Espírito Santo
Vitória, Espírito Santo

Nedinalva de Araujo Sellin
neide@vixsystem.com.br
Vix System
Serra, Espírito Santo

ABSTRACT

Lysa is a guide-robot whose goal is to assist people with visual impairment, improving their autonomy and mobility. Lysa is able to detect potholes, ground obstacles and risks of collision with high objects, warning the user of such dangers by voice. However Lysa is still unable to guide a person to a place within an environment. Therefore, the aim of this project will be to develop MobiLysa: a service to perform localization, navigation and control of the robot Lysa so it can guide a person to a desired location inside a building. That will be done using an Intelligent Space based on computer vision. The user will be able to indicate the desired destination, either by voice or other form of interaction, and be guided to the place independently and safely.

KEYWORDS

Computer Vision, ARUco, Intelligent Space

1 INTRODUÇÃO

No Brasil, estima-se que mais de seis milhões e meio de pessoas sejam portadoras de deficiência visual [9] e, embora não existam estatísticas oficiais, estima-se que existam menos de duzentos cães-guias no país.

O treinamento de um cão-guia é diferenciado e custa, em média, R\$ 50.000,00. Além disso, o treinamento é realizado somente em alguns lugares do Brasil, gerando assim uma longa fila para conseguir um cão treinado. O longo tempo de espera e o alto custo reduzem significativamente as possibilidades de pessoas com deficiência visual terem esse recurso.

O cão-guia robô Lysa [25], Figura 1, tem funções semelhantes às de um cão-guia convencional. É equipado com dois motores e cinco sensores que avisam aos deficientes visuais por meio de mensagens de voz gravadas, quando há buracos, obstáculos e riscos de colisões em altura. Para dar maior autonomia a esses, estão sendo desenvolvidos algoritmos de navegação com o uso de GPS, onde o usuário pode informar sua rota e a Lysa o levará até o seu destino. Entretanto, o uso de GPS é capaz de auxiliar apenas na localização e navegação da Lysa em ambientes externos, já que tal equipamento

não apresenta funcionamento adequado em ambientes internos. A localização e controle em ambientes internos pode ser conseguida através dos atualmente chamados Espaços Inteligentes [16].



Figura 1: Cão-guia Robô Lysa [25]

Um Espaço Inteligente [8] pode ser definido como um ambiente equipado com uma rede de sensores, que obtém informações sobre o mundo que observa (câmeras, microfones, termômetros, por exemplo), e uma rede de atuadores (robôs móveis, telas de informação, eletrodomésticos automatizados, entre outros), que permite sua interação com os usuários e alteração do próprio ambiente. Tanto os sensores quanto os atuadores devem ser governados por um sistema capaz de coletar e analisar informações obtidas pelos sensores e tomar decisões.

Nesse contexto, o uso de um Espaço Inteligente para localizar e controlar o cão-guia robô Lysa pode contribuir para melhorar o atendimento aos seus usuários. O sensoriamento distribuído e a inteligência presente no ambiente permitem que informações mais amplas, dadas por câmeras por exemplo, e não apenas por sensores do próprio robô, sejam usadas para tarefas mais complexas. Assim a localização e navegação do robô podem ser realizados de forma mais fácil e robusta, aumentando a confiabilidade do sistema e conforto do usuário.

O Espaço Inteligente em questão, interliga todos os dispositivos em rede, abstraindo-se da infraestrutura física e disponibilizando os recursos como serviços em um modelo *Platform as a Service* (PaaS). Nesse tipo de arquitetura, a inclusão de novos equipamentos e aplicações ocorrem normalmente na forma de novos serviços.

Logo, pensando-se nas vantagens que se pode agregar ao robô Lysa, é que este projeto MobiLysa foi concebido. O principal objetivo será desenvolver um serviço de localização e navegação para que o robô Lysa guie uma pessoa até um local desejado dentro de

In: VI Workshop "O Futuro da Videocolaboração" (WCT-Video 2019), Rio de Janeiro, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (Web-Media). Porto Alegre: Sociedade Brasileira de Computação, 2019.
ISSN 2596-1683

um prédio público, através de uma arquitetura de Espaço Inteligente baseada em serviços. Desta forma, o usuário poderá indicar o destino desejado, seja por voz ou outra forma de interação, e ser guiado até o local de maneira independente e segura.

