APLICAÇÃO DE MAPAS DE SALIÊNCIA COMO LIMITADORES DE REGIÃO PARA DETECTORES CLÁSSICOS NA TAREFA DE ODOMETRIA VISUAL

Alana de Santana Correia*, Eduardo Oliveira Freire†, Élyson Ádan Nunes Carvalho‡, Lucas Molina§

*Universidade Federal de Sergipe
Departamento de Engenharia Elétrica
Grupo de Pesquisa em Robótica da UFS (GPRUFS)
Av. Marechal Rondom, s/n, Jardim Rosa Elze, 49.100-000 / São Cristóvão-SE

Emails: alana.stc@gmail.com, efreire@ufs.br, ecarvalho@ufs.br, lmolina@ufs.br

Abstract— This paper presents an evaluation of the use of saliency maps as search space limiters for classic feature detectors in visual odometry. For this, three distinct maps are used with the SURF detector and descriptor, whose objective is to evaluate if the limitation of point detection in smaller and more informative regions of the image can compromise the robustness of the visual odometry system. The results using different scenes show that saliency maps can be a promising alternative for this task, contributing with positive results for odometry. However, this factor is dependent on the saliency map in eliminating the regions that do not have the features searched by SURF and keep regions stable during the movement of the robot by the environment.

Keywords— saliency maps, SURF, visual odometry.

Resumo— Este trabalho propõe a avaliação do uso de mapas de saliência como limitadores de região de busca para detectores clássicos na tarefa de odometria visual. Para isso, são utilizados três mapas distintos em conjunto com o detector e descritor de características SURF, cujo objetivo é avaliar se a limitação de detecção de pontos em regiões menores e mais informativas da imagem pode comprometer a robustez do sistema de odometria visual. Os resultados obtidos utilizando diferentes cenas mostram que mapas de saliência podem ser uma alternativa promissora para esta tarefa, contribuindo com resultados positivos para a odometria. No entanto, esse fator é dependente do mapa de saliência ser capaz de eliminar as regiões que não possuem as características procuradas pelo SURF e manter as regiões procuradas estáveis durante a movimentação do robô pelo ambiente.

Palavras-chave— mapas de saliência, SURF, odometria visual.

1 Introdução

Localizar um robô consiste em determinar sua posição e orientação no espaço em um determinado instante de tempo (Borenstein, Everett, Feng et al., 1996). Os métodos de localização de robôs móveis podem ser classificados em duas grandes categorias: métodos de localização relativa e métodos de localização absoluta (Borenstein et al., 1997). Os métodos de localização relativa utilizam as localizações obtidas em instantes anteriores para estimar a localização atual do robô, sendo a odometria e navegação inercial dois métodos baseados nesse princípio.

A odometria é um método de localização bastante utilizado na robótica. Quando utilizado em robôs móveis, esse método faz uso de sensores para medir a rotação das rodas do robô e a partir do modelo cinemático deste, calcular a sua posição e orientação. Desta forma, a odometria pode gerar erros que se propagam cumulativamente com a distância percorrida (Fraundorfer and Scaramuzza, 2011).

Na busca por técnicas mais robustas de odometria que possam melhorar a precisão de localização do robô, juntamente com o avanço da tecnologia das câmeras, foram desenvolvidas diversas soluções de localização relativa utilizando visão computacional, que são normalmente chamados de odometria visual.

A odometria visual é menos susceptível a ser afetada por derrapagens e outros problemas inerentes à odometria clássica, e por isso tem sido bastante estudada nos últimos anos (Fraundorfer and Scaramuzza, 2011). Grande parte dos sistemas de odometria visual utilizam o rastreamento de pontos de interesse para estimar a movimentação do robô entre quadros subsequentes em um ambiente. Para a extração de características de uma imagem as abordagens mais conhecidas se referem ao trabalho realizado pelos detectores e descritores de pontos-chave. Os detectores são responsáveis por selecionar os pontos e os descritores por atribuir uma identidade única aos pontos selecionados. Algoritmos que possuem detectores e descritores são conhecidos como extratores de características.

Nesta área já existem algoritmos robustos para realizar tal tarefa. Um dos líderes é o algoritmo SIFT (*Scale-Invariant Feature Transform*) proposto por Lowe (Lowe, 1999), e o algoritmo SURF (*Speed-Up Robust Features*) (Bay et al., 2008). A diferença básica entre eles é que o SURF traz uma versão mais acelerada do SIFT por reduzir algumas operações computacionalmente custosas. A redução do custo computacional dos extratores é essencial em aplicações práticas, na odometria visual por exemplo é preferível que o extrator tenha uma boa precisão de

detecção, selecione uma boa quantidade de pontos de modo a não prejudicar a estimativa de posição do robô e não seja lento para garantir tarefas em tempo real.

