

SLAM EM AMBIENTES INTERNOS BASEADO NO RECONHECIMENTO VISUAL DE LUZES NO TETO

LUIZ HENRIQUE*, DIOGO PEDROSA*, ADELARDO MEDEIROS*

*DCA – CT – UFRN
Natal, RN, Brazil

Emails: luiz@dca.ufrn.br, diogo@dca.ufrn.br, adelardo@dca.ufrn.br

Abstract— The research about simultaneous localization and mapping (SLAM) has a great importance in robotics. Mobile robots need to know their position in workspace whatever be their applications. This article shows an approach of visual SLAM. The robot embedded camera is pointed to the ceiling. The ceil lamps have good properties like visibility and immobility therefore they were adopted as marks. The Extended Kalman Filter is applied to solve the SLAM problem. The article tries to explore the parallelism between the ceiling and floor, the scale invariance in mathematical modelling. To validate our SLAM approach a test is made.

Keywords— Robot Navigation and SLAM, Computer Vision, Localization, Ceiling Vision, Lamps, EKF.

Resumo— A pesquisa sobre localização e mapeamento simultâneos (SLAM em inglês) é de grande importância na robótica. Robôs móveis em quaisquer aplicações precisam conhecer seu posicionamento dentro de seu espaço de trabalho. Este artigo apresenta uma abordagem de SLAM visual. A câmera embarcada no robô é direcionada para o teto do ambiente. As lâmpadas de teto tem boas propriedades como visibilidade e imobilidade portanto foram adotadas como marcas. O Filtro Kalman Estendido é aplicado para solução do SLAM. O artigo procura explorar o paralelismo entre teto e chão como a invariância de escala durante a modelagem matemática. Para validação da SLAM proposto é realizado um teste real.

Palavras-chave— Navegação robótica, SLAM, Visão Computacional, Localização, Visão do Teto, Lâmpadas, EKF.

1 Introdução

Em robótica móvel, a tarefa de realizar simultaneamente a localização e a representação do ambiente de trabalho (mapa) é uma das mais importantes. Várias propostas já foram apresentadas para a resolução do SLAM, ou *Simultaneous Localization And Mapping* (Durrant-Whyte and Bailey, 2006a; Durrant-Whyte and Bailey, 2006b), em diversos tipos de ambientes nos quais os robôs atuam.

Atualmente, busca-se a aplicação de métodos de SLAM com sensores de baixo custo e, em algumas situações, estas soluções podem usar conhecimentos pré-existentes do ambiente de trabalho do robô, o que simplifica a execução do algoritmo. Uma abordagem interessante para o SLAM se dá por meio do uso de câmera monocular direcionada para o teto do ambiente a ser mapeado. Através das imagens, o robô pode capturar e extrair informações relevantes para utilizar como referência, em um mapa concorrentemente construído enquanto realiza seu processo de auto-localização dentro desta representação. Contudo, segundo (Hwang and Song, 2011), imagens de teto sofrem pela baixa quantidade de informações que podem ser adotadas como características em um mapa. Neste trabalho, optou-se por usar as luzes do teto em ambientes internos como marcas naturais para o robô. Em ambientes residenciais ou executivos, as luminárias são regularmente distribuídas no teto e as distâncias relativas entre elas são valores fixos.

Utilizar a câmera direcionada para o teto

evita variações de escala das características observadas. Outro ponto destacado é que a visão das luminárias do teto não é obstruída por obstáculos. A informação de distância existente entre o plano de imagem monocular e a marca não é utilizada pois adota-se a restrição de que o plano do chão (onde o robô navega) e o plano do teto (onde as marcas estão) são paralelos e têm distância conhecida. Como deseja-se apenas o conhecimento da postura do robô e da marca para fins de mapa de referências, alguns procedimentos comumente usados em visão computacional deixam de ser executados. Adota-se apenas os métodos necessários para processamento de imagens do teto e resolução do SLAM através de filtragem gaussiana.

Tratando-se de processos de reconhecimento de marcas para fins de localização, o trabalho de (Olivi et al., 2012) utiliza o reconhecimento visual de marcas (os quais são os nós de uma rede de sensores sem fio) postas no teto do ambiente por meio de processamento da imagem no domínio *hue, saturation, value* (HSV) com morfologia matemática. Um procedimento similar é feito por (Xu et al., 2009). Eles usam uma câmera monocular voltada para o teto a fim de realizar a extração de linhas regulares como características do mapa (o forro do teto possui um padrão quadrangular) e adotam procedimentos algébricos para determinar a postura do robô. O artigo de (Jeong and Lee, 2010) também procede com análise de imagens do teto do ambiente interno para SLAM. No entanto, os autores utilizam marcas mais generalistas, como sub-regiões da imagem com gradiente significativo. O rastreamento destas marcas é feito por

FKE (Filtro de Kalman Estendido) e a postura do robô por meio de um processo de votação (algoritmo de Hough).