A solução proposta neste trabalho e que se encontra em desenvolvimento fará uso de uma rede de câmeras para localizar e controlar o robô Lysa. Além de se considerar a instalação de câmeras nos ambientes, espera-se que a solução proposta possa ser adaptada também a ambientes que já possuem câmeras de vigilância instaladas, aproveitando-se os sensores e infraestrutura física disponíveis. Caso a proposta seja validada com sucesso, essa poderá ser estendida para diversos outros ambientes sejam esses públicos, privados ou comerciais.

2 ESTADO DA ARTE

Uma vez que este projeto visa realizar a localização e navegação do robô Lysa em ambientes internos, o estado da arte a ser percorrido é com relação às diferentes tecnologias disponíveis e métodos para localização de dispositivos, bem como a estrutura necessária para a correta implementação.

O trabalho apresentado em [4], mostra como a localização se tornou um fator chave para serviços específicos que são oferecidos ou executados de acordo com o dispositivo ou localização do usuário.

É notório que atualmente, abordagens que exploram cooperação e coordenação com mecanismos de localização, para tecnologias sem fio/móveis, são certamente uma tendência a ser seguida em pesquisas nas áreas de redes e robótica. Abordagens como as apresentadas nos trabalhos [1][17] mostram como o uso de modelos de erro de alcance e a cooperação entre dispositivos podem melhorar a localização em redes de comunicação sem fio. Além disso, no trabalho [18], os autores demonstram que a fusão multissensorial e a cooperação espacial podem aumentar significativamente a localização em um cenário de grande escala com centenas de agentes móveis.

No entanto, problemas como obstrução de sinais entre emissores e receptores geralmente dificultam o uso de métodos tradicionais de localização baseados em radiofrequência (baseados em RF). A intensidade do sinal, o atraso do sinal e o ângulo de chegada da propagação da onda eletromagnética podem ser prejudicados por níveis severos de impedimentos e interferências em determinados ambientes [23].

Dessa forma, uma abordagem promissora para obter precisão de localização na ordem de poucos centímetros é usar métodos de visão computacional em ambientes inteligentes [23][7]. Imagens podem fornecer uma riqueza de informações que, a partir do devido processamento, são capazes de gerar estimativas de localização bastante precisas.

Um comparativo entre as principais técnicas de localização em ambientes internos é mostrado na Tabela 1, onde pode-se ver que técnicas de localização por visão computacional possuem uma melhor precisão quando comparadas com outras técnicas tradicionais de localização em ambientes internos.

No entanto, existem alguns desafios em relação ao uso de visão computacional e processamento de imagens para se localizar dispositivos móveis. Apesar de fornecer informações e localizações precisas, extrair tais dados a partir de um conjunto de imagens

Tecnologia	Precisão Aproximada	Desvantagem
Tecnologias com codificação de sinal		
Infravermelho	57cm ~ 23m	Interferência do Sol
VLC	10cm	Alta custo
Ultrassom	1cm ~ 2m	Interferências
WiFi	1.5m	Vulnerável à mudanças no AP
Bluetooth	30cm ~ 10m	Necessita mapeamento de sinal
ZigBee	25cm	Necessita equipamento específico
RFID	1m ~ 5m	Precisão muito baixa
UWB	15cm	Alto custo
Tecnologias passivas sem codificação de sinal		
Geomagnético	2m	Necessita de mapeamento
Inercial	2m	Acumula erro
Som ambiente	-	Sem precisão
Luz ambiente	10cm ~ alguns metros	Sensibilidade a luminosidade
Visão Computacional	1cm ~ 1m	Sensibilidade a luminosidade

Tabela 1: Comparativo tecnológico, adaptado de [23]

geralmente requer que grandes larguras de banda e algoritmos complexos sejam executados em tempo real.