Neste sentido, ainda existe uma dificuldade na área de extração de características em estabelecer o equilíbrio entre tempo de processamento e robustez sem comprometer a qualidade da detecção.

Por outro lado, existe uma área da visão computacional, pouco explorada em aplicações práticas, voltada a desenvolver mecanismos que reduzam o custo computacional no processamento de cenas visuais em máquinas. Essa área é conhecida como atenção visual, justamente por buscar inspiração no processo de atenção humana (Itti and Koch, 2001)(Itti et al., 1998).

O sistema visual humano utiliza uma estratégia baseada na economia do esforço, dando prioridade à extração das informações necessárias à execução de uma determinada tarefa, de modo que somente essas informações são processadas em níveis cognitivos mais altos e as informações irrelevantes são descartadas (Itti and Koch, 2001)(Itti et al., 1998).

No contexto da visão computacional, uma representação das informações mais importantes de uma cena é realizada através da composição de um mapa topográfico, amplamente conhecido como mapa de saliência (Itti et al., 1998). Esse mapa atribui valores altos para regiões consideradas mais importantes da cena, e valores baixos para regiões consideradas menos informativas.

Os mapas de saliência foram criados a *priori* com o objetivo de reproduzir o processo de atenção humana, sendo que alguns modelos são apenas teóricos. Por isso, eles não são tão explorados em aplicações práticas como outras ferramentas da visão computacional. Alguns trabalhos de maior destaque envolvendo o mapa de saliência clássico em uma aplicação de localização de um robô móvel foram propostos por (Siagian and Itti, 2009), (Newman and Ho, 2005) e (Frintrop et al., 2006).

No trabalho proposto por Siagian o mapa é utilizado como limitador de região de busca para o algoritmo SIFT. Segundo Siagian o principal objetivo é utilizar o mapa de saliência para selecionar apenas as regiões mais importantes, aliviando o custo de processamento do SIFT. Segundo Siagian, para viabilizar o uso do mapa em conjunto com o SIFT foram feitas algumas restrições de acordo com o tamanho e a quantidade das regiões que deve ser detectada em uma cena. Nesse sentido, o uso dos mapas de saliência pode afetar diretamente o resultado do sistema de localização do robô. Além disso, foi notada a ausência de uma avaliação prévia do impacto dos modelos de atenção visual quando utilizados como limitadores de espaço de busca para extratores clássicos. Assim, este artigo tem como objetivo avaliar a aplicabilidade dos mapas de saliência em um sistema de odometria visual de um robô móvel. O objetivo é avaliar se a redução da área de detecção utilizada pelo SURF afeta significativamente o resultado da odometria visual.

Este artigo está organizado da seguinte forma: Na seção 2 são apresentados os mapas de saliência utilizados neste trabalho. Na seção 3 é apresentado o extrator clássico SURF. Na seção 4 são descritos detalhes de segmentação e limiarização dos mapas de saliência. Ne seção 5 são descritos os detalhes do sistema de odometria e como foi realizada a estimativa de posição do robô. Na seção 6, são apresentados os resultados experimentais. Por fim, na seção 7 são apresentadas as conclusões e trabalhos futuros.

2 Mapas de Saliência

Segundo (Itti et al., 1998), mapas de saliência geralmente são robustos na presença de alguns tipos de ruído e os alvos indicados por eles normalmente são bons candidatos a marcos naturais do ambiente. Com o crescimento da área, surgiram diversos tipos de mapas de saliência. Cada abordagem propõe a construção do mapa utilizando diferentes técnicas para detectar as áreas mais relevantes de uma cena.

Este trabalho utiliza três modelos distintos de saliência. O primeiro modelo utilizado é o modelo clássico, proposto por Laurent Itti (Itti et al., 1998). Além do modelo clássico de saliência, este trabalho também utiliza um modelo que utiliza a teoria de grafos, conhecido como GBVS (Harel et al., 2007) e um modelo mais atual que utiliza aprendizado de máquina, conhecido como LDS (Fang et al., 2017).

As seções 2.1, 2.2 e 2.3 apresentam maiores detalhes a respeito desses mapas.

2.1 Mapa de Saliência Clássico

O modelo de atenção visual desenvolvido por Laurent Itti e Christof Koch é um método clássico de atenção visual, classificado como um modelo cognitivo e amplamente estudado por muitos pesquisadores. Uma visão geral do assunto pode ser vista em (Itti and Koch, 2001) (Itti et al., 1998).

Este modelo consiste na decomposição da imagem em três características primitivas: cor, intensidade e orientação. Para cada uma delas são criados diferentes mapas. As diferentes regiões desses mapas são qualificadas segundo a resposta gerada pelas características e as regiões que mais se destacam são chamadas de áreas de atenção. Os mapas são então combinados para a criação do mapa de saliência final, definindo os focos da atenção. Para isso, é necessário executar uma sequência de etapas, como será descrito a seguir.

