

Visão Computacional

Relatório do Projeto 01

George Oliveira Barros (190052911)
george.barros@ifgoiano.edu.br
Prof. Dr. Teófilo de Campos

Departamento de Ciência da
Computação
Universidade de Brasília
Campus Darcy Ribeiro, Asa Norte
Brasília-DF, CEP 70910-900, Brazil,

Abstract

Este artigo descreve a elaboração das etapas do projeto 01, da disciplina de visão computacional, ministrada pelo professor dr. Teófilo de Campos. O projeto permitiu a compreensão prática das etapas de um sistema de visão estéreo e como realizar reconstrução 3d de uma cena a partir de uma imagem 2d. Os resultados obtidos para os 4 requisitos do projeto foram: a execução total dos requisitos de calibração das câmeras e estimação das poses das câmeras nas coordenadas do mundo (1 e 2), execução parcial do requisito de cálculo de mapas de disparidade e profundidade (3), e não execução do rastreamento de objetos 3d no vídeo (4). Os resultados das tarefas realizadas se mostraram aceitáveis e coerentes em análise feita na conclusão deste relatório.

1 Introdução

A Visão Computacional é um ramo da Inteligência Artificial que tem como objetivo extrair informação de fotos e vídeos. Parte fundamental da tarefa de um sistema completo de visão computacional é a etapa de aquisição desses dados visuais. Existem diferentes dispositivos de captura de vídeo e foto. No entanto, alguns conceitos básicos são essenciais para entender, em aspectos gerais, como uma imagem digital é adquirida e representada[1].

O processo de formação de uma imagem inicia-se a partir de uma câmera. Consideremos aqui como câmera o modelo *pinhole*. Os raios de luz emitidos sobre um dado objeto sofrem do fenômeno de reflexão. A partir da captura de um feixe de luz oriundo de um dado objeto, é possível capturar esse sinal representando-o em uma forma matricial (bidimensional) do objeto ou cena em questão. Ou seja, ocorre a conversão da cena no mundo real, em três dimensões, para a imagem ou vídeo, em duas dimensões [2].

A utilização de uma única câmera, no entanto, não permite ao sistema de visão computacional realizar uma reconstrução ou representação tridimensional de uma cena. Em outras palavras, a partir de uma única câmera, dada uma imagem capturada, não é possível reprojeta a cena do mundo real. Esta tarefa é importante para diversas aplicações atuais, como utilização de veículos autônomos, visão robótica, aplicações industriais e muito mais. Para realização desta tarefa, uma das soluções possíveis é a utilização não apenas de uma, mas, um par de câmeras. A este sistema se dá o nome de Visão Estéreo (*Stereo Vision*).

A partir da utilização de duas câmeras é possível através de um algoritmo de casamento de pontos, por exemplo, mapear para cada pixel de uma imagem sua respectiva posição nas coordenadas do mundo [1].

Os estudos sobre aquisição de imagens e especialmente relacionados a visão estéreo ainda possuem fronteiras a serem rompidas, portanto, atualmente há muitas pesquisas em andamento, sejam do ponto de vista de aplicações ou desenvolvimento de técnicas. A seguir alguns trabalhos publicados:

Em [2], ainda no ano 2000, foi proposto pelos autores, um sistema de visão stereo para navegação de robôs em tempo real. Em trabalhos mais recentes, [3] apresentaram um método de autocalibração para um sistema de câmera robótica, que pode ser calibrado sem qualquer intervenção manual, permitindo que a calibração do robô seja totalmente online e adequada à programação rápida da célula de trabalho combinada do robô e da visão computacional. [4] realizam a comparação entre o sistema de visão estéreo humano e de insetos e [5], ainda, utilizam um sistema de visão estéreo para rastrear cavernas subaquáticas.

Tendo em vista a importância do estudo sobre visão estéreo e as variadas possibilidades de aplicação destas técnicas, este projeto teve como objetivo a compreensão e implementação de técnicas de visão estéreo em um conjunto de dados visuais disponibilizados na disciplina de Visão Computacional, ministrada pelo Prof. Dr. Teófilo de Campos. As etapas do projeto foram: (1) Calibração das câmeras, (2) estimação da pose das câmeras nas coordenadas do mundo, (3) casamento de pontos e mapas de dissimilaridade e profundidade e por fim, (4) rastreamento 3D de objetos nos vídeos.

A seguir é apresentado, para cada etapa do projeto, a metodologia adotada e os resultados obtidos. No final ainda é realizada uma breve conclusão sobre o projeto. O código fonte pode ser encontrado na pasta enviada em anexo pela plataforma moodle (no caso do professor e monitores da disciplina).