O trabalho apresentado neste artigo é inspirado nos resultados de (Hwang and Song, 2011). Eles procederam com um algoritmo de SLAM monocular, com câmera voltada para o teto dos ambientes internos. Como marcas de referência para seu mapa, (Hwang and Song, 2011) utilizaram lâmpadas circulares, portas e cantos de parede no teto para promover uma navegação estável do robô no ambiente em questão. Já a proposta aqui adotada baseia-se apenas nas luzes acesas como marcas. A posição de cada marca é obtida por meio de processamento de imagem seguindo os passos utilizados por (Fabrizi and Saffiotti, 2000) e (Fabrizi and Saffiotti, 2002) para determinar posição e orientação da região de interesse. A pose do robô é atualizada através da aplicação do FKE com as informações das marcas encontradas e dos seus *encoders*.

Este artigo é estruturado da seguinte forma: a seção 2 mostra a estrutura do robô utilizada no trabalho; a seção 3 tem a descrição dos procedimentos adotados para o processamento das imagens das lâmpadas e a consequente extração das informações de interesse para o robô; Na seção 4 aborda-se o Filtro de Kalman Estendido, o qual foi implementado para proceder com a resolução do problema de SLAM tratado neste artigo. Resultados experimentais são apresentados na seção 5 e, por fim, considerações sobre os resultados e os métodos utilizados estão na seção 6.

2 Estrutura do Robô

O robô utilizado para este trabalho possui duas rodas motoras com *encoders* e uma terceira roda de apoio, cujos efeitos dinâmicos sobre o robô é desprezível. Cada roda motora tem um circuito para controle de posicionamento. Este dispositivo pode receber comandos como (a) girar a roda por uma distância especificada e (b) retornar o valor lido do *encoder*. O principal sensor embarcado é uma câmera com resolução 640×480 pixels com taxa de aquisição de 30 imagens por segundo. A câmera fica orientada para o teto para observação das lâmpadas, de tal forma que o plano da imagem seja o mais paralelo possível com o plano do teto (ver figura 1). O sistema de coordenadas do robô é baseado em seu modelo cinemático. O ponto de origem é exatamente o centro do eixo das rodas motoras. O eixo das abcissas aponta para a frente do robô enquanto que o eixo das ordenadas aponta para a roda esquerda. O sistema de coordenada da câmera tem a mesma orientação, mas apresenta uma translação na direção das abcissas.

O robô possui acionamento diferencial. A pose \mathbf{p} do robô é descrita por X_R , Y_R e θ_R . Após percorrer distâncias ΔE e ΔD com as rodas es-

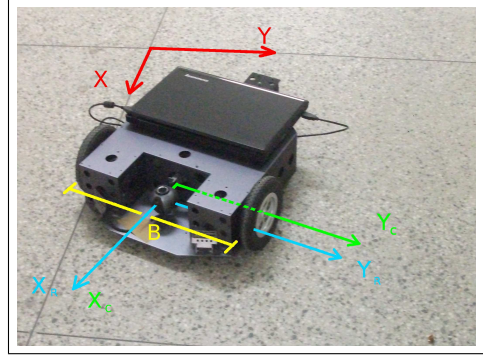


Figura 1: Plataforma robótica “Nomad”

querda e direita respectivamente, o robô terá uma nova pose descrita por X'_R , Y'_R e θ'_R . As duas poses do robô estão relacionadas pelo modelo cinemático $g(\mathbf{p}, \Delta E, \Delta D)$. O valor B é distância entre as rodas motoras. A equação 1 mostra o modelo matemático do robô.

$$g(\mathbf{p}, \Delta E, \Delta D) = \begin{bmatrix} x' \\ y' \\ \theta' \end{bmatrix} = \begin{bmatrix} x \\ y \\ \theta \end{bmatrix} + \begin{bmatrix} \Delta s \cos(\theta + \frac{\Delta\theta}{2}) \\ \Delta s \sin(\theta + \frac{\Delta\theta}{2}) \\ \Delta\theta \end{bmatrix} \quad (1)$$

$$\Delta s = \frac{\Delta E + \Delta D}{2} \quad \Delta\theta = \frac{\Delta D - \Delta E}{B}$$

3 Processamento das Imagens do Teto

O sensoreamento do ambiente é feito apenas com uma câmera para observação das marcas escolhidas (lâmpadas acesas). Estas marcas são de fácil visualização pois a sua imagem normalmente aparece saturada. Como consequência, tem-se uma região de brilho envolvendo a lâmpada sem uma definição clara do seu contorno como ilustrado na figura 2.