A primeira etapa consiste na formação de mapas individuais que quantificam a presença das características extraídas da cena. São criados 4 mapas de cor, o R (vermelho), G (verde), B (azul) e Y (amarelo), levando em consideração a teoria de oposição de cores. É gerado também um mapa de intensidade I, que consiste na média dos canais R, G e B da imagem. E por fim, são criados quatro mapas de orientação gerados a partir do filtro de Gabor com valores distintos de ângulos, que neste modelo são 0°, 45°, 90° e 135°.

A partir de cada mapa formado são criados nove mapas em escalas distintas para cada característica, formando assim pirâmides gaussianas. Com as pirâmides definidas, os mapas de características são obtidos efetuando operações de diferença entre os mapas com escalas distintas. Este processo, segundo Itti (Itti et al., 1998), tenta representar a ação dos neurônios dos campos visuais, que são estimulados em uma pequena região do espaço visual central e são inibidos em áreas vizinhas.

No modelo, estas operações são realizadas como diferenças entre escalas finas (níveis de maior resolução na pirâmide) e as escalas grossas (níveis de menor resolução na pirâmide). No total, 42 mapas são computados: 6 de intensidade, 12 de cor e 24 de orientação, pois são construídos 6 mapas para cada uma das quatro orientações utilizadas.

Para combinar os mapas de características em somente um é necessário utilizar um operador de normalização N(.). Neste modelo de atenção existem quatro métodos para realizar a normalização, quais sejam (i) Somatório Simples, (ii) Normalização não linear, (iii) Competição Iterativa, (iv) Normalização com pesos aprendidos. Em (Itti and Koch, 1999) é feita uma comparação entre os resultados de mapas de saliência obtidos utilizando cada operador de normalização e, por conta da simplicidade e dos resultados satisfatórios encontrados, é mais comum utilizar nas aplicações a competição iterativa, que consiste na aplicação de filtros 2D de diferenças gaussianas (DoG).

Uma vez escolhida a estratégia de normalização que melhor se adapte aos objetivos de utilização do mapa de saliência, os mapas de características são então normalizados e combinados em três mapas de conspicuidade (I, C e O), gerados através de uma adição entre escalas, que consiste em expandir cada mapa para a escala 4 e efetuar uma soma *pixel* a *pixel* (Itti et al., 1998). Para compor o mapa de saliência os três mapas (I, C e O) são normalizados e uma combinação linear é realizada entre eles, o que produz a saída final S.

2.2 GBVS - Graph-Based Visual Saliency

Neste modelo os mapas são extraídos de forma semelhante à apresentada pelo modelo clássico na seção 2.1. No entanto, para cada mapa de característica é gerado um grafo totalmente conectado, de modo que nodos são responsáveis pela representação de todas as localizações. Pesos entre os nodos são associados proporcionalmente à similaridade em relação às suas características e ponderadas por suas distâncias espaciais. Em um processo de normalização, nodos que representem altos valores de dissimilaridade em relação à vizinhança são definidos com altos valores de saliência. Os mapas são finalmente normalizados para enfatizar os detalhes importantes, e por fim combinados em um único mapa de saliência. Mais detalhes desse modelo podem ser encontrados em (Harel et al., 2007).

2.3 LDS - Learning Discriminative Subspaces

O modelo de atenção LDS é um modelo de atenção visual mais elaborado, pois tem como objetivo distinguir os alvos e distrações que compartilham certos atributos visuais (Fang et al., 2017).

Para isso, este modelo utiliza um grande conjunto de imagens de treinamento. Junto com as imagens de treinamento também são utilizados os mapas de fixação humana disponibilizados publicamente por diversos pesquisadores em http://saliency.mit.edu/. Esses mapas definem os alvos corretos para a cena.

Inicialmente são gerados diferentes mapas de características, e com o auxílio da técnica PCA os elementos salientes de cada mapa são ressaltados. Esses mapas são gerados em diferentes escalas. Dispondo do mapa de fixação da imagem de entrada e dos mapas de características para a imagem, os mapas são separados em cinco grupos distintos: (i) mapas que ressaltam apenas os elementos salientes, (ii) mapas que ressaltam apenas os distratores, (iii) mapas que ressaltam tanto distratores como elementos salientes, (iv) mapas que apresentam respostas baixas e elevadas para partes diferentes dos elementos salientes, (v) mapas que não conseguem distinguir quem são os alvos e os distratores. Esse último grupo é eliminado da fase de treinamento (Fang et al., 2017). Posteriormente, é aplicado um fator de normalização nos mapas de características escolhidos para elevar o contraste entre alvos e distratores.