Essa é, na verdade, a principal desvantagem de usar câmeras como sensores. Processar informações complexas, como vídeo, que contém três dimensões, para o caso de imagens em escala de cinza (X, Y e tempo) ou seis dimensões, para o caso de imagens coloridas (R, G, B, X, Y e tempo) geralmente representa alto custo computacional. Nessas situações, os algoritmos para extração de dados se tornam mais complexos e normalmente exigem *hardware* robusto para processar a quantidade de informações com rapidez suficiente.

Entretanto, os algoritmos de visão computacional têm experimentado grandes avanços [15]. Além disso, atualmente já existem *hardware* mais potentes e mais baratos, com várias soluções desenvolvidas especificamente para processamento de imagens [2], o uso de unidade de processamento gráfico [3] e microcontroladores dedicados [13]. Tudo isso, permite o desenvolvimento de novas e variadas aplicações para visão computacional, as quais tornam o seu uso cada vez mais popular.

Especificamente no caso de localização, se as câmeras estiverem calibradas corretamente e sua posição relativa for conhecida, é possível obter uma precisão de alguns centímetros na estimativa da posição de um objeto ou indivíduo. Quando uma das dimensões do objeto é dada, estimar sua localização pode se tornar ainda mais simples, apesar de continuar dependendo da qualidade da calibração.

Todavia, uma vez que a calibração for feita, não há necessidade de recalibração, a menos que as câmeras tenham sua postura (posição e orientação) alterada. Além disso, não há erro acumulativo no processo de localização por visão computacional, ou seja, mesmo que o robô se mova pelo local de trabalho, o erro de localização é constante e não muda com o seu deslocamento.

Vale ainda ressaltar que, normalmente em ambientes multicâmera, quanto mais sobreposição houver entre as regiões visualizadas pelas câmeras, melhor será a precisão da localização.

Conforme mencionado anteriormente, devido à grande quantidade de dados, o uso da visão computacional requer alta largura de banda conjuntamente à necessidade de baixa latência entre os

nós de processamento, principalmente quando a aplicação deve ser executada em tempo real. Porém, se a infraestrutura física e virtual permitir que o processamento pesado seja paralelizado e executado na nuvem, a visão computacional se torna uma excelente solução, além de estratégica para localização e controle de dispositivos móveis.

Em [5], pode-se observar como um Espaço Inteligente Programável baseado em visão computacional permite que os requisitos de latência e controle em tempo real desse tipo de aplicação sejam atendidos.

A arquitetura utilizada em [6] e [5] permite que todos os serviços sejam executados de forma distribuída, em um ambiente de nuvem computacional com alta capacidade de processamento. Além disso, a orquestração dos recursos é realizada de forma a atender aos requisitos elásticos dos serviços implementados que necessitem de alto poder computacional. Para isso, são usados juntamente com o conceito de arquitetura baseada em serviço, conceitos e ferramentas de computação em nuvem [19].

3 PROPOSTA DE TRABALHO

Para se desenvolver o serviço de localização e navegação do robô Lysa em ambientes internos pretende-se utilizar uma arquitetura de Espaço Inteligente Programável baseado em visão computacional. A ideia é usar uma rede de câmeras para localizar o robô no ambiente e realizar o controle do dispositivo principalmente por realimentação visual.

A localização do robô será feita em tempo real usando-se um método de marcadores visuais ou método de reconhecimento de objetos. Dada a localização, a navegação do robô Lysa poderá ser controlada fazendo-se uso de um mapa do ambiente. Tal mapa será previamente construído e conterá todas as informações relevantes para condução dos usuários aos locais de destino disponíveis, assim como indicações de elevadores, escadas, bebedouros, etc. Além disso, esse mapa deverá ser atualizado sempre que alterações significativas acontecerem no ambiente.

