

Reconstrução 3D baseada em estereoscopia com a utilização de detectores de características

3D reconstruction based on stereoscopy utilizing feature detectors

Jonas Cordeiro do Nascimento Neto
Escola Politécnica de Pernambuco
Universidade de Pernambuco
50.720-001 - Recife, Brasil
jcn@ecom.poli.br

Bruno José Torres Fernandes
Escola Politécnica de Pernambuco
Universidade de Pernambuco
50.720-001 - Recife, Brasil
bjtf@ecom.poli.br

Resumo *A construção de mapas 3D é importante em diversas áreas do conhecimento, são algumas delas: topografia, exploração de terrenos desconhecidos, ambientes subaquáticos, em geral, onde a presença humana se torna inviável. Esse trabalho desenvolveu um modelo capaz de construir mapas 3D de um cenário, a partir de duas imagens 2D do mesmo ambiente com a utilização da estereoscopia, que torna possível o cálculo da distância que se encontram os objetos no cenário. A pesquisa também teve o objetivo de comparar os algoritmos: Scale Invariant Feature Transform (SIFT)[15], Speed Up Robust Features (SURF)[13] e Oriented FAST and Rotated BRIEF (ORB)[16] no processo de identificação de pontos nas imagens, e os algoritmos: Localitysensitive hashing (LSH), k-d tree e BruteForce no processo de correlação dos pontos identificados. Foram obtidas duas imagens do mesmo cenário com uma pequena distância entre elas e foram aplicados algoritmos de identificação de pontos, em seguida, foram aplicados algoritmos para correlacioná-los, após esse processo, foram utilizadas técnicas estatísticas para remoção de falsas correlações, por fim, através de geometria epipolar foi calculada a distância de cada ponto para a câmera e foi construído o mapa 3D. Durante a comparação dos algoritmos utilizados, foram analisados fatores como: número de pontos correlacionados corretamente, distribuição espacial dos pontos correlacionados e tempo de execução.*

Palavras-Chave: *Estereoscopia; Reconstrução 3D; Detetor de Característica; Visão Computacional*

Abstract *The 3D map construction is important in many study fields, some of those are: topography, unknown land exploration, subaquatic environments, in general, places where the human presence becomes inviable. This study developed a model capable of built 3D maps of a scene[1], from two 2D images of the same environment using stereoscopy, which makes possible the distance calculus of the objects in the scenario. The research also had the objective of comparing algorithms: Scale Invariant Feature Transform (SIFT)[2], Speed Up Robust Features (SURF) [3] and Oriented FAST and Rotated BRIEF (ORB) [4] in the process of identification points in the images, and the algorithms: Locality-sensitive hashing (LSH), k-d tree and BruteForce in the process of correlation of the identified points. Two images were obtained of the same scenario with a small distance between them, and were applied the algorithms for point identification, then, were applied algorithms to the point correlation, after those processes were utilized elastic techniques to remove false correlations, then, using epipolar geometry was calculated the distance between the each point and the camera, and the 3D map was built. During the algorithm comparisions factors like: number of points correctly correlated, spacial distribution of the correlated points and execution time were analised.*

Keywords: *Stereocopy, 3D Reconstruction, Feature Detector, Computer Vision*

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 Ranging[5] e o Microsoft Kinect[6].

Esse trabalho visa à investigação do modelo de Visão estereoscópica com o auxílio de algoritmos de detecção e correlação de características, para serem aplicados na reatuação 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 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 algoritmos detectores de características como o SIFT, SURF e ORB 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 simulasse a visão humana com duas câmeras alinhadas, mas com angulações diferentes, a partir dessas câmeras obtém-se duas imagens, uma esquerda e outra direita. Após a aquisição das imagens o algoritmos de detecção de características

são aplicados, definindo pontos de interesse em ambas as imagens, feita a identificação desses pontos de interesse, um algoritmo para correlação dessas características é aplicado, em seguida, é aplicado o algoritmo do RANSAC que elimina correlações erradas que não seguem um padrão e por fim é feito o cálculo das disparidades. O fluxograma abaixo resume os passos utilizados neste trabalho para a reconstrução 3D:

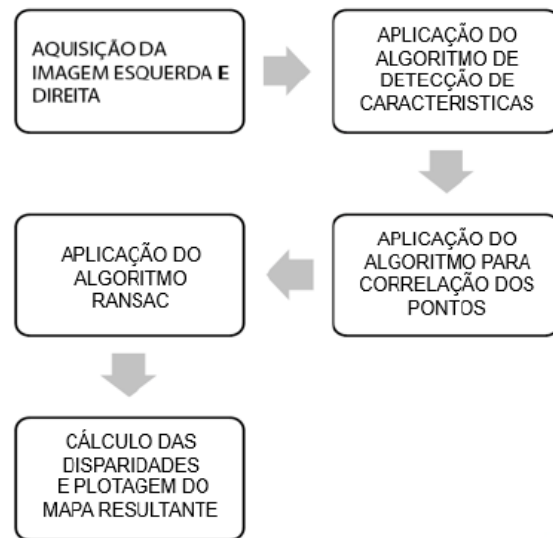


Figura 1: Fluxograma de representação da técnica proposta.

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.

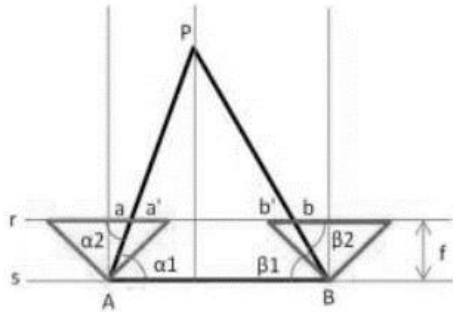


Figura 2: Representação do posicionamento das câmeras e do ponto de interesse.

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

Tabela 1: Símbolos e Notações

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.

$$D = \frac{fX}{|Xe - Xd|} \tag{1}$$

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.

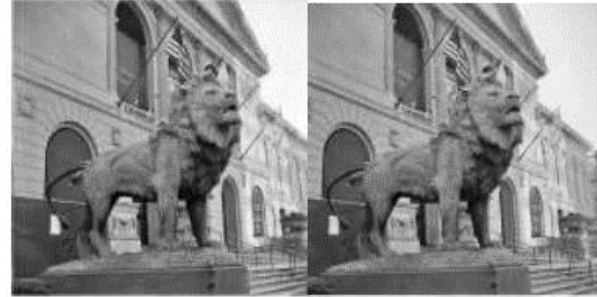


Figura 3: Imagens exemplo para a reconstrução 3D.

2.2 Identificação de Pontos de Interesse

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 ocluso, 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.

Nesse trabalho, 3 diferentes algoritmos de identificação de características vão ser utilizados:

- Scale Invariant Feature Transform (SIFT)
- Speed Up Robust Features(SURF)
- Oriented FAST and Rotated BRIEF (ORB)

Na figura 4 pode ser visualizado o resultado da aplicação do algoritmo SURF.

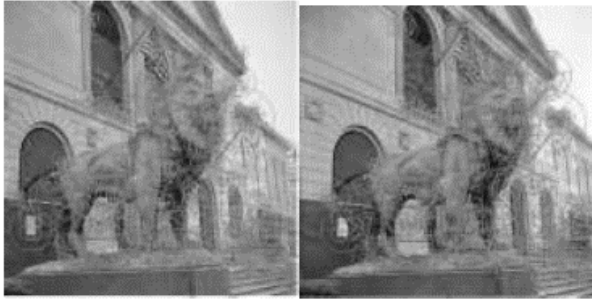


Figura 4: Imagens após aplicação do algoritmo SURF.

2.3 Correlação dos Pontos de Interesses

Após a eleição de diversos pontos de interesse o mesmo 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 algoritmos detectores de características 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], com a comparação de 3 técnicas:

- BruteForce
- Locality Sensitive Hashing (LSH)
- k-D Tree.

2.4 RANSAC

Após a etapa de correlação dos pontos alguns pares de pontos podem estar fora do padrão, essas correlações equivocadas afetam o cálculo de disparidades 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 5.