Em seguida, os mapas de características e a imagem são quebrados em *patches* e com a informação do mapa de fixação humano é possível separar os *patches* da imagem que indicam os alvos e os *patches* que indicam os distratores. Cada *patch* da imagem possui n vetores de características, que são resultantes dos n mapas criados. A partir de então, esses vetores são utilizados em um algoritmo de treinamento que gera como saída uma escolha otimizada de características e pesos para cada mapa escolhido.

3 SURF - *Speed-Up Robust Features*

O algoritmo SURF é, na verdade, uma versão mais acelerada do SIFT (Lowe, 1999). Para gerar o espaço de escalas, o SIFT faz uma aproximação do Laplaciano da Gaussiana com a Diferença de Gaussianas. O SURF (Bay et al., 2008) vai um pouco mais além, e aproxima o LoG com um filtro retangular conhecido como filtro de caixa (*box filter*). Uma vantagem dessa aproximação é que a operação de convolução com o filtro retangular pode ser facilmente calculada com a ajuda da técnica de Imagem Integral. Através dessa técnica, operações como médias e somas são realizadas de forma mais rápida, otimizando o custo computacional. Essa operação de convolução pode ser feita em paralelo para diferentes escalas (Bay et al., 2008).

Para assinalar a orientação de um ponto de interesse, o SURF usa respostas da aplicação da Transformada de Haar (*Haar-wavelet*) nas direções x e y, dentro de uma região de vizinhança especificada. Uma vez que as respostas à *wavelet* são calculadas e ponderadas por uma gaussiana centrada no ponto de interesse, as respostas são plotadas em um dado espaço. A orientação dominante é estimada calculando-se a soma de todas as respostas que estão dentro de uma área do espaço com ângulo de 60 graus. Para a composição do descritor, o SURF define uma vizinhança em torno do ponto em questão. Essa janela é subdividida em regiões de 4x4 *pixels*. Para cada elemento da sub-região, as respostas da *Haar-wavelet* nas direções de x e y são tomadas, e um vetor é formado com esses elementos. No final, o descritor será formado pela concatenação dos vetores de todas as sub-regiões, formando um vetor descritor de 64 elementos.

4 Pré-Processamento das Áreas de Saliência

A saída dos dois modelos de atenção visual utilizados neste trabalho é um conjunto de áreas de interesse. No entanto, as áreas de saliência dentro do mapa são pontuadas com diferentes valores, que variam de acordo com a relevância de cada elemento da cena. Como o objetivo deste trabalho é selecionar apenas as regiões mais interessantes, foi utilizando um método de limiarização automático para definir qual o limiar de corte entre as regiões mais e menos salientes dos mapas.

O algoritmo de limiarização escolhido foi o de Otsu (Gonzalez and Woods, 2000). Essa limiarização contribuiu para selecionar apenas as regiões com maiores focos de atenção.

Após a limiarização foi utilizado um método de segmentação por crescimento de regiões (Gonzalez and Woods, 2000). Esse método permite separar as regiões mais importantes acima de um determinado limiar e eliminar as regiões de saliência muito pequenas. Após a segmentação, as regiões mais salientes são identificadas na imagem original. Realizado esse processo, apenas as regiões consideradas interessantes são processadas pelos extratores de características.

5 Odometria Visual

A solução para o problema de localização com imagens tem princípio no início dos anos 80, sendo um dos primeiros trabalhos o proposto por (Moravec, 1980). Com o passar dos anos, o trabalho de (Nistér et al., 2004) foi um marco divisório nas pesquisas de odometria visual. Nesse trabalho foi utilizado um algoritmo para eliminação de valores extremos (*outliers*) e estimativa de pose 3D-2D para calcular a nova posição. Mais recentemente, com a popularidade adquirida pelos métodos de SLAM e outras técnicas de navegação autônoma, surgiram trabalhos que integram técnicas de odometria visual em um sistema mais completo, como o proposto por (Engel et al., 2014). Neste trabalho, os autores solucionaram os problemas de estimativa de postura em ambientes desconhecidos, sem a utilização de sensores possivelmente ruidosos e sem outros métodos de localização, por meio apenas da odometria visual monocular em um *drone*.

Para realizar tais tarefas, grande parte dos algoritmos de odometria visual se utilizam de pontos de interesse que podem ser facilmente localizados na cena para estimar a movimentação do robô entre quadros subsequentes. Esses sistemas necessitam de três componentes básicos: um sensor; métodos para detecção e correspondência dos pontos de interesse; e um método para calcular a posição do robô (Borenstein, Everett and Feng, 1996).

