

## **CONSTRUÇÃO DE MAPAS LOCAIS 3-D UTILIZANDO VISÃO MONOCULAR PARA FINS DE NAVEGAÇÃO DE ROBÔS MÓVEIS AUTÔNOMOS**

**Guilherme Leal Santos, gsantos@dca.ufrn.br<sup>1</sup>**

**Pablo Javier Alsina, pablo@dca.ufrn.br<sup>1</sup>**

<sup>1</sup>Departamento de Engenharia de Computação e Automação, Centro de Tecnologia, Universidade Federal do Rio Grande do Norte, CEP 59078-900 - Natal - RN - Brasil

**Resumo.** A localização e mapeamento simultâneos (*Simultaneous Localisation and Mapping - SLAM*) é uma abordagem em que um robô móvel autônomo utiliza observações de marcas no ambiente para construir um mapa local e para localizar-se. O uso de uma única câmera como sensor primário no SLAM possui algumas particularidades, pois é um desafio determinar a profundidade de uma característica utilizando visão monocular, por haver somente medidas 2-D de estruturas 3-D. Neste trabalho é apresentado uma técnica de SLAM visual monocular, que utiliza características encontradas em uma sequência de imagens através do descritor SURF (*Speeded Up Robust Features*) para determinar quais características podem ser utilizadas como marcas. Para iniciar o mapeamento e localização, é necessário primeiro determinar a posição das marcas presentes no ambiente. Para isso, tem-se disponível apenas as coordenadas  $(u, v)$  das características detectadas na imagem e os parâmetros intrínsecos e extrínsecos da câmera. Assim, para a inicialização das marcas, foi utilizada a técnica de inicialização atrasada. Com a posição estimada das marcas, foi possível estimar a posição atualizada do robô a cada passo; obtendo resultados satisfatórios que comprovam a eficiência do sistema de SLAM visual monocular proposto neste trabalho.

**Palavras-chave:** Mapas Locais 3-D; Speed up Robust Features (SURF); Visão Monocular; Simultaneous Localisation and Mapping (SLAM)

### **1. INTRODUÇÃO**

A visão é uma fonte sensorial extremamente importante tanto para humanos quanto para robôs, pois fornece informações detalhadas sobre o ambiente. Um sistema de visão robusto deve ser capaz de detectar objetos e fornecer uma representação adequada do mundo para processos de nível mais alto. O sistema de visão deve ser também altamente eficiente, para responder rapidamente a uma mudança de ambiente.

A localização e mapeamento simultâneos é o problema em que o robô é capaz de construir um mapa de um ambiente desconhecido e simultaneamente localizar-se no mesmo. Esta habilidade torna o robô verdadeiramente autônomo. Enquanto realiza o SLAM, o robô observa o seu ambiente de navegação e detecta a posição de alguns pontos de interesse. Alguns destes pontos de interesse são utilizados como *marcas* para o processo de SLAM.

A estimação da posição destas marcas constitui a parte de mapeamento do processo de SLAM. A medida que o robô se move, novas marcas são observadas e então é realizada uma correlação com as marcas já encontradas anteriormente e a diferença entre o valor esperado e o valor medido é utilizado para estimar a posição do robô. Esta parte é a etapa de localização do processo de SLAM.

Na literatura, o telêmetro *laser* e o sensor sonar têm sido utilizados para a percepção do ambiente e assim realizar o SLAM. Entretanto, a situação mudou rapidamente durante a última década pois uma quantidade considerável de pesquisas foram direcionadas para a utilização de sensores visuais para o SLAM. Estes sensores proveem uma rica quantidade de informações sobre o ambiente, possibilitando a detecção de pontos de interesse. Por isso, é muito interessante a utilização de câmeras como sensor principal de um sistema de SLAM para um robô móvel autônomo.

O desenvolvimento e aperfeiçoamento de técnicas que façam simultaneamente o mapeamento e a localização de um robô móvel autônomo e a criação de mapas locais 3-D, a partir de uma sequência de imagens, é bastante estudado no meio científico. A principal razão é a acessibilidade, porque o uso de um sistema de visão é uma alternativa barata em relação ao uso de outros tipos de sensores, como sonar, sensor infravermelho, telêmetro laser (preços mais elevados). A captura de imagens pode ser realizada por uma simples webcam. Além do baixo custo, a quantidade de informações úteis

fornecidas por uma sequência de imagens é grande.