Apesar de se considerar o uso de um Espaço Inteligente Programável baseado em visão computacional para o desenvolvimento do serviço MobiLysa, o mapa a ser utilizado não é formado pelas imagens capturadas pelas câmeras. Esse mapa será do tipo híbrido (topológico e métrico). Dessa forma, mesmo que as câmeras do ambiente não apresentem sobreposição em todas as regiões, ou seja, apresentem zonas mortas, a navegação será possível devido aos sensores embarcados no robô e estratégia de navegação adotada.

Vale ressaltar que as tarefas de detecção e desvio de obstáculos são funcionalidades embarcadas no próprio robô e, portanto, não representam uma preocupação para o serviço de localização e navegação proposto. Caso algum desvio aconteça por causa de um obstáculo, assim que possível o robô será redirecionado à trajetória planejada.

Além de imagens, dados de odometria ou de outros sensores locais do robô podem ser usados para melhorar o desempenho do controlador, principalmente nas zonas mortas mencionadas anteriormente. É importante que a navegação no ambiente possa ocorrer de forma suave e estável, uma vez que o robô estará guiando uma pessoa com deficiência visual através de seu movimento.

Tanto a localização quanto a navegação do robô serão implementados como serviços no Espaço Inteligente. Isso dará ao sistema modularidade, escalabilidade e versatilidade, pois novos serviços e aplicações poderão ser adicionados ao sistema sempre que necessário, além da utilização imediata de serviços já existentes na arquitetura.

Assim, a solução oferecida para os clientes que quiserem implantar o sistema poderá ser fornecida a um custo menor, já que não será necessário investir em *hardware* dedicado, uma vez que o processamento poderá ser executado em nuvem.

Inicialmente, será realizada uma revisão bibliográfica de métodos de localização em sistemas multicâmeras [20] e métodos de detecção e reconhecimento de objetos [10]. Serão também implementados os *gateways* para o robô Lysa e as câmeras a serem utilizadas no sistema. Através dos *gateways*, esses novos elementos serão incorporados ao Espaço Inteligente e assim poderão ser acessados e controlados.

Como primeira abordagem, será aplicado um método de detecção que faz uso de marcadores visuais como o ArUco [24], Figura 2. Portanto, um marcador visual desse tipo será fixado na plataforma robótica para permitir sua localização em tempo real. A vantagem de se usar esse tipo de marcador está no fato de que seu formato e codificação permitem sua detecção e localização de forma robusta e precisa. Os pesquisadores envolvidos neste projeto já empregaram abordagem semelhante, onde robôs móveis puderam ser localizados e controlados satisfatoriamente [5]. Feito isso, os serviços de localização e controle do robô Lysa serão implementados e adicionados aos serviços já existentes do Espaço Inteligente.

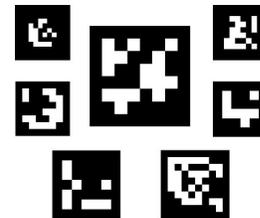


Figura 2: Marcadores ArUco

A seguir, serão estudados métodos de detecção e reconhecimento de objetos com objetivo de avaliar a possibilidade de se eliminar o uso de um marcador visual atrelado ao robô Lysa. Afinal, ter um marcador visual acoplado à plataforma pode ser esteticamente indesejável ou até mesmo comprometer o processo de localização do dispositivo.

Atualmente as técnicas mais promissoras de detecção e reconhecimento de objetos fazem uso de processamento através de várias camadas de uma rede neural convolucional (CNN)[12]. A camada final da CNN produz um valor probabilístico, correspondente a cada uma das possíveis categorias a qual o objeto pode ser classificado. Dentre os algoritmos mais utilizados para detecção de objetos pode-se listar o YOLOv3 [14], Mask R-CNN [11] e Faster R-CNN [22], os quais são capazes de realizar detecção de objetos em tempo real, de forma rápida e precisa. Técnicas como transfer learning [21] podem ainda ser aplicadas para adaptar reconhedores de objetos como YOLO, para um problema mais específico como a detecção do robô Lysa.

Dessa forma, será gerado um banco de imagens do robô Lysa, onde o robô deverá aparecer em diferentes posições e sob diferentes pontos de vista, tanto isoladamente quanto sendo utilizado por usuários. Essas imagens serão utilizadas para adaptação e treinamento de uma rede neural, a fim de que a detecção e reconhecimento da plataforma possa acontecer satisfatoriamente.