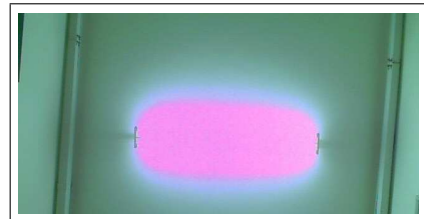


Figura 2: Uma luminária com duas lâmpadas tubulares.

Esta característica apresenta uma vantagem do trabalho pois simplifica o casamento de marcas durante a estimação do FKE. Alguns testes foram feitos para ajustar os parâmetros fundamentais da câmera e do processamento de imagens com o objetivo de melhorar a qualidade da detecção das marcas. Tais testes foram: definição do tamanho da máscara de suavização (parâmetro para o software de processamento), definição do limiar para

binarização (outro parâmetro para o software de processamento) e, por fim, o tempo de exposição (parâmetro da câmera).

A tabela 1 sintetiza os testes realizados com lâmpadas fluorescentes sem caixa de luminária que direcione o fluxo de luz. O comportamento de algumas propriedades da região brilhante foram analisadas quando cada um dos parâmetros citados anteriormente foi alterado no teste. Cada linha da tabela resulta da combinação de 800 amostras. A primeira coluna mostra os valores da máscara utilizada (em pixels) seguida do limiar (intensidade de pixel) e do tempo de exposição (s^{-1}). A segunda e a terceira coluna mostram as variâncias vertical e horizontal do ponto central da região (em pixel), a quarta coluna mostra a variância angular (radianos) da região, a quinta coluna exibe a área média da região (pixels) e a sexta coluna exibe o percentual da imagem que a marca ocupa. Os valores “X” e “E” representam situações onde ocorreu deterioração no processamento.

	Param	σ_v^2	σ_h^2	σ_a^2	área	%
1	5-150-800	5,4769	1,90687	3,952258E-005	51998,9	16,9
2	5-170-800	0,3356	0,17385	2,501294E-006	41883,4	13,6
3	5-200-800	X	X	X	X	
4	9-150-800	5,3620	0,635699	0,000015114	52604,7	17,1
5	9-170-800	0,2493	0,202125	2,154289E-006	40556	13,2
6	9-200-800	X	X	X	X	
7	13-150-800	5,2467	0,002478	1,039336E-005	52553,3	17,1
8	13-170-800	0,31480	0,051674	0,00000228	41278,6	13,4
9	13-200-800	X	X	X	X	
10	5-150-1200	0,15165	0,102197	0,00000551	41745,2	13,5
11	5-170-1200	E	E	E	E	
12	5-200-1200	X	X	X	X	
13	9-150-1200	0,10787	0,447198	7,503895E-006	41511,2	13,5
14	9-170-1200	E	E	E	E	
15	9-200-1200	X	X	X	X	
16	13-150-1200	0	0,034728	3,532696E-006	41698,7	13,5
17	13-170-1200	0	0	2,377779E-006	34223,2	11,1
18	13-200-1200	X	X	X	X	
19	5-150-1400	E	E	E	E	
20	5-170-1400	E	E	E	E	
21	5-200-1400	X	X	X	X	
22	9-150-1400	0,24873	0,243917	5,562473E-006	37898,4	12,3
23	9-170-1400	E	E	E	E	
24	9-200-1400	X	X	X	X	
25	13-150-1400	0,19698	0,0086201	4,585880E-006	38272,2	12,4
26	13-170-1400	0,24164	0,0721437	2,481770E-006	31308,9	10,1
27	13-200-1400	X	X	X	X	

Tabela 1: Propriedades geradas com 800 amostras

As linhas com “X” tiveram a região brilhante corrompida pelo alto nível do limiar de binarização, de forma que a região se fragmentou em pequenas outras regiões. As linhas com “E” tiveram a região brilhante corrompida pelo baixa dimensão da máscara de suavização da imagem, caracterizada por algumas fragmentações nas extremidades da região brilhante (coincidentes com a região de extremidade das lâmpadas).