## 2. TÉCNICAS DE SLAM VISUAL

A visão é a fonte de informação mais rica de nosso ambiente. Por isso que os algoritmos de SLAM começaram a ser usados com informações visuais Artieda *et al.* (2009). A informação fornecida por sistemas de visão consiste em uma vasta quantidade de dados a serem processados para que os algoritmos de SLAM sejam providos de informações úteis. Portanto, os algoritmos de processamento de imagens tem grande importância nos resultados e no sucesso da resolução do problema de SLAM visual.

Os telêmetros *lasers* são sensores ativos de grande precisão, mas a sua característica de medição ponto-a-ponto limita a sua capacidade de reconhecer e seguir objetos. Os sistemas de localização baseados em sonar são rápidos, baratos e tem uma capacidade de reconhecer e seguir objetos semelhantes à dos sistemas que se baseiam na visão, mas tendem a produzir medições menos precisas e mais suscetíveis ao ruído. Os sensores inerciais, assim como os odômetros, são úteis porém o acúmulo de erros de medição geram grande influência nas estimativas de posição posteriores.

Entretanto, os sistemas de localização baseados em visão são baratos, pesam pouco e permitem medições de grande alcance e com grande resolução Chen *et al.* (2007). A visão é muito utilizada nas técnicas de percepção acopladas ao SLAM porque existem técnicas eficazes de visão computacional que consistem em extrair e comparar características visuais. Uma das limitações para o uso de um sistema de visão é o elevado esforço computacional associado, tornando difícil a criação de sistemas para funcionamento em tempo real.

A maioria das técnicas de SLAM visual existentes na literatura podem ser divididas em quatro etapas: percepção do ambiente utilizando um sensor visual, extração e seleção de pontos de interesse, inicialização de marcas, implementação de um algoritmo de estimação. Nas seções a seguir, serão descritas as principais características destas partes.

### 2.1 Sistema de visão

Quanto à etapa de percepção do ambiente, o SLAM visual pode ser dividido em sistemas que utilizam somente uma câmera (monocular) ou com duas ou mais câmeras (estéreo).

#### 2.1.1 Monocular

Os sistemas de visão monocular utilizam somente uma câmera para capturar as imagens do ambiente, não fornecendo portanto a informação de profundidade. O SLAM visual monocular foi apresentado por Davison *et al.* (2007), onde o sistema proposto faz uma reconstrução de ambientes internos. Em [Artieda et al. 2009] foi apresentado um sistema que utiliza a técnica de SLAM visual 3-D para VANTs (Veículos Aéreos Não- Tripulados) onde são detectadas características em um ambiente parcialmente estruturado e depois são calculadas as distâncias até o VANT. O SLAM visual monocular também obteve bons resultados nos trabalhos de Sim *et al.* (2005); Mouragnon *et al.* (2006); Lemaire and Lacroix (2007a); Zhang *et al.* (2008).

#### 2.1.2 Estéreo

O SLAM visual estereo utiliza duas ou mais câmeras para capturar as imagens do ambiente, e através da técnica de triangulação de pontos é possível obter a informação de profundidade. Entretanto, o problema da correlação de características é mais complicado que no caso do SLAM visual monocular, porque há um aumento na complexidade computacional, dado que as características tem que ser correlacionadas entre as duas imagens do par estereo e entre as aquisições consecutivas no tempo Lemaire *et al.* (2007b).

Em Salvi *et al.* (2008) o SLAM visual foi implementado em um AUV (*Autonomous Underwater Vehicle*) equipado com um sistema de câmeras estereo; onde o sistema proposto fornece como saída uma aquisição 3-D em larga escala do ambiente navegado. Lemaire *et al.* (2007b) implementou um sistema de SLAM visual estereo para robôs terrestres e aéreos.