Assim como anteriormente, novos serviços para localização e controle do robô em tempo real serão implementados, agora usando-se o método de detecção e reconhecimento baseado em redes neurais profundas.

Em todas as etapas, os primeiros experimentos serão feitos considerando-se apenas um robô para que se obtenha uma avaliação técnica detalhada dos serviços criados. Depois outros robôs serão incluídos, para que o sistema seja avaliado quando vários usuários estiverem presentes e quando for necessário o controle de vários dispositivos ao mesmo tempo. A partir dos valores de precisão e desempenho alcançados, será possível avaliar se a localização e navegação dos robôs Lysa no ambiente poderá ser realizada usando-se ou não um marcador visual atrelado à plataforma. Ao final, espera-se obter uma solução tecnicamente e economicamente viável para a criação do sistema MobiLysa: um serviço de localização, navegação e controle do robô Lysa.

4 CONCLUSÃO

O sistema MobiLysa, apresentado neste projeto, traz uma proposta de solução para localização e navegação em ambiente internos que melhora a acessibilidade para as pessoas com deficiência visual, permitindo a sua condução a locais desejados por meio do uso do robô Lysa.

Essa é uma demanda constante recebida pela startup parceira neste projeto, VixSystem (proprietária e desenvolvedora do robô Lysa), por parte de clientes interessados na utilização do robô cão-guia nos seus ambientes. Tais clientes, representam espaços que desejam melhorar a acessibilidade e mobilidade de usuários com deficiência visual. Entretanto, há sempre o desejo que o robô seja capaz de guiar a pessoa até um determinado local, o que só é conseguido através de um sistema de localização e navegação.

Dessa forma, o MobiLysa se apresenta como um serviço interessante aos espaços públicos, com grande circulação de pessoas, com boas perspectivas de aceitação pelo público alvo e pela sociedade.

O MobiLysa encontra-se atualmente em desenvolvimento. A arquitetura do Espaço Inteligente a ser utilizada já foi definida, validada em outros projetos e será em seguida testada neste sistema. Além disso, algumas estratégias de localização e navegação foram implementadas e testadas com outras plataformas robóticas, as quais serão agora testadas e adaptadas ao robô Lysa. Essa validação técnica representa um dos próximos passos do projeto.

Além do desenvolvimento técnico do sistema, tem-se também como passo futuro tornar esse sistema economicamente viável para sua comercialização.

Existem ainda outras propostas de solução como óculos ou bonés inteligentes. Entretanto, estes dispositivos possuem apenas abrangência local e não oferecem ao usuário a autonomia de se dirigir a um local ainda desconhecido sem a necessidade do auxílio de um acompanhante. Por esses e outros motivos, acredita-se que o

sistema MobiLysa pode fazer grande diferença, contribuindo para maior mobilidade de seus usuários.

AGRADECIMENTOS

Este projeto conta com o suporte da RNP - Rede Nacional de Ensino e Pesquisa através do GT-MobiLysa aprovado no Edital de 2019. Agradecimentos também ao CNPq pelo concessão de bolsa de mestrado a um dos pesquisadores participante neste projeto.

