PROCESSAMENTO DE IMAGEM E VISÃO

Academic Year 2017-2018

Dept Electrical and Computer Engineering – Instituto Superior Técnico

Test: #2 Duration: 2 hours

---

1) **[5 pts, Features]** We want to design a computer vision system that extracts image features to classify image categories (e.g., beach, forest, urban scenes).

2 a) Describe the advantages and drawbacks of SIFT features when compared to HARRIS features

2 b) Describe the steps of the SIFT features pipeline: (i) keypoint detection; and (ii) patch description;

1 c) Describe how to use the SIFTs for image recognition using the bag-of-features approach and discuss its limitations.

2) **[5pts, Texture and colour]** In order to identify pedestrians walking inside a large area covered by a network of cameras with no overlapping fields of view, we will use information about colour and texture

2 a) Explain how the colour histogram could be used to complement the description of the different individuals (describe the method and how to compare different patches)

2 b) Explain the meaning of "texture" and how it can be characterized with gradient information

1 c) Explain how the co-occurrence matrices can be used to characterise image patches, which metrics can be used to compare local texture, and calculate the co-occurrence matrices in the following example:

$$Img = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 2 & 2 & 2 & 3 \\ 1 & 3 & 2 & 3 \\ 3 & 2 & 3 & 3 \end{bmatrix} \quad P_{(0,1)} = \begin{bmatrix} \cdot\cdot & \cdot\cdot & \cdot\cdot \\ \cdot\cdot & \cdot\cdot & \cdot\cdot \\ \cdot\cdot & \cdot\cdot & \cdot\cdot \end{bmatrix} \; and \; P_{(1,0)} = \begin{bmatrix} \cdot\cdot & \cdot\cdot & \cdot\cdot \\ \cdot\cdot & \cdot\cdot & \cdot\cdot \\ \cdot\cdot & \cdot\cdot & \cdot\cdot \end{bmatrix}$$

3) **[5pts, Stereo vision]** Consider a pair of calibrated cameras characterized by their projective camera models $\tilde{x}_i = \tilde{P}_i \tilde{X}$ and $i = 1, 2$, with:

$$\tilde{P}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ and } \tilde{P}_2 = \begin{bmatrix} 1 & 0 & 0 & -10 \\ 0 & 1 & 0 & -10 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

2 a) Explain the concept of epipolar lines and epipoles, and how they are useful for stereo analysis.

2 b) Calculate the epipoles of the two cameras and explain the orientation of the epipolar lines.

1 c) Calculate the fundamental matrices F_{12} and F_{21} and explain how they can be used for stereo.

4) **[5pts]** Classify the following statements, as false or true, and justify your choices in detail.

1 a) T The Hough transform can be used to detect elliptic shapes in images.

2 b) F At each image point, the optical flow can be estimated from local spatio-temporal image derivatives.

2 c) F To estimate the epipolar geometry, it is necessary to have the cameras calibrated.

Please read the questions carefully and derive the answers using mathematical models when appropriate.

TÉCNICO LISBOA

PIV
2017-18

TEST #2
SOLUTIONS

Problem 1

a) The Harris features are not scale invariant and can only tolerate a certain amount of image rotation, contrast. The ~~Harris~~ SIFT can also show invariance to scale change. On the other hand the SIFTs are more demanding computationally and include a representation for the local image patches which the Harris method does not consider.

b) SIFT's pipeline

(i) Key point detection: Define a set of scales and filter the image with Difference-of-Gaussian filter applied in the scale space. Keypoints correspond to extrema of the output signal. The Difference-of-Gaussian is an approximation of the scale function Laplacian of Gaussian, and is used for speeding up the execution.
Those keypoints where the gradient is low are eliminated (This step is similar to Harris analysis).

(ii) Each keypoint is associated to a reference orientation. The reference orientation is extracted from the histogram of the gradient orientation from 16×16 windows around the keypoint.