Diante dos valores obtidos, foi escolhido trabalhar com os parâmetros da linha 17 por causa dos baixos valores apresentados (incluindo a área). Com os valores desta linha, foi formada a matriz de covariância das marcas $Q_{3 \times 3}$.

O processamento é feito com imagens em nível de cinza. Inicialmente, um filtro de suavização (filtro espacial passa-baixa) é aplicado na imagem capturada. Após isso, ocorre a binarização

da imagem, com limiar definido anteriormente, e a extração do contorno da região. A existência de contornos secundários internos ao contorno externo principal são descartados. Contornos que tangenciam os limites da imagem também são descartados. A partir deste ponto, contornos remanescentes são considerados regiões brilhantes legítimas prontas para análise. Os momentos inerciais de cada região são extraídos para que se possa calcular propriedades como coordenadas do centro da região i e j , o ângulo da região com relação à imagem, a excentricidade e a área da região. Este procedimento é similar ao trabalho (Fabrizi and Saffiotti, 2000) e (Fabrizi and Saffiotti, 2002), onde as mesmas informações são obtidas de regiões contínuas de uma grade de ocupação para extração de informações métricas locais.

Das propriedades encontradas, as coordenadas do centro da região e o ângulo são usadas no FKE. Para isso, é necessário transformá-las para o sistema de representação do mundo. As coordenadas i e j do centro da região passam por uma transformação de homografia. Define-se uma matriz $H_{3 \times 3}$ para mapear pontos do plano de imagem em pontos no plano do teto no referencial de câmera, obtendo assim os valores X_C e Y_C da marca (equação 2). Esta matriz é calculada por calibração (seção 5).

$$\begin{bmatrix} i \\ j \\ 1 \end{bmatrix} = H_{3 \times 3} \begin{bmatrix} X_C \\ Y_C \\ 1 \end{bmatrix} \quad (2)$$

A utilização de homografia exige que algumas considerações sejam adotadas: (a) o teto é plano; (b) o chão é plano; e (c) os dois planos são paralelos. Estas três considerações são válidas para grande parte dos ambientes internos.

Uma vez que a marca esteja no referencial de câmera, ela é facilmente transformada para o referencial do robô usando o operador de translação, definida em $[C]_{3 \times 1}$. Por último, a marca encontra-se no referencial de mundo pela transformação $[T]_{3 \times 3}$ que utiliza a pose atual do robô \mathbf{p} (equação 3).

$$\begin{bmatrix} X_m \\ Y_m \\ 1 \end{bmatrix} = [T]_{3 \times 3} \left([C]_{3 \times 1} + [H]_{3 \times 3}^{-1} \begin{bmatrix} i \\ j \\ 1 \end{bmatrix} \right) \quad (3)$$

Esta sequência de transformações pode ser invertida matematicamente, afim de se obter uma estimativa da pose da marca em relação à câmera, dada a sua posição no mundo. O vetor $\mathbf{m} = [X_m, Y_m, \theta_m]$ é a pose da marca em referencial de mundo.

$$\begin{bmatrix} i \\ j \\ 1 \end{bmatrix} = \begin{bmatrix} si(\mathbf{p}, \mathbf{m}) \\ sj(\mathbf{p}, \mathbf{m}) \\ 1 \end{bmatrix} = [H]_{3 \times 3} \left(-[C]_{3 \times 1} + [T]_{3 \times 3}^{-1} \begin{bmatrix} X_m \\ Y_m \\ 1 \end{bmatrix} \right) \quad (4)$$

Definindo a função matricial $h(\mathbf{p}, \mathbf{m})$ que retorna a posição (i, j) e a orientação θ da marca na imagem, tem-se:

$$h(\mathbf{p}, \mathbf{m}) = \begin{bmatrix} i \\ j \\ \theta \end{bmatrix} = \begin{bmatrix} si(p, m) \\ sj(p, m) \\ \theta_M - \theta_R \end{bmatrix} \quad (5)$$

Dessa função deriva-se então a matriz Jacobi-ana, denotada por $J_{3 \times 6}$, que será utilizada no Filtro de Kalman Estendido (resolução do SLAM).

