

Mapeamento e Reconstrução 3D com a utilização da Estereoscopia e SURF

Cordeiro, J. N. N.

Escola Politécnica de Pernambuco
Universidade de Pernambuco
50.720-001 - Recife, Brasil
jcnn@ecomp.poli.br

Fernandes, B. J. T.

Escola Politécnica de Pernambuco
Universidade de Pernambuco
50.720-001 - Recife, Brasil

Resumo *A visão estereoscópica simula a visão humana computacionalmente utilizando duas imagens distintas do mesmo cenário, neste trabalho é utilizada essa técnica para realizar a reconstrução 3D de um cenário. É feito o uso de algoritmos de visão computacional, tais como o detector de características, SURF, utilizado para marcar pontos nas duas imagens e para refinar os resultados e garantir uma reconstrução mais próxima do real é utilizado o algoritmo RANSAC.*

Abstract *The stereoscopic vision simulates human vision computationally using two different images of the same scene, in this study this technique is used, to perform the 3D reconstruction of a scene. The use of computer vision algorithms, such as how the detector characteristics, SURF, is used to score points made in the two images and to refine the results and ensure reconstruction of nearest real the RANSAC algorithm is used.*

1 Introdução

A construção de mapas 3D é importante em várias linhas de pesquisa e áreas do conhecimento, são algumas delas: topografia, na exploração de terrenos desconhecidos[1], exploração e reconhecimento do ambiente de outros planetas[2] ambiente subaquáticos[3] em geral, onde a presença humana se torna inviável.

As últimas décadas foram marcadas pelo rápido crescimento da tecnologia do escaneamento 3D, entre as quais destacam-se: câmeras estereoscópicas[4] câmeras LIDAR (Câmeras de Light Detection and Raging[5] e o Microsoft Kinect[6].

Esse trabalho visa à investigação do modelo de Visão estereoscópica com o auxílio do algoritmo SURF, para ser aplicado na área de mapeamento e reconstrução 3D. Utilizaremos ainda o algoritmo RANSAC para refinar os resultados.

A estereoscopia é uma técnica que possui motivação biológica, com sua atuação seguindo os mesmos princípios que regem a formação da imagem 3D pelo olho humano, de forma computacional, é feita utilizando um par de câmeras digitais posicionadas do a lado. As imagens captadas deverão estar direcionadas para um mesmo local, com isso terão pequenas diferenças entre si, essas diferenças são conhecidas como disparidades. Através dessas disparidades pode-se calcular a profundidade de um determinado ponto encontrado no cenário e montar um mapa 3D a partir disso.

Um problema enfrentado para o uso da estereoscopia é encontrar o mesmo ponto em ambas as imagens[7], uma das soluções encontradas tem como base achar pontos de interesse na primeira imagem e na segunda imagem, em seguida buscar tentar combiná-los, porém mudanças de iluminação, escala e rotação acabam dificultando a marcação desse mesmo ponto, ou aumentando o custo computacional[8]. Para resolver esse problema a ideia é utilizar o SURF para marcar o mesmo ponto de interesse nas duas imagens, independente da mudança de escala, rotação ou iluminação.

2 Técnica Proposta

Neste artigo é proposta uma abordagem específica para realização de uma reconstrução 3D, baseando-se, primeiramente na técnica da estereoscopia[9], onde simula-se a visão humana com duas câmeras alinhadas, mas com angulações diferentes, a partir dessas câmeras obtêm-se duas imagens, uma esquerda e outra direita. Após a aquisição das imagens o algoritmo do SURF é aplicado,

definido pontos de interesse nas duas imagens, feita a identificação desses pontos de interesse a correlação dos pontos e os cálculos de disparidade são feitos. Para correção e refinamento do mapa de pontos é aplicado o algoritmo do RANSAC que elimina outliers dos pontos correlacionados. O fluxograma abaixo resume os passos utilizados neste trabalho para a reconstrução 3D:



Fig 1. Fluxograma de representação da técnica proposta

2.1 Visão Estereoscópica

A visão computacional é formada por algoritmos que retiram informações a partir de imagens para aplicações específicas como o reconhecimento de cores ou de objetos. A visão estereoscópica é uma subárea da visão computacional que possui motivação biológica, baseando-se na visão humana, em que para enxergar um ponto, os dois olhos devem estar fixados neste, de maneira que eles apontem diretamente para o ponto em questão, para essa simulação, um hardware deve ser dotado de duas câmeras capturando imagens de pontos distintos, simulando a noção geométrica da visão.

Para a sensação de profundidade de um objeto[10], é feita uma análise matemática baseada no modelo representado na figura 2, utilizando-se artifícios de triangulação e trigonometria, com seus valores descritos de acordo com a tabela 1.