### 2.2 Extração de pontos de interesse

Em um sistema de SLAM visual, as características tem que ser extraídas do ambiente para serem utilizadas como marcas. Estas características devem ser estáveis e observadas de diferentes pontos de visão e ângulos. Mouragnon *et al.* (2006) utilizou o detector de cantos de Harris em um sistema de SLAM onde o objetivo é encontrar a posição e orientação em um sistema de referência global da câmera em tempos  $t$ , assim como a posição 3-D das características vistas ao longo da cena. Davison *et al.* (2007) utilizou o detector de Shi e Tomasi.

Salvi *et al.* (2008) criou um sistema de SLAM visual onde as características são extraídas com um algoritmo misto de SIFT e SURF sincronizados, para assim obter imagens pré-processadas com um equilíbrio entre quantidade, robustez e complexidade computacional. Em Zhang *et al.* (2008) foi proposto um sistema de SLAM visual monocular que detecta

características no ambiente utilizando um extrator SURF (*Speeded Up Robust Feature*), e utiliza um FKE (Filtro de Kalman Estendido) para calcular os estados da câmera e das marcas visuais encontradas. Sünderhalf *et al.* (2007) também utilizou o detector SURF para obter os pontos de interesse.

### 2.3 Inicialização das marcas

Após a etapa de extração dos pontos de interesse na imagem, parte desses pontos serão determinados como marcas que serão utilizadas pelo sistema de SLAM visual. Existem duas técnicas principais para realizar tal inicialização, que serão mostradas a seguir.

#### 2.3.1 Inicialização direta

Para conseguir informações como direção e profundidade de características logo na inicialização do sistema de SLAM, uma alternativa é utilizar visão estéreo. Lemaire *et al.* (2007b) apresentou um sistema SLAM visual que utiliza visão estéreo para robô terrestres e aéreos. Uma outra forma de fazer uma inicialização direta de marcas é usar um alvo artificial de aparência conhecida para inicializar o sistema de SLAM.

Davison *et al.* (2007) usa um retângulo sólido impresso em um papel como alvo para inicialização. Civera *et al.* (2008) mostrou um sistema de SLAM visual 3-D totalmente automático que utiliza somente uma câmera hand-held sem sensores adicionais. Os experimentos mostraram a robustez do sistema em ambientes internos e externos com um alcance alto para profundidade da cena, movimento variado e também um giro 360° em tempo real.

#### 2.3.2 Inicialização atrasada

Quando utiliza-se um sistema de SLAM visual monocular, não é possível determinar a profundidade das características com a primeira imagem adquirida. Para isso, a câmera deve mover-se para diferentes pontos de visão e utilizar a correlação de características entre diferentes quadros, para assim utilizar alguma técnica para a inicialização das marcas para o sistema de SLAM. Os trabalhos de Kim and Sukkarieh (2003); Bryson and Sukkarieh (2005) utilizaram inicialização atrasada.

Zhang *et al.* (2008) utiliza um filtro de partículas para inicializar as marcas. Partículas são distribuídas com a mesma prioridade em uma linha reta 3-D que passa pela característica encontrada, denominadas de marcas virtuais, onde cada uma destas possui uma matriz de covariância. Esta matriz de covariância define uma área elíptica para pesquisa na imagem. Nos passos seguintes, a prioridade das partículas é alterada calculando-se a verossimilhança de cada partícula. Finalmente, a marca virtual com maior probabilidade é transformada em uma marca real.

### 2.4 Técnicas de SLAM

Existem diferentes técnicas que foram criadas para solucionar o problema de SLAM. A seguir, serão apresentadas as técnicas de filtro de Kalman estendido e filtro de partículas, e outras técnicas.

#### 2.4.1 SLAM com Filtro de Kalman Estendido - FKE SLAM

A utilização do algoritmo de SLAM baseado no filtro de Kalman estendido é uma das alternativas mais utilizadas para solucionar o problema de SLAM. O FKE estima a posição do robô através da observação de características presentes no ambiente. As características a serem utilizadas pelo FKE neste trabalho são detectadas e rastreadas, utilizando a técnica SURF Bay *et al.* (2006).

O FKE utilizado para solucionar o problema de SLAM está sujeito a muitos dos problemas da solução do FKE para navegação de robô móveis como: complexidade computacional, associação dos dados e não-linearidades. Sob condições ideais para o FKE SLAM, a covariância da estimativa da localização do robô e as posições individuais das marcas convergiram para zero. Contudo a complexidade computacional da etapa de correção cresce quadraticamente com o número de marcas Sim *et al.* (2005).