4 Aplicação FKE

O Filtro de Kalman Estendido foi implementado com auxílio da biblioteca *OpenCV*. A etapa de predição incorpora o modelo cinemático do robô. Já a etapa de atualização é baseada no modelo de observação da câmera. Então, para cada distância percorrida pelas rodas, realiza-se a predição da nova pose do robô, seguida da observação. Caso alguma marca já armazenada no mapa seja vista, a pose do robô e a pose da marca têm seus valores atualizados pela nova observação, encerrando-se, assim, o processo de estimação. A pose do robô e de cada marca são armazenadas no vetor de estado \mathbf{x} do filtro (equação 6). Inicialmente contém apenas a pose do robô, com a exploração do ambiente marcas novas são incluídas neste vetor. Tanto a pose do robô como a pose de cada marca estão em referencial de mundo.

$$\mathbf{x} = [X_R, Y_R, \theta_R, X_{m_1}, Y_{m_1}, \theta_{m_1}, \dots, X_{m_n}, Y_{m_n}, \theta_{m_n}]^T \quad (6)$$

O filtro pode ser visto como uma função recursiva com argumentos de entrada e com retorno de valores. Os principais argumentos do filtro implementado são o vetor de estado da iteração anterior, a matriz de covariância correspondente, as distâncias percorridas por cada roda do robô desde o estado anterior, a lista das marcas observadas durante este percurso e um vetor de casamento para as marcas vistas. O filtro retorna o novo estado estimado e sua matriz de covariância associada.

A matriz de covariância da pose do robô $R_{3 \times 3}$ (equação 8) é obtida pelo mapeamento da matriz de covariância das distâncias trafegadas pelas rodas, $M_{2 \times 2}$, para o espaço de estados. O mapeamento é feito com o Jacobiano de $g(\mathbf{p}, \Delta E, \Delta D)$

com relação às distâncias ΔE e ΔD , o qual é denotado pela matriz $V_{3 \times 2}$. A definição da matriz $M_{2 \times 2}$ (equação 7) envolve duas constantes K_L e K_R , estimadas por testes descritos na seção de 5.

$$M = \begin{bmatrix} |\Delta E|K_L & 0 \\ 0 & |\Delta D|K_R \end{bmatrix} \quad (7)$$

$$R = V \cdot M \cdot V^T \quad (8)$$

5 Experimentos com Robô

Para a realização dos testes com o FKE foi necessário a definição da matriz de covariância do robô $M_{2 \times 2}$ e da matriz de covariância das marcas $Q_{3 \times 3}$. Para esta segunda matriz, os valores utilizados foram da linha escolhida na tabela 1.

$$Q = \begin{bmatrix} \sigma_i^2 & 0 & 0 \\ 0 & \sigma_j^2 & 0 \\ 0 & 0 & \sigma_\theta^2 \end{bmatrix} \quad (9)$$

Na matriz $M_{2 \times 2}$, as constantes K_L e K_R foram obtidas de maneira experimental. O procedimento consistiu em girar a roda direita por uma determinada distância e, depois, comparar a distância real percorrida com a distância calculada pela odometria. Foram feitas 20 repetições deste procedimento. A variância dos erros é igualada a $|\Delta D|K_R$, onde ΔD é a distância estipulada. Nesta equação tem-se K_R de incógnita. Assim, o valor K_R é facilmente encontrado. O valor de K_L foi calculado de forma análoga. A distância real percorrida por cada roda foi estimada com auxílio da estrutura de visualização da equipe POTI-UFRN de futebol de robôs (ver figura 3).



Figura 3: Estrutura utilizada para obter a covariância das rodas do robô.

A definição da matriz homografia $H_{3 \times 3}$ precede o experimento com o filtro e foi feita com associação de pixels com pontos conhecidos do

mundo. O robô foi posicionado de tal forma que a câmera observou alguns pontos notáveis no teto, cujas distâncias em relação à câmera eram conhecidas. Estas associações são descritas por um conjunto de equações, onde as suas incógnitas formam a matriz $H_{3 \times 3}$. Quando o sistema de equações é solucionado, definem-se os valores da homografia. Esta matriz representa o mapeamento de pontos do teto para a imagem. A sua matriz inversa $H_{3 \times 3}^{-1}$ realiza o mapeamento inverso, ou seja, pontos do teto são convertidos em pixels na imagem. Uma vez realizada esta calibração de câmera, pode-se aplicar o FKE.

O teste para validar o SLAM proposto foi realizado no corredor do prédio do Departamento de Engenharia de Computação (DCA) da UFRN. O robô foi manualmente controlado para percorrer um caminho com formato de “U” (figura 4).



Figura 4: Corredor usado para validação do SLAM proposto.