Os sensores normalmente utilizados são câmeras e o método de detecção de pontos é responsável por detectar pontos de interesse que possuam características distintas e que sejam fáceis de localizar na imagem. Após a detecção dos pontos de interesse em cada imagem, é criado um conjunto de pontos para cada imagem processada. Tal conjunto possui um vetor identidade para cada ponto de interesse. A próxima etapa é estabelecer as correspondências corretas entre os pontos de interesse da primeira imagem com os da segunda imagem.

Realizada a correspondência dos pontos, utiliza-se os pontos correspondidos entre dois quadros obtidos pela câmera para se estimar a transformação de corpo rígido, ou seja, rotação e translação do robô. Existem diversas técnicas para a estimativa da relação de transformação entre dois conjuntos de pontos baseados na minimização por mínimos quadrados. Neste trabalho, foi utilizada a abordagem de decomposição em valor singular (SVD) (Arun et al., 1987).

As transformações de rotação e translação ob-

tidas através da decomposição em valor singular podem ser representadas na forma matricial, e a relação de transformação do robô nos instantes *k-1* e *k* é dada por

$$T_{k-1}^{k} = \begin{pmatrix} R_{k-1}^{k} & t_{k-1}^{k} \\ 0 & 1 \end{pmatrix} \quad (1)$$

em que R_{k-1}^{k} representa a rotação e t_{k-1}^{k} representa a translação entre os instantes *k-1* e *k*.

Após obter a transformação T_{k-1}^{k}, é necessário atualizar a posição do robô com relação às coordenadas de referência, que é normalmente a posição do robô no instante $k = 0$. A conversão é feita da seguinte forma

$$M_{k}^{0} = M_{k-1}^{0} T_{k}^{k-1} \quad (2)$$

Considerando que possa haver correspondências incorretas e ruído na detecção dos pontos ou na captura da imagem, é importante a implementação de uma técnica de remoção de pontos considerados *outliers*, ou seja, pontos que não correspondem com a tendência do conjunto e que podem afetar o resultado final ocasionando erros grosseiros na estimativa da posição. Tais erros podem comprometer todo sistema de localização, considerando que a odometria é um sistema que possui acumulação de erros inerente à técnica.

Uma abordagem muito utilizada em diferentes áreas de pesquisa e que é utilizada neste trabalho para eliminar os *outliers* é o RANSAC (*RANdom SAmple Consensus*)(Nistér et al., 2006). O RANSAC é responsável pela busca aleatória do modelo que melhor se encaixa entre os pontos testados. Como esta é uma abordagem aleatória, o resultado será diferente para cada execução, porém, bastante próximo do valor real.

Definido como foi realizada a estimativa de posição do robô, a seção 6 apresenta os resultados obtidos.

6 Resultados Experimentais

Para a realização dos experimentos foram utilizados três ambientes distintos, cujo objetivo é verificar se a mudança do conteúdo das cenas pode impedir o rastreamento do robô por conta das áreas selecionadas pelos mapas de saliência. Algumas cenas dos três ambientes escolhidos podem ser visualizadas na Figura 1. O robô utilizado nos experimentos foi o Pioneer-3DX disponibilizado pelo grupo de pesquisa em robótica da UFS (GPRUFS), e para a aquisição das imagens foi utilizado o Kinect XBOX 360. A utilização do Kinect tem como objetivo facilitar a aquisição dos dados de profundidade da cena evitando a utilização de visão estéreo ou de cenas apenas planares.

O objetivo dos experimentos é verificar a aplicabilidade e as fragilidades dos mapas de saliência na tarefa de odometria visual. Por isso, todos os

(a)

(b)

(c)

Figura 1: Algumas cenas dos ambientes utilizados nos experimentos. (a) são cenas do ambiente 1. (b) são cenas do ambiente 2. (c) são cenas do ambiente 3.

experimentos foram simplificados fazendo o robô seguir em linha reta, paralelo e há uma distância fixa de 1,5 m dos objetos de interesse da cena.

Para verificar se os resultados da odometria visual são condizentes com a correta movimentação do robô no ambiente, foram utilizados os dados de odometria do robô como padrão para comparação, já que em baixa velocidade e pequenos percursos a odometria do Pioneer é bastante confiável e pode ser utilizada como referência.

Além de analisar os dados da odometria também foram realizados testes para verificar alterações no tempo de processamento do experimento utilizando apenas o SURF e o SURF em composição com as técnicas de saliência. Para isso, todos os experimentos foram executados utilizando um mesmo computador com processador Intel Core i5-4200U, CPU de 1,60 GHz, 8 GB de memória RAM, sistema operacional Windows 10 de 64 bits.