A técnica de FKE SLAM é extremamente sensível à associação incorreta das marcas com as novas observações. O problema da associação dos dados é muito difícil para o SLAM, no caso em que as referências são re-observadas por pontos de observação distintos; e a linearização do modelo não-linear do movimento pode potencializar resultados com soluções inconsistentes Thrun *et al.* (2006).

O filtro de Kalman estendido é uma das técnicas de SLAM mais utilizadas na literatura. Algumas implementações de FKE para SLAM visual podem ser encontradas em Davison *et al.* (2007); Lemaire and Lacroix (2007a); Lemaire *et al.* (2007b); Zhang *et al.* (2008); Salvi *et al.* (2008); Artieda *et al.* (2009).

### 2.4.2 SLAM com Filtro de Partículas - FP SLAM

O filtro de partículas foi aplicado com sucesso para a resolver o problema de SLAM. O algoritmo FastSLAM, introduzido por Montemerlo *et al.* (2002), utilizou um filtro de partículas para estimar a pose do robô e um filtro de Kalman estendido para estimar a localização das marcas. Este algoritmo é baseado no problema de que a posição do robô é conhecida, e a localização das marcas é independente.

Para um estado do robô no tempo  $t$ , é calculado para cada partícula o novo estado do robô usando o estado no tempo  $t-1$  e a entrada de controle  $u_t$ . Isto cria um conjunto temporário de partículas, cada uma recebe um peso diferente, que representa o estado do robô no instante  $t$ . Para cada partícula, cada marca é atualizada utilizando um filtro de Kalman estendido.

Os trabalhos de Nieto *et al.* (2003); Montemerlo *et al.* (2002); Hähnel *et al.* (2003); Sim *et al.* (2005); Elinas *et al.* (2006) utilizaram um sistema de SLAM visual baseado no filtro de partículas. Zhang *et al.* (2008) utilizou um filtro de partículas para a inicialização atrasada de marcas visuais.

### 2.4.3 Outras técnicas

No trabalho de Sünderhalf *et al.* (2007) foi apresentado um sistema de SLAM visual monocular que utiliza a técnica de UKF (*Unscented Kalman Filter*) com uma parametrização de profundidade invertida para controle de dirigíveis autônomos. O UKF realiza uma linearização estocástica através do uso de um processo de regressão linear estatística utilizando pesos. Ao invés de aproximar uma função  $g$  através de uma expansão de série de Taylor, como no filtro de Kalman estendido, o filtro de Kalman descentrado extrai deterministicamente os pontos chamados *sigma* de Gaussianas e passam estes através de  $g$ . Estes pontos sigma são localizados na média e simetricamente no eixo central de covariância (dois por dimensão). Para maiores informações consultar Thrun *et al.* (2006).

## 3. MÉTODO PROPOSTO

Para iniciar o mapeamento e localização, é necessário primeiro determinar a posição das marcas presentes no ambiente. Para isso, tem-se disponível apenas as coordenadas  $(u, v)^T$  das características detectadas na imagem e os parâmetros intrínsecos e extrínsecos da câmera. É possível determinar a posição das marcas somente com a disponibilidade da profundidade.

Com a utilização de um sistema de visão monocular, não é possível obter a profundidade de forma direta, como em um sistema de visão estéreo. Para poder estimar a profundidade de uma característica, são necessárias informações sobre o movimento da câmera e de várias medidas da mesma marca visual realizadas de diferentes pontos de vista. Entretanto, não é adequado demorar muitos quadros de imagem para poder inicializar uma marca. Esta técnica é denominada de *inicialização atrasada de marcas*.

Nas técnicas propostas por Davison *et al.* (2007); Zhang *et al.* (2008), um filtro de partículas é utilizado para calcular a profundidade da marca. Um conjunto de partículas são linearmente espaçadas ao longo de uma linha 3-D entre o centro ótico da câmera e o ponto  $(X, Y, Z)^M$  no mundo que corresponde à característica  $(u, v)^T$  da imagem, com uma distância máxima estabelecida em metros. Todas as partículas são inicializadas com a mesma probabilidade. A probabilidade para cada uma dessas partículas é atualizada no próximo passo, quando é feito uma estimativa de onde tal característica  $(u, v)^T$  apareceu na imagem anterior. A Figura (1) mostra o ambiente utilizado para testar esta técnica.