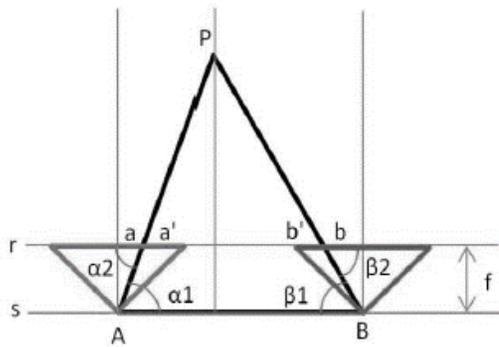


Fig 2. Representação do posicionamento das câmeras e do ponto de interesse

Tabela 1. Símbolos e Notações.

Símbolo	Legenda
<i>A</i>	Posição da Câmera Esquerda
<i>B</i>	Posição da Câmera Direita
<i>f</i>	Distância Focal das Câmeras
<i>X</i>	Distância entre as Câmeras
<i>P</i>	Ponto de Interesse
<i>Xe</i>	Distância do ponto <i>P</i> projetado no plano da imagem esquerda, até a origem
<i>Xd</i>	Distância do ponto <i>P</i> projetado no plano da imagem direita, até a origem
<i>D</i>	Distância do ponto <i>P</i> ao plano formado pelas duas câmeras

Após algumas análises, e desenvolvimento de equações, Ribeiro et. Al[7] chegou a equação 1 que descreve a distância de um ponto de interesse ao plano que se encontram as duas câmeras, sendo possível então utilizarmos essa informação para sabermos a profundidade desse ponto até onde a imagem foi tirada.

$$D = \frac{fX}{|Xe - Xd|}$$

Reconstrução 3D[11] é um conjunto de técnicas e métodos para transformar dados em imagem através de um dispositivo gráfico. Neste artigo iremos tratar da reconstrução 3D como sendo a utilização da técnica de estereoscopia para transformar uma imagem em um mapa de pontos 3D.

Dessa forma, utilizando a equação 1 e as imagens da figura 3, fazendo uso da técnica da estereoscopia, é possível construir um mapa de pontos 3D do cenário.

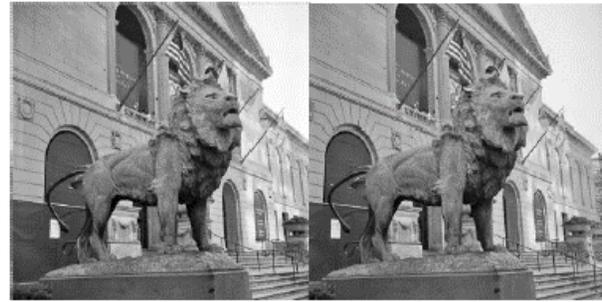


Fig 3. Imagens utilizadas para a Reconstrução 3D

2.2 Identificação de Pontos de Interesse e Correlação

Para utilizar a técnica de estereoscopia faz-se necessário a identificação de um mesmo ponto nas duas imagens e correlacioná-los[12]. Os pontos de interesse são portanto pontos candidatos a correlação, de modo que possam ser localizados nas duas imagens. É necessário a marcação de vários pontos de interesse para tentar correlacioná-los nas duas imagens.

A estereoscopia enfrenta algumas adversidades, entre elas estão a oclusão de pontos e a marcação e correlação dos mesmos pontos nas duas imagens. A oclusão de pontos ocorre quando na imagem obtida por uma das câmeras é possível a identificação de um certo ponto, mas na imagem obtida pela segunda câmera não é possível encontrar o mesmo ponto, tornando-o ocluído, problema até então, não solucionado.

O problema da correlação dos mesmos pontos em ambas as imagens pode se dar de duas formas, a partir da não identificação do mesmo ponto nas duas imagens e da afirmação que um ponto de uma imagem é o mesmo ponto na segunda imagem.

Para solucionar o problema da marcação do mesmo ponto nas duas imagens são utilizados vários pontos de interesse para aumentar as chances de sucesso na correlação dos pontos.

Para se encontrar os pontos de interesse, diversas técnicas podem ser utilizadas, tais como a identificação por cor, máxima variância, quinas ou luminosidade.

Após a eleição de diversos pontos de interesse os mesmos devem ser propriamente correlacionados com a maior precisão possível para que se possa efetuar os cálculos de profundidade com precisão.

Duas abordagens podem ser utilizadas para a realização da correlação dos pontos de interesse, a primeira abordagem consiste em achar pontos de interesse apenas na primeira imagem e procurar seus correspondentes na segunda imagem, porém esse processo torna-se custoso devido a quantidade de combinações existentes, a segunda consiste em encontrar pontos de interesse em ambas as imagens e depois tentar correlacioná-los, contudo alguns pontos encontrados podem não ter um par correto.