Durante o percurso, o robô observou três lâmpadas. Na terceira lâmpada, o robô realizou um giro aproximado de 180° , acionando apenas a roda esquerda. Depois seguiu em linha reta até retornar às proximidades da posição de partida. O caminho percorrido é apresentado nas figuras a seguir. São apresentados os resultados da odometria (figura 5) e do SLAM (figura 6).

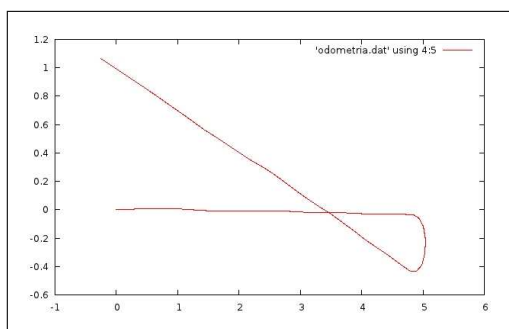


Figura 5: Trajeto estimado por odometria

6 Conclusões

Verificou-se que as luminárias utilizadas no ambiente de testes do robô possuem duas lâmpadas

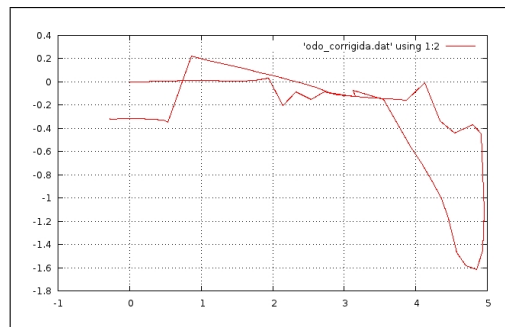


Figura 6: Trajeto estimado por FKE

fluorescentes. Durante o desenvolvimento inicial do trabalho, procedeu-se com ajustes manuais nos parâmetros da câmera para detectar o formato de cada uma das lâmpadas tubulares instaladas no suporte. A ideia original era que o formato da lâmpada pudesse ser usado para extração dos parâmetros de interesse. No entanto, percebeu-se que trabalhar com as *manchas luminosas* fez com que toda a informação necessária para o SLAM pudesse ser mais facilmente obtida.

Para estes resultados preliminares, a associação das marcas foi realizada manualmente. Em breve, espera-se melhorar o programa de resolução de SLAM usado para o robô e ambiente apresentados neste artigo de forma que esta associação ocorra automaticamente.

Com relação aos parâmetros da marca, alguns deles não foram usados ainda para este artigo (excentricidade e a área da região branca). Também espera-se obter uma melhoria do algoritmo de SLAM ao incluir estes parâmetros para fins de associação entre marcas vistas e marcas já armazenadas no mapa. Outras abordagens também poderão ser testadas, como as que se baseiam em medidas estatísticas (distância de Mahalanobis, por exemplo).

Mais testes precisam ser realizados com o robô. Verificou-se que a rotação feita implicou erros no SLAM, principalmente na pose do robô.

Agradecimentos

Agradecemos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pelo incentivo financeiro, aos professores do laboratório de robótica e também aos colegas de laboratório pelas discussões de idéias.

Referências

- Durrant-Whyte, H. and Bailey, T. (2006a). Simultaneous localization and mapping: Part I, *IEEE Robotics & Automation Magazine* pp. 99–108.
- Durrant-Whyte, H. and Bailey, T. (2006b). Simultaneous localization and mapping (SLAM):

Part II, *IEEE Robotics & Automation Magazine* pp. 108–117.

- Fabrizi, E. and Saffiotti, A. (2000). Extracting topology-based maps from gridmaps, *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.
- Fabrizi, E. and Saffiotti, A. (2002). Augmenting topology-based maps with geometric information, *Robotics and Autonomous Systems* **40**(2): 91–97.
- Hwang, S.-Y. and Song, J.-B. (2011). Monocular vision-based SLAM in indoor environment using corner, lamp, and door features from upward-looking camera, *IEEE Transactions on Industrial Electronics* **58**: 4804–4812.
- Jeong, W. Y. and Lee, K. M. (2010). CV-SLAM: A new ceiling vision-based SLAM technique, *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*.
- Olivi, L., Souza, R., Paolieri, F., Guimarães, E. and Cardozo, E. (2012). A distributed navigation strategy for mobile robots based on wireless sensor networks, *Anais do XIX Congresso Brasileiro de Automática, CBA 2012*.
- Xu, D., Han, L., Tan, M. and Li, Y. F. (2009). Ceiling-based visual positioning for an indoor mobile robot with monocular vision, *IEEE Transactions on Industrial Electronics* **56**: 1617–1628.