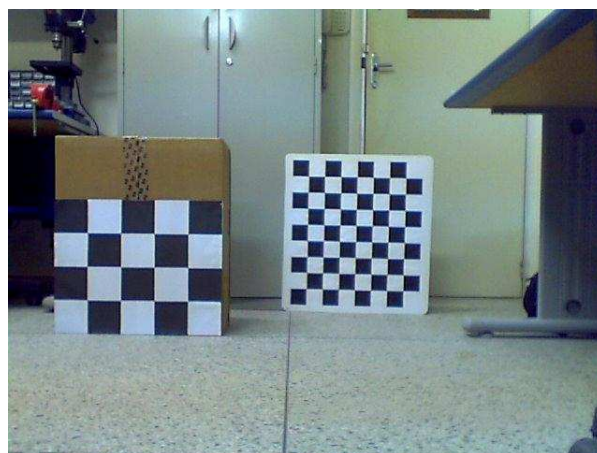


Figura 1. Ambiente de teste da técnica de inicialização atrasada.

Se a probabilidade das partículas formarem uma distribuição gaussiana aproximada, a partícula com maior probabilidade é tornada uma marca. Obtendo-se as marcas no ambiente, é verificado se esta é uma marca nova ou uma marca já conhecida. Se for uma marca nova, é adicionada ao vetor de estados do FKE. Senão, é atualizada a pose do robô e das marcas. A Figura (2) mostra o resultado obtido para uma determinada marca, onde a distância entre  $(X, Y, Z)$  real e o  $(X, Y, Z)$  calculado de uma marca foi de  $0,52m$ , com um erro de  $5,2\%$ .

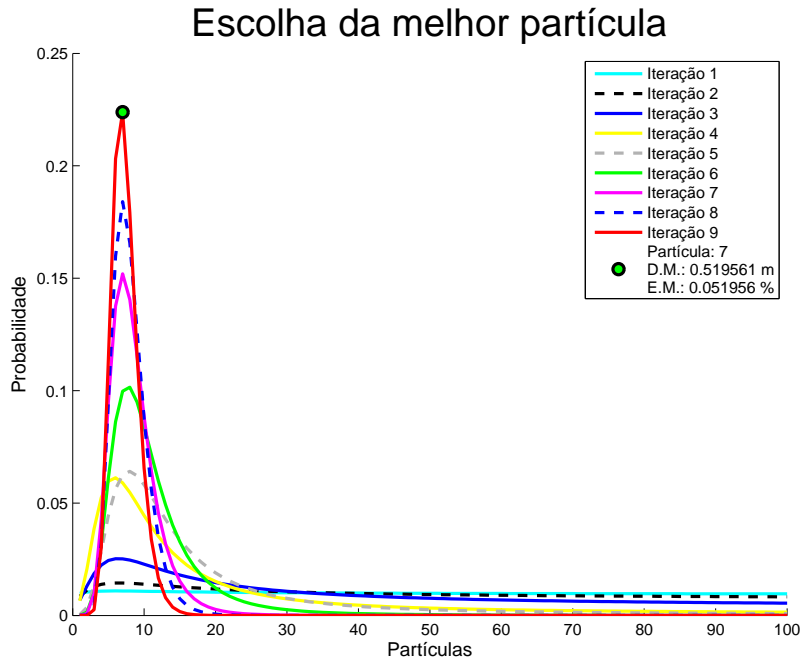


Figura 2. Probabilidade das partículas após 9 iterações, calculada utilizando 100 partículas que estimam a profundidade de uma marca em até 10 metros.