Nesse trabalho será utilizada a segunda abordagem, que é mais compatível com a função empregada pelo algoritmo SURF para encontrar os pontos de interesse, isso se dá devido ao seu bom desempenho mesmo com mudanças de escala, rotação, iluminação ou transformações geométricas[8].

2.3 Algoritmo SURF

O algoritmo SURF (Speeded Up Robust Features) [13] é uma algoritmo baseado nas somas de Haar-Wavelet, que funciona como um detector de características comumente utilizado dentro dos ramos da visão computacional.

Neste trabalho o algoritmo do SURF é aplicado nas duas imagens para a marcação de pontos de interesse, por ele ser robusto e ter um bom desempenho. O resultado da aplicação do algoritmo de SURF pode ser visualizado na figura 4.

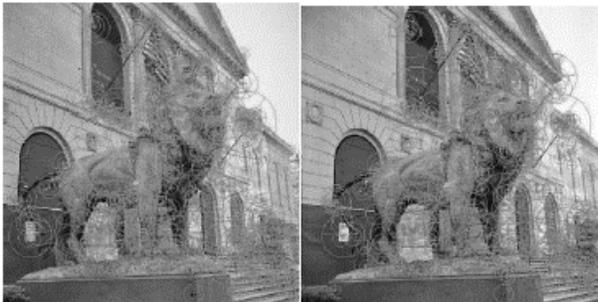


Fig 4. Imagens após aplicação do algoritmo SURF

Após a aplicação do SURF e passada a etapa de correlação, pode ocorrer de alguns pontos estarem correlacionados de forma errada, alguns pontos podem não encontrar o seu correspondente, ou acharem um correspondente errado, na figura 5 podemos verificar o resultado da correlação de forma bruta, com essas marcações equivocadas, que geram outliers.

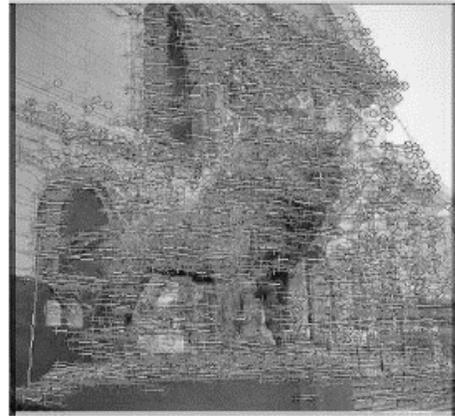


Fig 5. Correlação das imagens com imperfeições

2.4 Algoritmo RANSAC

De acordo com o que foi observado na figura 5 há correlações que se encontram fora do padrão, chamadas de outliers, essas correlações equivocadas afetam o cálculo de diparidades que por sua vez pode resultar em uma reconstrução 3D que foge ao padrão ideal.

O algoritmo do RANSAC (Random Sample Consensus) [14] é um método iterativo para estimar parâmetros de um modelo matemático, na perspectiva adotada pela reconstrução 3D nesse artigo o algoritmo do RANSAC realiza um aprimoramento do resultado das correlações de pontos de interesse.

O algoritmo do RANSAC foi utilizado para realizar um refinamento do resultado, eliminando as correlações outliers, de modo que dado um conjunto de pontos correlacionados, o algoritmo procura estabelecer um modelo matemático que agrupe o maior número de pontos(inliers), excluindo os pontos que tiveram a correlação fora do padrão(outliers) como pode ser visto na figura 6.

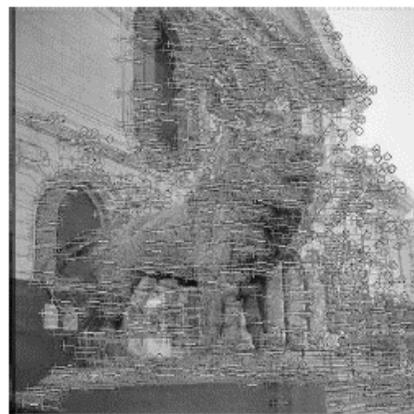


Fig 6. Correlação das imagens sem imperfeições

3 Resultados

Como descrito em seções anteriores o SURF obteve sucesso na marcação de pontos de interesse nas duas imagens, como observado na figura 4, contudo ao realizar correlação dos pontos alguns se correlacionaram de forma equivocada. Ao calcular a distância a esses pontos, valores fora do ideal são obtidos.