REFERÊNCIAS

- [1] Conti A., Guerra M., Dardari D., Decarli N., and Win M.Z. 2012. *Network experimentation for cooperative localization*. IEEE. <http://ieeexplore.ieee.org/document/7762095/>.
- [2] Jaimes A. 2016. *Computer vision startups tackle AI*. IEEE.
- [3] Nieto A., Vilariño D.L., and Brea V.M. 2016. *PRECISION: a reconfigurable SIMD/MIMD coprocessor for computer vision systems-on-chip*. IEEE.
- [4] Yassin A., Nasser Y., Awad M., and et al. 2019. *Recent advances in indoor localization: a survey on theoretical approaches and applications*. IEEE. <http://ieeexplore.ieee.org/document/7762095/>.
- [5] do Carmo A.P., Vassallo R.F., de Queiroz F.M., and et al. 2019. *Programmable intelligent spaces for Industry 4.0: Indoor visual localization driving attocell networks*. Trans. Emerg. Tel. Technologies. <https://doi.org/10.1002/ett.3610>.
- [6] Almonfrey D., do Carmo A.P., de Queiroz F.M., R. Picoreti, Vassallo R.F., and E.O.T.m Salles. 2018. *A flexible human detection service suitable for intelligent spaces based on a multi-camera network*. International Journal of Distributed Sensor Networks. <https://doi.org/10.1177/1550147718763550>.
- [7] Dardari D., Closas P., and Djurić P.M. 2015. *Indoor tracking: theory, methods, and technologies*. IEEE.
- [8] ĀNurov D., HaluAka R., Hucec T., Puheim M., VaAjk J., and SinAjk P. 2017. *Intelligent space at center for intelligent technologies – system proposal*. Symposium on Applied Machine Intelligence and Informatics (SAMMI).
- [9] Instituto Brasileiro de Geografia e Estatística IBGE. 2011. *CENSO DEMOGRFICO 2010 - Características gerais da populao, religio e pessoas com deficincia*. Instituto Brasileiro de Geografia e Estatística - IBGE. ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Caracteristicas_Gerais_Religiao_Deficiencia/caracteristicas_religiao_deficiencia.pdf.
- [10] Singh H. 2019. *Practical Machine Learning and Image Processing: For Facial Recognition, Object Detection, and Pattern Recognition Using Python*. Apress.
- [11] Kaiming He, Georgia Gkioxari, Piotr Dollr, and Ross Girshick. 2017. *Mask R-CNN*. In *Proceedings of the International Conference on Computer Vision (ICCV)*.
- [12] Goodfellow I., Bengio Y., and Courville A. 2016. *Deep Learning*. MIT Press.
- [13] Fung J. and S. Mann. 2008. *Using graphics devices in reverse: GPU-based image processing and computer vision*. IEEE.
- [14] Redmon J. and Farhadi A. 2018. *Yolo V3: An incremental improvement*. arXiv.
- [15] Schlessman J. and Wolf M. 2015. *Tailoring design for embedded computer vision applications*. Computer.
- [16] Rampinelli M., Covre V.B., de Queiroz F.M., Vassallo R.F., Bastos-Filho T.F., and Mazo M. 2014. *An intelligent space for mobile robot localization using a multi-camera system*. Sensors. <http://www.mdpi.com/1424-8220/14/8/15039>.
- [17] Win M.Z., Gillford W.M., Conti A., and et al. 2011. *Network localization and navigation via cooperation*. IEEE.
- [18] Win M.Z., Gillford W.M., Conti A., and et al. 2018. *Efficient multisensor localization for the Internet of Things: exploring a new class of scalable localization algorithms*. IEEE.
- [19] Dua R., Raja A.R., and Kakadia D. 2014. *Virtualization vs containerization to support paas*. IEEE.
- [20] Raman R., Bakshi S., and Pankaj K. Sa. 2013. *Multi-camera localisation: a review*. International Journal of Machine Intelligence and Sensory Signal Processing. <https://dx.doi.org/10.1504/IJMISP.2013.052876>.
- [21] Shanmugamani R. 2018. *Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Keras*. Packt Publishing.
- [22] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. *Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks*. Cambridge, MA, USA.
- [23] Brena R.F., Garca-Vzquez J.P., Galvn-Tejada C.E., MuAoz-Rodrguez D., Vargas-Rosales C., and Fangmeyer J. 2017. *Evolution of indoor positioning technologies: a survey*. Sensors.
- [24] Garrido-Jurado S., MuAoz-Salinas R.M., Madrid-Cuevas F.J., and Marn-Jimnez M.J. 2014. *Automatic generation and detection of highly reliable fiducial markers under occlusion*. Pattern Recognition. Pattern Recognition.
- [25] VixSystem. 2016. *Quem  Lysa? VixSystem*. <http://www.caoguiarobo.com.br/>.