A Figura 2 ilustra os resultados de odometria obtidos no primeiro experimento utilizando o ambiente 1. Os resultados da odometria visual utilizando somente o SURF e o SURF em conjunto com o mapa GBVS foram bastante semelhantes

com o resultado da odometria do robô, apresentando um erro no eixo Y de aproximadamente 15 cm, já no eixo X o erro foi bem mais reduzido, de apenas 1,5 cm. Esses resultados mostram que mesmo com a redução da área de busca do SURF pelo mapa de saliência, foi possível manter bons resultados. Por outro lado, a utilização dos mapas LDS e do modelo clássico proposto por Itti prejudicaram bastante o desempenho do sistema de odometria, sendo que com o mapa de saliência clássico de Itti não foi possível construir a trajetória do robô pela ausência de pontos na etapa de correspondência, impossibilitando assim a estimativa de posição. Com o mapa LDS os resultados também não foram favoráveis, pois os resultados obtidos foram bem diferentes da odometria real do robô.

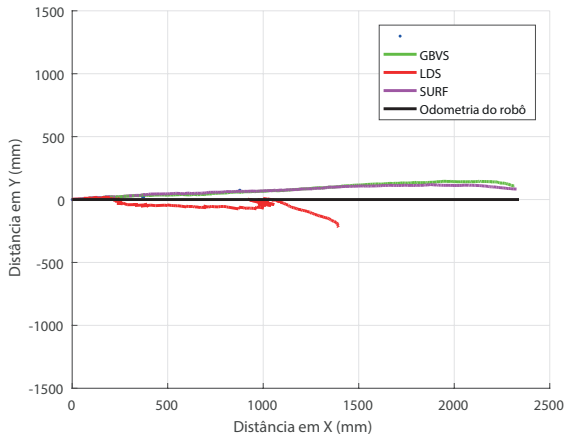

Figura 2: Trajetória em linha reta efetuada pelo robô Pioneer-3DX no ambiente 1.

Tabela 1: Tempo de processamento para todas as imagens do ambiente 1.

Técnica	Tempo (s)
SURF	1312,5
ITTI+SURF	535,5
LDS+SURF	762,5
GBVS+SURF	748,5

De acordo com os dados da Tabela 1 o SURF apresenta um tempo de processamento maior que quando utilizado em conjunto com as técnicas de saliência. O modelo clássico proposto por Itti foi o que obteve maior redução do tempo de execução, com 535,5 s, mas por selecionar áreas muito pequenas, o que explica a maior redução do tempo de processamento, prejudicou os resultados de localização. O modelo LDS, apesar de reduzir significativamente o tempo de execução, não obteve um resultado satisfatório na localização do robô. O GBVS foi o modelo que apresentou tanto resultados satisfatórios na localização como com relação à redução do tempo de execução.

No experimento realizado no ambiente 2 também ocorreram resultados positivos utilizando o mapa de saliência GBVS em conjunto com o SURF, que como ilustrado na Figura 3, foi a abordagem que obteve os resultados mais próximos da odometria do robô. Isto ilustra a robustez desse mapa com relação à mudança de conteúdo da cena sem interferir no desempenho do algoritmo de detecção de pontos. Já os resultados obtidos utilizando o modelo LDS e o modelo proposto por Itti não foram satisfatórios.

Utilizando esses últimos dois métodos não foram encontrados muitos pontos de correspondência, pelo fato deles selecionarem regiões de saliência muito pequenas no ambiente. Desse modo, foi observado que em regiões muito pequenas o desempenho do SURF é prejudicado, pois os detectores clássicos trabalham com imagens completas em busca de bordas abruptas. Se uma região selecionada pelo mapa de saliência for muito pequena de modo a apresentar somente superfícies sem bordas, a taxa de correspondências do SURF cai significativamente, prejudicando a odometria. A Figura 6 ilustra o problema da ausência de pontos na etapa de correspondência utilizando o modelo clássico proposto por Itti.

A Tabela 2 apresenta os resultados obtidos com relação ao tempo de execução. Com o uso dos mapas de saliência o tempo de execução foi reduzido em até 50%. No entanto, apenas o modelo GBVS apresentou bons resultados na localização, o que demonstra a eficiência da técnica reduzindo o tempo de execução sem prejudicar os resultados da odometria.

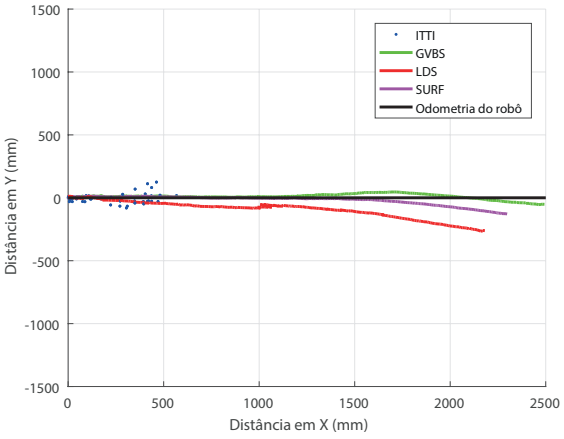

Figura 3: Trajetória em linha reta efetuada pelo robô Pioneer-3DX no ambiente 2.