Após a aplicação do algoritmo RANSAC, os pontos que tiveram uma correlação diferente do padrão da maioria foram desprezados, com isso o vetor de pontos de profundidade, diminuiu em cerca de 5%. Observa-se na figura 7, o resultado final da reconstrução 3D

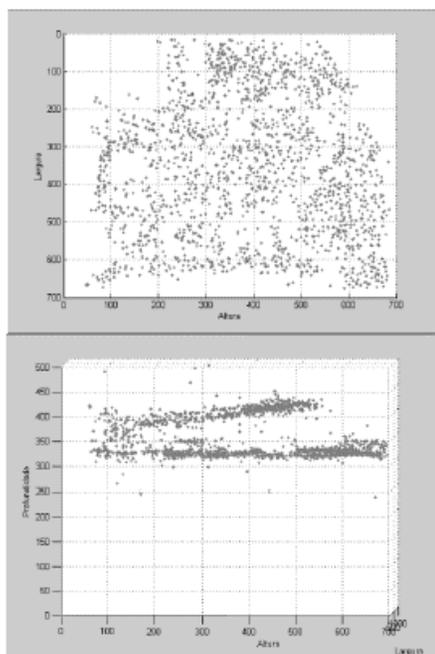


Fig 7. Reconstrução 3D

4 Conclusão

Este trabalho apresentou um algoritmo para reconstrução 3D utilizando a técnica da estereoscopia aliada aos algoritmos já existentes de visão computacional, o SURF, que marca pontos de interesse e o RANSAC que refina resultados.

Um dos problemas apresentados para a técnica de estereoscopia foi a marcação de pontos de interesse, em que foi utilizada uma abordagem que marca pontos de interesse nas duas imagens e tenta correlacioná-los, feita pelo SURF.

Outro problema explanado foram os outliers, correlações equivocadas realizadas por pontos marcados pelo algoritmo do SURF, e que foram removidas com o algoritmo do RANSAC que estima parâmetros para as correlações e remove as que fujam a esse padrão.

Futuramente visa-se a possibilidade de utilizar uma plataforma robótica equipada com duas câmeras que aplique a técnica descrita neste trabalho para realizar uma reconstrução 3D.

Referências

- [1] Kapoutsis, A.C., C.S.D.L.B.d.S.J.e.K.E.: Autonomous navigation of teams of unmanned aerial or underwater vehicles for exploration of unknown static dynamic environments. In: 21st Mediterranean Conference on Control Automation (MED). (2013), pp.1181-1188.
- [2] Salamuniccar, G., L.S.G.A.e.W.C.: Hybrid method for detection of lunar craters based on topography reconstruction from optical images. In: 2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA). (2011). pp.597-602.
- [3] Nygren, I e Jansson, M.: Terrain navigation for underwater vehicles using the correlator method. In: IEEE Journal of Oceanic Engineering, vol 29.(2004) pp. 906-915.
- [4] S. Liu, L.Z.e.J.L.: The applications and summary of three dimensional reconstruction based on stereo vision. In: International Conference on Industrial Control and Electronics Engineering, IEEE (2012) pp. 620-623.
- [5] Khattak, S.R.; GAMERLab. U.o.O.I.o.T.O.O.C..B.D..H.A.: Reconstructing 3d buildings from lidar using level set methods. In: Computer and Robot Vision (CRV), 2013 International Conference on, IEEE (2008) pp.151-158.
- [6] J. Han, L. Shao, D.X.e.J.S.: Enhanced computer vision with microsoft Kinect sensor: A review. In: IEEE Transactions on Cybernetics, vol. 43, IEEE (2013). Pp.1318-1334.
- [7] Ribeiro, G.P.E.: Estereoscopia no cálculo de distância e controle de plataforma robótica. (2012).
- [8] Szczuko, P.: Inuence of image transformations and quality degradations on surf detector

- e_cieny. In: Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), IEEE (2013). pp.285-290.
- [9] McAllister, D.F.: Stereo and 3-d display technologies. In: Encyclopedia of Imaging Science and Technology, John Wiley Son. (2002) pp.1327-1344.
- [10] Lipton, L.: Foundations of the stereoscopic cinema a study in depth. In: Van Nostrand Reinhold Company. (1982).
- [11] Foley, J.D.: Introduction to computer graphics. In: Addison-Wesley. (1993)
- [12] Galindo, J.R.F.: Analise da reconstrucao 3d a partir de um par estereoscopico hrccd/cbers-2, usando dois modelos matematicos. In: Disserta de Mestrado,UNESP (2008).
- [13] H. Bay, A. Ess, T.T., Gool., L.V.: Surf: Speeded up robust features. In: Computer Vision and Image Understanding. (2008). pp.346-359.
- [14] Fischler, M., Bollers., R.: Random sample consensus:a paradigm for model setting with applications to image analysis and automated cartography. In: Communications of the ACM. (1981). pp.381-395.