



FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Departamento de Engenharia Eletrotécnica e de Computadores

Visão por Computador

2017/2018

Trabalho Prático Nº1

Introdução à Parte Prática

Eduardo Filipe da Fonseca Rodrigues
2007105369

Parte 1: *Monocular Visual Odometry*

a)

Este exemplo tem como objetivo calcular uma estimativa da trajetória 3D de uma câmara através de uma sequência de vistas 2D capturadas pela mesma, neste caso, as vistas 2D são geradas por computador, não tendo sido capturadas por uma câmara física.

A trajetória estimada é então representada, assim como a trajetória real, para se comparar o erro entre ambas.

b)

A pose da câmara é definida como a sua posição espacial e orientação. No caso da câmara real ("ground truth camera"), são posições e orientações já incluídas no exemplo.

c)

Para calcular o deslocamento da câmara, são utilizados os deslocamentos dos objetos, ou pontos, representados nas imagens. No entanto, o tamanho desses objetos ou a distância a que se encontram da câmara é desconhecida.

Num sistema de odometria visual seria utilizado outro sensor para determinar distâncias desconhecidas de modo a fazer a relação entre a distância percorrida e a distância a que um objeto se encontra, ou o seu tamanho real.

Por exemplo, ao se observar um deslocamento de uma caixa na imagem para a esquerda, poderemos afirmar que a câmara se moveu para a direita, mas apenas sabendo o tamanho da caixa e a distância a que esta está da câmara, poderia ser calculada a distância que a câmara foi deslocada.

d)

Ao correr o exemplo, a trajetória estimada da câmara é representada em ambiente 3D, indicando a posição da câmara através de pontos e a orientação através do ícone representativo de câmara.

Duas trajetórias são apresentadas, a trajetória calculada através do algoritmo de estimativa, representada a verde, e a trajetória real da câmara, representada a azul.

Apesar de toda a informação calculada, posição e orientação, ser mostrada em tempo real, no final do exemplo pode apenas visualizar-se a trajetória e a orientação final de ambas as câmaras.

Parte 2: *Structure From Motion From Multiple Views*

a)

Este exemplo tem como objetivo calcular a estrutura 3D de um objeto através de um conjunto de vistas 2D do mesmo.

b)

Para calcular a posição espacial de um ponto através de várias vistas, é necessário antes saber a posição e orientação dessas mesmas vistas.

O exemplo anterior utiliza apenas os pontos de interesse das imagens para calcular a nova pose da câmara em relação à anterior, este exemplo estima também a pose da câmara entre as várias vistas para depois representar em ambiente 3D os pontos de interesse observados das imagens.

c)

O problema do factor de escala existe também neste exemplo, sendo a pose da câmara calculada através da pose da câmara anterior. No entanto, como neste exemplo são apenas utilizados 5 vistas. O exemplo anterior apenas corrige o factor de escala após as 3 primeiras vistas.

d)

Na segunda parte do “script” as imagens são analisadas de novo, no entanto, é utilizada a função “*detectMinEigenFeatures*” para detetar os pontos de interesse e criado um “*tracker*” para seguir esses pontos através das várias imagens. Ao contrário da primeira metade do “script” onde os pontos de interesse da imagem eram calculados e comparados com os pontos de interesse da imagem anterior, na segunda metade, os pontos de interesse são calculados através do deslocamento que tiveram entre a imagem anterior e actual.

e)

O “script” começa por guardar a diretoria onde se encontram as imagens utilizadas e representar as mesmas uma matriz de 3×2 .

Estas imagens são então convertidas para escala de cinza.

Através dos parâmetros da câmara, a primeira imagem é recuperada, eliminando a distorção que a câmara possa ter provocado.

As características da primeira imagem são então extraídas e é criada uma primeira vista com posição na origem e orientação segundo o eixo z.

Para as seguintes imagens, o processo é repetido, no entanto, após a extração das características da imagem, estas são relacionadas com as características extraídas da imagem anterior. Estes pares de características comuns entre as vistas são então utilizados para calcular a posição relativa da segunda vista em relação à primeira e guardados como ligação entre as duas vistas.

Utilizando os pontos de interesse entre as várias vistas é efetuada a triangulação da posição espacial destes e refinada a pose e posição dos pontos.

Os pontos são então apresentados em ambiente 3D, excluindo pontos com alto ruído, assim como as várias poses de câmara.

Na segunda metade do “script” é efetuada a mesma operação utilizando um processo diferente, já descrito na alínea anterior.