Com relação ao ambiente 3 foi observada uma queda de desempenho de todos os métodos, ocasionada pelo acúmulo do erro no eixo Y, pois até mesmo nos testes utilizando somente o SURF o erro de odometria no eixo Y foi de aproximadamente 40 cm. Isto demonstra que a perda de robustez utilizando o mapa GBVS não foi ocasio-

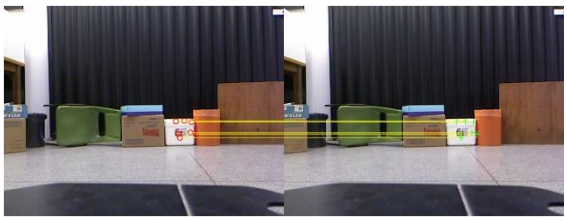

Figura 4: Correspondência de pontos entre dois *frames* do ambiente 2 utilizando o mapa de saliência clássico proposto por Itti.

Tabela 2: Tempo de processamento para todas as imagens do ambiente 2.

Técnica	Tempo (s)
SURF	1612,9
ITTI+SURF	738,6
LDS+SURF	989,2
GBVS+SURF	893,7

nada pela utilização da técnica de limitação de região, mas sim pelo acúmulo do erro de orientação, possivelmente ocasionado por um mal posicionamento do Kinect sobre o robô. No entanto, do mesmo modo como observado nos experimentos anteriores, não foi possível realizar o rastreamento utilizando o modelo clássico proposto por Itti por conta da sensibilidade do método em variar constantemente as regiões de saliência durante o percurso do robô nesse ambiente, e quando selecionadas regiões consistentes em um par de imagens, elas são muito pequenas, dificultando a busca de pontos realizada pelo SURF.

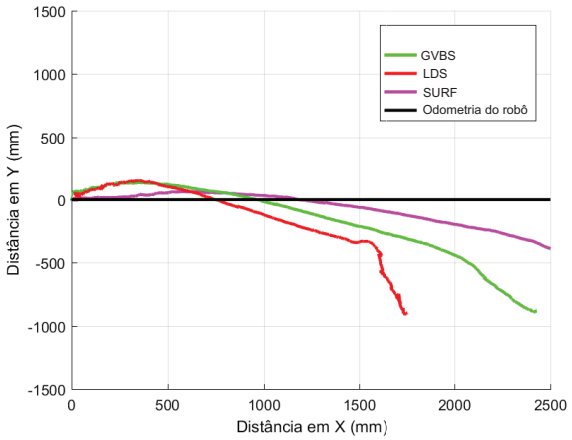

Figura 5: Trajetória em linha reta efetuada pelo robô Pioneer-3DX no ambiente 3.

De acordo com a Tabela 3, o uso da técnica GBVS apresentou uma redução no tempo de execução de 1222,8 segundos quando comparado ao tempo de execução do SURF. Essa é a maior redução de tempo entre todas as técnicas utilizadas no experimento 3. Nesse experimento, apesar da técnica GBVS não ter obtido um resultado de odometria tão próximo da odometria do robô, seus resultados foram melhores que o SURF com um tempo de execução bem menor, demonstrando a aplicabilidade dos mapas de saliência como limitadores de região de busca.

Tabela 3: Tempo de processamento para todas as imagens do ambiente 3.

Técnica	Tempo (s)
SURF	1977,1
ITTI+SURF	952,1
LDS+SURF	998,2
GBVS+SURF	754,3

7 Conclusões

Neste artigo foi apresentada uma avaliação do uso de mapas de saliência para limitar regiões de busca do extrator clássico SURF. Para isso, foram utilizados três modelos de atenção visual distintos para gerar as áreas de saliência: o modelo clássico proposto por Itti, um modelo que utiliza a teoria de grafos chamado GBVS e um modelo que utiliza aprendizado de máquina chamado LDS.

De acordo com os resultados obtidos é possível notar que o uso de mapas de saliência pode alterar bastante a resposta do sistema de odometria visual. A reposta do sistema de odometria pode ser melhorada a partir do mapa ou ela também pode ser prejudicada por completo. Isso ocorre devido ao tamanho das áreas selecionadas e às mudanças de focos de atenção em uma cena dinâmica.

De acordo com os resultados obtidos, os mapas que selecionam uma quantidade muito restrita de focos na cena, ou seja, selecionam regiões muito pequenas, prejudicam o processo de correspondência de pontos. Como são encontrados poucos pontos, o sistema de odometria se torna menos robusto e mais suscetível a erros. Esse problema foi notado utilizando o modelo clássico proposto por Itti e o modelo LDS. Neste sentido, propõe-se como trabalhos futuros investigar os fatores que podem contribuir para uma melhor seleção de regiões de saliência para mapas que apresentam esse problema em aplicações práticas.