2 Calibração das câmeras (Requisito 1)

A tarefa de calibração das câmeras também pode ser chamada de cálculo dos parâmetros intrínsecos. Nesta etapa o objetivo foi estimar uma série de informações relativas às câmeras. Há importantes informações associadas aos dispositivos de captura das imagens.

As informações estimadas nesta etapa foram: coeficientes de distorção e parâmetros intrínsecos e extrínsecos de cada uma das câmeras. Os coeficientes de distorção são importantes pois algumas câmeras *pinhole* apresentam um certo nível de distorção nas imagens, sendo a distorção radial e tangencial os principais tipos de distorção apresentados [6]. Os 5 parâmetros que formam o vetor de coeficientes de distorção são dados a partir do cálculo isolado dos coeficientes de distorção radial e tangencial, gerando assim: (*coefficients*=($k_1 \ k_2 \ p_1 \ p_2 \ k_3$)).

Entre os parâmetros intrínsecos encontram-se a distância focal (f_x, f_y) e centro óptico (c_x, c_y). Estes dados são comumente representados em conjunto, formando assim a matriz de intrínsecos, mostrada a seguir:

$$camera\ matrix = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}.$$

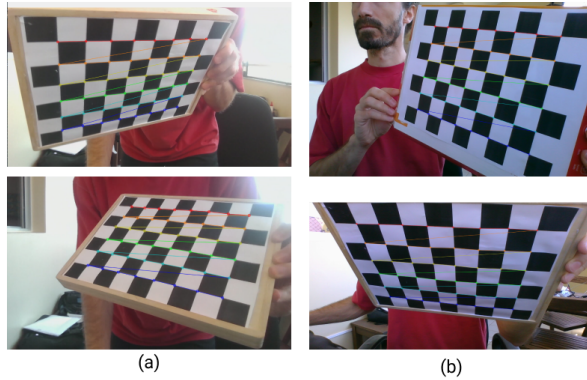


Figure 1: Calibração das imagens da câmera 1 (b) e câmera 2(a).

2.1 Metodologia

Para obter a matriz de intrínsecos e os coeficientes de distorção foi utilizado o método de utilização de um tabuleiro como referência de coordenadas do mundo. Neste método, inicialmente é realizada uma série de fotos do tabuleiro (*chess board*) em diferentes posições, ângulos e inclinações diante das câmeras. Como são duas câmeras distintas esse processo foi realizado para cada uma das câmeras. As fotos foram obtidas pelo professor e disponibilizadas junto com as instruções deste projeto. A ideia que está por trás é correlacionar pontos do mundo (3d) com pontos na imagem (2d), logo aí está a razão da utilização do tabuleiro, por facilitar a identificação destes pontos pelos métodos de identificação das esquinas e pontos no tabuleiro [1] e [2].

2.2 Resultados

A Figura 1 mostra os resultados da aplicação de método de identificação dos pontos na superfície do tabuleiro em imagens de cada uma das câmeras. Foram utilizadas 31 imagens para calibrar a camera 1 (da direita) e 15 imagens para calibrar a câmera 2 (da esquerda). A seguir também é apresentado o resultado obtido com a aplicação do método, obtendo-se a matriz de intrínsecos e os vetores de distorção das câmeras 1 (b - direita) e 2 (a - esquerda).

Resultados da câmera 1:

$$coefficients_1 = (1.84 \quad -1.09 \quad 7.62 \quad 1.99 \quad 1.46).$$

$$camera\ matrix_1 = \begin{bmatrix} 1.34 & 0 & 9.45 \\ 0 & 1.35 & 5.34 \\ 0 & 0 & 1 \end{bmatrix}.$$

Resultados da câmera 2:

$$coefficients_2 = (4.36 \quad 2.27 \quad 7.56 \quad -9.64 \quad -1.23).$$

$$camera\ matrix_2 = \begin{bmatrix} 953.0 & 0 & 627.4 \\ 0 & 954.8 & 344.4 \\ 0 & 0 & 1 \end{bmatrix}.$$

3 Estimando a pose das câmeras (Requisito 2)

Esta tarefa também pode ser chamada de cálculo dos parâmetros extrínsecos. Diferente da etapa anterior, onde o interesse era estimar as informações relacionadas às câmeras, nesta etapa, o objetivo foi calcular, a partir dos dados da calibração, onde as câmeras estavam situadas nas coordenadas do mundo.

3.1 Metodologia

Para estimar as poses das câmeras nas coordenadas do mundo foi necessário: (1) calcular os vetores de translação e rotação das câmeras, (2) encontrar 4 pontos no plano de coordenadas do mundo e suas respectivas localizações em um par de frames das duas câmeras e por fim, (3) calcular posição 3d das câmeras.