A inicialização atrasada utilizando partículas fornece resultados satisfatórios para o SLAM visual, porém pode apresentar os seguintes problemas:

- Como o tempo computacional é da ordem de  $n$  (número de partículas), se o número de partículas aumentar, o esforço computacional também cresce;
- Se existir a necessidade de identificar marcas a grandes distâncias (50 metros, por exemplo), existem dificuldades porque estas marcas irão demorar para serem adicionadas ao vetor de estados ou até mesmo não irão ser detectadas por possuírem uma baixa parallax;
- Nesta técnica há a necessidade de inicialização prévia de parâmetros, como o número de partículas e a distância máxima de alcance de marcas. Ou seja, para um funcionamento correto da técnica de SLAM visual, deve-se informar parâmetros de acordo com o ambiente de navegação do robô.

O método proposto neste trabalho calcula a posição das marcas com inicialização atrasada, porém com uma maior rapidez devido a estimação da profundidade ser feita de forma analítica. No passo inicial, é calculada a equação paramétrica de uma reta 3-D  $R_1$  que passa pela marca e pelo centro ótico da câmera. Nos passos seguintes, são obtidos os novos parâmetros das retas 3-D  $R_N$  que passam pela mesma marca, e a partir do método dos mínimos quadrados encontram-se os pontos  $P_1, P_2, \dots, P_N$ , que representam a menor distância entre estas retas. O ponto a ser inicializado como marca é o ponto médio entre os pontos encontrados.

A técnica de inicialização proposta neste trabalho resolve estes problemas existentes na inicialização atrasada com partículas, porque não há a necessidade de uma inicialização prévia de parâmetros como número de partículas e distância máxima de alcance. A Figura (3) mostra como as retas são representadas nas coordenadas de mundo, e mostra que para a marca para a qual foi estimada a profundidade pelo método das partículas, a distância média entre  $(X, Y, Z)$  real e o  $(X, Y, Z)$  calculado foi de  $0,14m$ .

Os testes foram realizados com um robô móvel que possui dois encoders acoplados aos eixos dos motores e uma câmera dedicada para o sistema de localização. Os dados obtidos com o FKE são utilizados para atualizar a pose do robô e são comparados aos dados fornecidos pela odometria do robô móvel, obtida com os dados fornecidos pelos encoders óticos. Durante o percurso, o robô móvel detecta a presença de características nas imagens e, através do filtro de partículas,

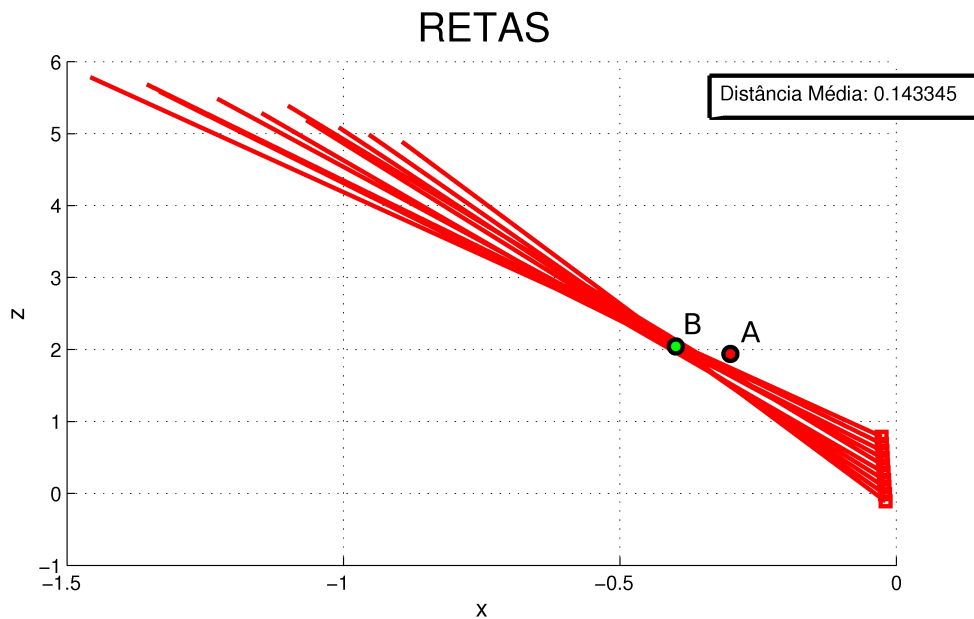


Figura 3. Feixe de retas 3-D traçadas a partir do centro óptico da câmera, em direção a uma marca visual encontrada no ambiente. O ponto A corresponde à posição real da marca, e o ponto B corresponde à posição calculada.

adiciona centenas de novas marcas ao vetor de estados do FKE após estimar a profundidade das características. Para algumas imagens não foi possível detectar e rastrear características, devido ao ruído nas imagens causado pela própria movimentação do robô e devido às pequenas variações de iluminação. Com a posição estimada das marcas, foi possível estimar a posição atualizada do robô a cada passo; obtendo resultados satisfatórios que comprovam a eficiência da técnica de inicialização atrasada proposta neste trabalho.