Por outro lado, sistemas de atenção que apresentam uma transição suave dos focos durante a movimentação do robô, e tendem a selecionar as áreas de atenção visual eliminando regiões que não tem tantas características informativas para o SURF contribuem para o processo de filtragem de pontos para o sistema de odometria. Esse resultado foi observado utilizando o modelo GBVS, que apresentou resultados satisfatórios nos experimentos 1 e 2.

Com relação ao tempo de execução, foi observada uma redução significativa de cerca de até

50% ao utilizar alguma técnica de atenção de visual juntamente com o SURF. No entanto, apenas a técnica GBVS apresentou resultados satisfatórios na localização do robô e na redução do tempo de execução, o que demonstra que esses dois critérios precisam ser corretamente analisados antes de adotar uma técnica de atenção visual como limitador de espaço de busca.

Com a validação do uso da técnica em pequenos percursos, propõe-se como trabalho futuro, analisar o desempenho da abordagem proposta em percursos mais longos. Para tanto podem ser utilizados bases de dados públicas para a realização dos testes e comparação de resultados.

Agradecimentos

O presente trabalho foi realizado com o apoio da CAPES, entidade do Governo Brasileiro voltada para a formação de recursos humanos.

Referências

Arun, K. S., Huang, T. S. and Blostein, S. D. (1987). Least-squares fitting of two 3-d point sets, *IEEE Transactions on pattern analysis and machine intelligence* (5): 698–700.

Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L. (2008). Speeded-up robust features (surf), *Computer vision and image understanding* **110**(3): 346–359.

Borenstein, J., Everett, H. and Feng, L. (1996). *Navigating mobile robots: Systems and techniques*, AK Peters, Ltd.

Borenstein, J., Everett, H., Feng, L. et al. (1996). Where am i? sensors and methods for mobile robot positioning, *University of Michigan* **119**(120): 15.

Borenstein, J., Everett, H. R., Feng, L. and Wehe, D. (1997). Mobile robot positioning-sensors and techniques, *Technical report*, NAVAL COMMAND CONTROL AND OCEAN SURVEILLANCE CENTER RDT AND E DIV SAN DIEGO CA.

Engel, J., Sturm, J. and Cremers, D. (2014). Scale-aware navigation of a low-cost quadrocopter with a monocular camera, *Robotics and Autonomous Systems* **62**(11): 1646–1656.

Fang, S., Li, J., Tian, Y., Huang, T. and Chen, X. (2017). Learning discriminative subspaces on random contrasts for image saliency analysis, *IEEE transactions on neural networks and learning systems* **28**(5): 1095–1108.

Fraundorfer, F. and Scaramuzza, D. (2011). Visual odometry: Part i: The first 30 years and fundamentals, *IEEE Robotics and Automation Magazine* **18**(4): 80–92.

Frintrop, S., Jensfelt, P. and Christensen, H. I. (2006). Attentional landmark selection for visual slam, *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, IEEE, pp. 2582–2587.

Gonzalez, R. C. and Woods, R. E. (2000). *Processamento de imagens digitais*, Edgard Blucher.

Harel, J., Koch, C. and Perona, P. (2007). Graph-based visual saliency, *Advances in neural information processing systems*, pp. 545–552.

Itti, L. and Koch, C. (1999). Comparison of feature combination strategies for saliency-based visual attention systems, *Human vision and electronic imaging IV*, Vol. 3644, International Society for Optics and Photonics, pp. 473–483.

Itti, L. and Koch, C. (2001). Computational modelling of visual attention, *Nature reviews neuroscience* **2**(3): 194.

Itti, L., Koch, C. and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis, *IEEE Transactions on pattern analysis and machine intelligence* **20**(11): 1254–1259.

Lowe, D. G. (1999). Object recognition from local scale-invariant features, *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, Vol. 2, Ieee, pp. 1150–1157.

Moravec, H. P. (1980). Obstacle avoidance and navigation in the real world by a seeing robot rover., *Technical report*, STANFORD UNIV CA DEPT OF COMPUTER SCIENCE.

Newman, P. and Ho, K. (2005). Slam-loop closing with visually salient features, *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, IEEE, pp. 635–642.

Nistér, D., Naroditsky, O. and Bergen, J. (2004). Visual odometry, *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, Vol. 1, Ieee, pp. I–I.

Nistér, D., Naroditsky, O. and Bergen, J. (2006). Visual odometry for ground vehicle applications, *Journal of Field Robotics* **23**(1): 3–20.

Siagian, C. and Itti, L. (2009). Biologically inspired mobile robot vision localization, *IEEE Transactions on Robotics* **25**(4): 861–873.