Os vetores de rotação e translação foram informações obtidas na calibração das câmeras, portanto, ao realizar a primeira etapa (requisito 1) esses dados foram calculados. Para relacionar os pontos reais (3d) com pontos nos frames foi necessário, inicialmente, encontrar a localização dos pontos nas imagens. Estes respectivos pontos foram obtidos através do click do mouse. Foram selecionados frames sincronizados, que tratassem de um mesmo momento do vídeo. Ao observar a Figura 2 pode-se ver que o carrinho está no tabuleiro, em ambas as imagens. A Figura 2 também mostra a localização bidimensional em cada um dos frames (câmera 2 e câmera 1). Os pontos dos objetos selecionados nas coordenadas do mundo (3D) foram informados pelo professor na descrição do trabalho.

Após a obtenção de todas essas informações, foi possível estimar a pose de um dado objeto em relação às câmeras. Para realizar esta tarefa foi utilizada uma função que encontra uma pose que minimiza o erro de reprojeção das câmeras, ou em outras palavras, minimiza a soma das distâncias quadradas entre as projeções observadas dos pontos nas imagens e as posições reais dos objetos. O resultado destas operações foram as matrizes de rotação e vetores de translação e rotação. Para encontrar a posição 3d de cada uma das câmeras bastou realizar a seguinte operação para cada uma delas, onde M é a matriz de rotação e R o vetor de translação: $posicao3d(x,y,z) = (-M^T * R)$.

3.2 Resultados

Para cada uma das câmeras foram calculadas os seguintes dados:

rotação₁ = (1.27 2.02 -1.10).
translação₁ = (0.32 -1.19 3.69).
cameraPose₁ = (3.51 1.34 0.99).
rotação₂ = (1.77 1.36 -0.74).
translação₂ = (-1.35 -0.96 3.18).
cameraPose₂ = (3.44 0.48 0.91).

4 Mapas de Disparidade e Profundidade (Requisito 3)

Após calibrar as câmeras e estimar as suas posições nas coordenadas do mundo, nesta etapa o objetivo do trabalho foi gerar os mapas de disparidade e profundidade relativos à câmera 1.

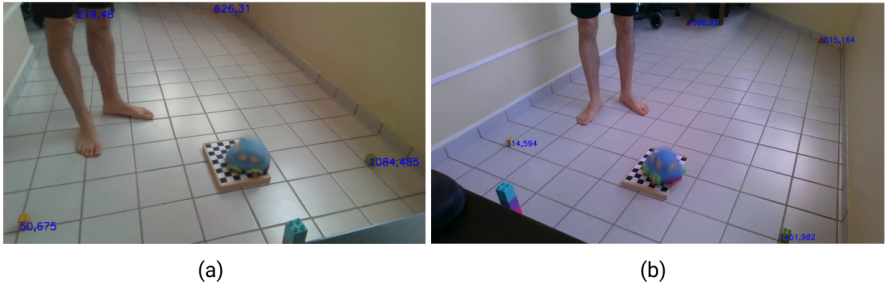


Figure 2: Localização dos objetos em um frame da câmera 1 (b) e câmera 2 (a).

A partir de duas imagens obtidas por um sistema de visão estéreo a disparidade é a diferença espacial entre os pixels das duas imagens, esquerda e direita. Também pode ser definida como a distancia horizontal entre pixels das duas imagens [8] [10]. O cálculo da disparidade se dá por: $d = |X_l - X_r|$.

A profundidade é obtida através da disparidade. O valor da disparidade entre os pixels da imagem podem ser convertidos em distâncias físicas[8]. Obtem-se os valores de profundidades a partir da seguinte equação, onde B representa a distância entre as câmeras e f representa a distância focal de cada câmera: $Z = (B \cdot f) / d$.

4.1 Metodologia

Esta tarefa foi realizada parcialmente, apenas foi gerado o mapa de disparidade. Para gerar o mapa de disparidade foi necessário antes de qualquer coisa redimensionar as imagens de calibração e recalibrar as câmeras. Em seguida, foi necessário utilizar os coeficientes de distorção das câmeras para remover eventuais distorções, depois aplicar o algoritmo de [10] para realizar o casamento de pontos e calcular a disparidade.

A primeira tarefa foi a de redimensionar as imagens de calibração das câmeras para facilitar o computo dos mapas de disparidade. As imagens foram redimensionadas para 600x380. Logo em seguido foi necessário recalibrar as câmeras com essas novas imagens.

Depois de recalibrar foi aplicada uma função de distorção, com a utilização dos coeficientes de distorção, para corrigir quaisquer distorções na imagem. A tarefa seguinte foi a aplicação do algoritmo de casamento de pontos de [10]. O resultado desse algoritmo foi um par de imagens mapeadas e prontas para realização do cálculo da disparidade.

4.2 Resultados

A Figura 3 mostra a projeção do mapa de disparidade obtido a partir da imagem da câmera 1 em relação a câmera 2.