#### 4. CONCLUSÃO

Neste trabalho foi apresentada uma técnica para construção de mapas locais 3-D, a partir do ambiente de navegação de um robô móvel, utilizando retas 3-D para estimar a profundidade de marcas visuais. Esta técnica comprovou ser mais eficiente computacionalmente que as técnicas de inicialização atrasadas que utilizam partículas para estimar a profundidade de uma marca visual. Também foi obtido êxito na sua utilização em um sistema de SLAM visual monocular para um robô móvel autônomo.

Na introdução, foi mostrado porque os sistemas de SLAM visual e a criação de mapas locais 3-D são bastante estudados no meio científico. Logo após, foram mostradas as principais características das técnicas de SLAM visual existentes na literatura, dando ênfase às suas partes principais: percepção do ambiente utilizando um sensor visual, extração de pontos de interesse, seleção de pontos de interesse e a inicialização como marcas a serem utilizadas pelo SLAM, implementação de um algoritmo de estimação. Por fim, foi apresentada a técnica de inicialização atrasada de marcas utilizando retas 3-D.



Figura 4. Robô móvel *Karel*

Para a implementação, foram utilizadas as funções disponíveis nas bibliotecas *OpenCV* e *OpenSURF*, com linguagem C++. Como suporte ao método implementado, foi utilizado para a navegação o robô móvel *Karel* (Figura (4)), considerado como um corpo rígido, sem peso e capaz de se mover num plano horizontal respeitando as restrições associadas à não holonomicidade.

## 5. REFERÊNCIAS

- Artieda, J., Sebastian, J.M., Campoy, P., Correa, J.F., Mondragón, I.F., Martínez, C. and Olivares, M., 2009. “Visual 3-D SLAM from UAVs”. *Journal of Intelligent Robots and Systems*, , No. 55, pp. 299–321.
- Bay, H., Ess, A., Tuytelaars, T. and Gool, L.V., 2006. “Surf: Speeded-up robust features”. *Lecture notes in computer science*, Vol. 3951, No. 404.
- Bryson, M. and Sukkarieh, S., 2005. “Bearing-only SLAM for an airborne vehicle”. *Australasian Conference on Robotics and Automation*. Sidney, Austrália.
- Chen, Z., Samarabandu, J. and Rodrigo, R., 2007. “Recent advances in simultaneous localization and mapbuilding using computer vision”. *Advanced Robotics*, Vol. 3, No. 4.
- Civera, J., Davison, A.J. and Montiel, J.M.M., 2008. *IEEE Transactions on Robotics*, Vol. 24, No. 5, pp. 932–945.
- Davison, A., Reid, I., Molton, N. and Stasse, O., 2007. “MonoSLAM: real-time single camera SLAM”. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 29, No. 6, pp. 1052–1067.
- Elinas, P., Sim, R. and Little, J., 2006. “ $\sigma$ -SLAM: Stereo vision SLAM using the rao-blackwellised particle filter and a novel mixture proposal distribution”. *International Conference on Robotics and Automation*. Orlando, FL, EUA.
- Hähnel, D., Fox, D., Burgard, W. and Thrun, S., 2003. “A highly efficient FastSLAM algorithm for generating cyclic maps of large-scale environments from raw laser range measurements”. In *Proceedings of the Conference on Intelligent Robots and Systems (IROS)*.
- Kim, J.H. and Sukkarieh, S., 2003. “Airborne simultaneous localisation and map building”. *IEEE International Conference on Robotics and Automation*, Vol. 1, pp. 406–411.
- Lemaire, T., Berger, C., Jung, I. and Lacroix, S., 2007b. “Vision-based SLAM: Stereo and monocular approaches”. *International Journal of Computer Vision*, Vol. 74, No. 3, pp. 343–364.
- Lemaire, T. and Lacroix, S., 2007a. “Monocular-vision based SLAM using line segments”. *International Conference on Robotics and Automation*. Roma, Itália.
- Montemerlo, M., Thrun, S., Koller, D. and Wegbreit, B., 2002. “FastSLAM: A factored solution to the simultaneous localization and mapping problem”. *National Conference on Artificial Intelligence*. Edmonton, Canadá.
- Mouragnon, E., Lhuillier, M., Dhome, M., Dekeyser, F. and Sayd, P., 2006. “Monocular vision based SLAM for mobile robots”. *International Conference on Pattern Recognition*. Hong Kong, China.
- Nieto, J., Guivant, J., Nebot, E. and Thrun, S., 2003. “Real time data association for FastSLAM”. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.
- Salvi, J., Petillo, Y., Thomas, S. and Aulinas, J., 2008. “Visual SLAM for underwater vehicles using video velocity log and natural landmarks”. *OCEANS*, pp. 1–6. 15 a 18 de Setembro.
- Sim, R., Elinas, P., Griffin, M. and Little, J.J., 2005. “Vision-based SLAM using the rao-blackwellised particle filter”. *IJCAI Workshop on Reasoning with Uncertainty in Robotics*, pp. 9–16.
- Sünderhalf, N., Lange, S. and Protzel, P., 2007. “Using the unscented kalman filter in Mono-SLAM with inverse depth parametrization for autonomous airship control”. *IEEE International Workshop on Safety, Security and Rescue Robotics*. Roma, Itália.
- Thrun, S., Burgard, W. and Fox, D., 2006. *Probabilistic robotics*. MIT Press.
- Zhang, Z., Huang, Y., Li, C. and Kang, Y., 2008. “Monocular vision simultaneous localization and mapping using surf”. *WCICA*, pp. 1651–1656. 25 a 27 de Junho.