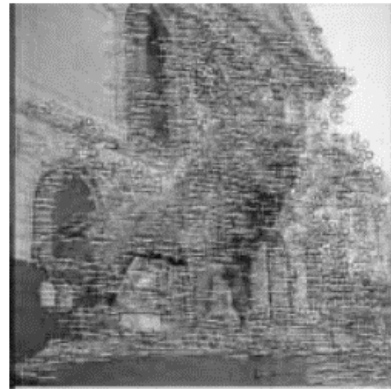


Figura 5: Correlação entre as duas imagens após aplicação do algoritmo RANSAC.

3 Resultados

Para comparação dos algoritmos, foram analisados fatores como:

- Número de pontos correlacionados corretamente;
- Distribuição espacial dos pontos correlacionados;
- Tempo de execução do processo;

As seguintes combinações de algoritmos de detecção e correlação de características existentes na biblioteca OpenCV[17] foram executadas :

- SIFT + BruteForce
- SIFT + k-d Tree
- SURF + BruteForce
- SURF + k-d Tree
- ORB + BruteForce

- ORB + LSH

Os resultados podem ser vistos na figura 6 e nas tabelas 2 e 3.

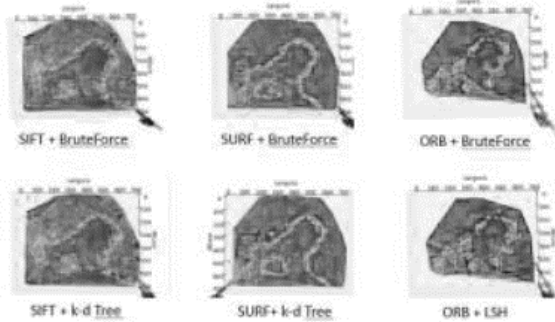


Figura 6: Resultados da Reconstrução 3D

Símbolo	SIFT + Brute-Force	SURF + Brute-Force	ORB + Brute-Force
Pontos identificados na Imagem 1	3506	3936	3500
Pontos identificados na Imagem 2	3266	3936	3500
Pontos Correlacionados	3266	1385	1421
Tempo de Execução (s)	2,18	1,94	0,58

Tabela 2: Resultados das combinações: SIFT + BruteForce, SURF + BruteForce e ORB + BruteForce

Símbolo	SIFT + k-D Tree	SURF + k-D Tree	ORB + LSH
Pontos identificados na Imagem 1	3506	3936	3500
Pontos identificados na Imagem 2	3266	3936	3500
Pontos Correlacionados	1377	1385	1445
Tempo de Execução (s)	1,92	1,67	0,453

Tabela 3: Resultados das combinações: SIFT + k-D Tree, SURF + k-D Tree e ORB + LSH

4 Conclusão

Este trabalho apresentou um modelo para reconstrução 3D utilizando a técnica da estereoscopia aliada aos algoritmos de detecção e correlação de características já para identificar pontos de interesse e o RANSAC para 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.

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

A combinação ORB+LSH foi a que mais correlacionou pontos de interesse e a mais rápida em tempo de execução, porém a dispersão dos pontos correlacionados foi muito concentrada, tornando o mapa final mais centrado em algumas regiões, é indicado para aplicações que necessitem de uma construção rápida do mapa.

A combinação SURF + k-D Tree foi um pouco mais lenta que a combinação acima, porém possuía uma distribuição espacial mais dispersa, tornando o mapa final mais detalhista, indicado para aplicações em que o mapa construído precisa fornecer informações sobre o maior número possível de objetos no cenário e o tempo não é um fator determinante.

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.: Influence of image transformations and quality degradations on surf detector efficiency. 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), 1991.
- [11] Foley, J.D.: Introduction to computer graphics. In: Addison-Wesley. (1993)
- [12] Galindo, J.R.F.: Análise da reconstrução 3d a partir de um par estereoscópico hrccd/cbers-2, usando dois modelos matemáticos. In: Dissertação 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.
- [15] Lowe, David G. Object recognition from local scale-invariant features. Proc. 7th International Conference on Computer Vision (ICCV'99) (Corfu, Greece): 1150-1157
- [16] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "Orb: An efficient alternative to sift or surf," in ICCV'11, 2011, pp. 2564–2571
- [17] OpenCV (Open Source Computer Vision). Disponível em:<http://opencv.org/>