5 Rastreamento de objetos (Requisito 4)

Requisito não concluído.

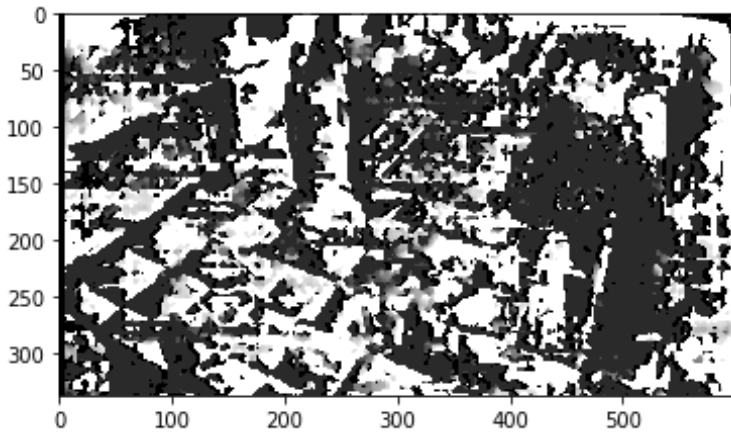


Figure 3: Mapa de disparidade obtido a partir da câmera 2 (da direita).

6 Conclusões

Apesar da conclusão parcial do requisito 3 e não realização do requisito 4, por problemas de implementação, o presente trabalho permitiu uma compreensão mais profunda do funcionamento de um sistema de visão estéreo. A estimação das poses das câmeras apresentou resultados bem coerentes, considerando a posição 3d dos objetos que foram utilizados como referência. O bom resultado de cálculo das poses também permitiu avaliar de maneira positiva a tarefa de calibração das câmeras, pois câmeras mal calibradas não permitiriam encontrar um bom resultado no cálculo das poses. Por fim, o resultado do cálculo do mapa de disparidade foi aceitável, apesar dos ruídos oriundos da alta variação de intensidade de pixels. Possíveis estratégias para dirimir o problema seria a aplicação de filtros de suavização e normalização com um intervalo menor de intensidade para cada pixel.

References

- [1] Kenji Hata and Silvio Savarese. CS231A Course Notes 1: Camera Models. Disponível em: http://web.stanford.edu/class/cs231a/course_notes/01-camera-models.pdf. Acessado dia 04 de outubro de 2020.
- [2] Kenji Hata and Silvio Savarese. CS231A Course Notes 4: Stereo Systems and Structure from Motion. Disponível em: http://web.stanford.edu/class/cs231a/course_notes/04-stereo-systems.pdf. Acessado dia 04 de outubro de 2020.
- [3] Kenji Hata and Silvio Savarese. CS231A Course Notes 3: Epipolar Geometry. Disponível em: http://web.stanford.edu/class/cs231a/course_notes/03-epipolar-geometry.pdf. Acessado dia 04 de outubro de 2020.
- [4] David A. Forsyth, Jean Ponce, Computer Vision: A Modern Approach, Second Edition, Prentice Hall, 2011

- [5] Xuanchen Zhang, Yuntao Song, Yang Yang, Hongtao Pan, Stereo vision based autonomous robot calibration, *Robotics and Autonomous Systems*, Volume 93, 2017, Pages 43-51, ISSN 0921-8890, <https://doi.org/10.1016/j.robot.2017.04.001>.
- [6] Vivek Nityananda, Ghaith Tarawneh, Sid Henriksen, Diana Umeton, Adam Simmons, Jenny C.A. Read, A Novel Form of Stereo Vision in the Praying Mantis, *Current Biology*, Volume 28, Issue 4, 2018, Pages 588-593.e4, ISSN 0960-9822, <https://doi.org/10.1016/j.cub.2018.01.012>.
- [7] Murray, D., Little, J.J. Using Real-Time Stereo Vision for Mobile Robot Navigation. *Autonomous Robots* 8, 161–171 (2000). <https://doi.org/10.1023/A:1008987612352>
- [8] N. Weidner, S. Rahman, A. Q. Li and I. Rekleitis, "Underwater cave mapping using stereo vision," 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 2017, pp. 5709-5715, doi: 10.1109/ICRA.2017.7989672.
- [9] Camera Calibration. Página Oficial do OpenCV. Disponível em: https://docs.opencv.org/master/dc/dbb/tutorial_py_calibration.html acessado dia 03 de outubro de 2020.
- [10] Ribeiro, Antônio Horta. Implementação de uma câmera estéreo. TCC, Engenharia Elétrica. Orientador: Guilherme Augusto Silva Pereira, UFMG. 2015.
- [11] Heiko Hirschmuller. Stereo processing by semiglobal matching and mutual information. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 30(2):328–341, 2008