## 6. DIREITOS AUTORAIS

Os autores são os únicos responsáveis pelo conteúdo do material impresso incluído no seu trabalho.



**VI CONGRESSO NACIONAL DE ENGENHARIA MECÂNICA**  
**VI NATIONAL CONGRESS OF MECHANICAL ENGINEERING**  
**18 a 21 de agosto de 2010 – Campina Grande – Paraíba - Brasil**  
*August 18 – 21, 2010 – Campina Grande – Paraíba – Brazil*

## **CONSTRUCTION OF 3-D LOCAL MAPS USING MONOCULAR VISION AND SENSOR FUSION FOR NAVIGATION PURPOSES OF AUTONOMOUS MOBILE ROBOTS**

**Guilherme Leal Santos, gsantos@dca.ufrn.br<sup>1</sup>**  
**Pablo Javier Alsina, pablo@dca.ufrn.br<sup>1</sup>**

<sup>1</sup>Departamento de Engenharia de Computação e Automação, Centro de Tecnologia, Universidade Federal do Rio Grande do Norte, CEP 59078-900 - Natal - RN - Brasil

**Abstract.** *The simultaneous localization and mapping (SLAM) is an approach where an autonomous mobile robot uses observations of landmarks in the environment to build a local map to find itself. The use of a single camera as the primary sensor has some particularities, because it is a challenge to determine the depth of a feature using monocular vision, existing only 2-D measures of 3-D structures. This paper presents a technique of monocular visual SLAM, which uses features found in a sequence of images using the Speeded Up Robust Features (SURF) descriptor to determine which features can be used as landmarks. To start the mapping and localization, it must be determined the position of the landmarks in the environment. For this, it's available only the coordinates (u, v) of the features found in the image and the intrinsic and extrinsic parameters of the camera. Thus, for the initialization of the landmarks, is used the technique of delayed initialization. With the estimated position of the landmarks, it's possible to estimate the updated position of the robot at each step; achieving important results that prove the efficiency of the monocular visual SLAM proposed in this paper.*

**Keywords:** *3-D Local Maps; Speed up Robust Features (SURF); Monocular Vision; Simultaneous Localisation and Mapping (SLAM)*