(iii) Patch description: We take a 16×16 window around the keypoint and divide it in 4×4 windows (each with 4×4 pixels). For each window we calculate the gradient magnitude and orientation and calculate the

histogram of orientation with 8 bins. This generate a $4 \times 4 \times 8 = 128$ feature vector that is used to describe the local patch.

c) The SIFT features can be compared with different metrics, e.g. the Euclidean distance, the nearest point (with a threshold) or the two closest neighbors

The bag of visual words or visual features consists in finding a set of features without imposing a particular structure between these features. The limitations result from the lack of a specific relation (e.g. geometry) between features (e.g. the order doesn't matter).

Problem 2

a) We can take an image patch and use a colour representation (RGB or HSV) to represent colour. Once we normalize w/ respect to intensity: $r = \frac{R}{R+G+B}$; $g = \frac{G}{R+G+B}$ we can build a (two-dimensional) histogram to compare two image patches and use an appropriate distance metrics: Euclidean distance, Minn distance or Kullback divergence.

b) Texture refers to the distribution of pixel values, in terms of their spatial organization, in an image.
The gradient histogram is obtained by estimating the gradient information of an image patch and calculate the histogram of the orientation a magnitude. One popular choice is the histogram of orientation modulated by the intensity.

TÉCNICO LISBOA

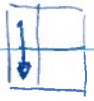
c) The co-occurrence matrix accounts how many times a certain pair of pixel values appear in a specific orientation.

$$Img = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 2 & 2 & 2 & 3 \\ 1 & 3 & 2 & 3 \\ 3 & 2 & 3 & 3 \end{bmatrix}$$

$P_{(0,1)}$ =

	1	2	3
1	0	1	2
2	1	2	2
3	0	2	2

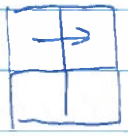
(0,1) →



$P_{(1,0)}$ =

	1	2	3
1	0	0	2
2	1	2	3
3	0	3	1

(1,0) :



Problem 3

a) Epipolar lines represent the possible location of homologous points in a stereo system. For each image projection in the first camera, the corresponding points in the second image are constrained to this line.
The epipolar lines have a common point - the epipole. The epipole results from the projection of each camera optic center in the other camera's image.

b)

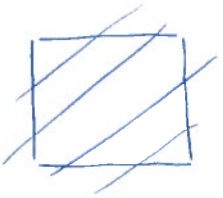
$P = K[R \mid T]$ in this case the intrinsics are $K = I$ and $R = I$.

We then have

$$O_1 = -RT_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad O_2 = -RT_2 = \begin{bmatrix} 10 \\ 10 \\ 0 \end{bmatrix}$$

$$e_1 = P_1 O_2 = P_1 \begin{bmatrix} 10 \\ 10 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \\ 0 \end{bmatrix} \qquad e_2 = P_2 O_1 = P_2 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -10 \\ -10 \\ 0 \end{bmatrix}$$

The epipoles are at infinity, meaning that the epipolar lines are parallel and have direction of 45°.



- epipolar lines.

c) $F_{12} = [e_2]_\times P_2 P_1^{-1}$

$P_2 P_1^{-1}$ → First 3 columns of matrices P_1 and P_2 which is identity

$[e_2]_\times$ → anti-symmetric matrix expressing the vector product.

$$= \begin{bmatrix} 0 & -e(3) & e(2) \\ e(3) & 0 & -e(1) \\ -e(2) & e(1) & 0 \end{bmatrix}$$

$$F_{12} = \alpha \begin{bmatrix} 0 & 0 & -10 \\ 0 & 0 & 10 \\ 10 & -10 & 0 \end{bmatrix} \quad \text{c/ } \alpha \neq 0 \qquad \underline{F_{21} = F_{12}^T}$$

TÉCNICO LISBOA

~~Problem~~

F_{12} can be used to calculate the epipolar lines corresponding to points in the first image:

$\underset{\sim}{m}_1$ - point in the first image

$\underset{\sim}{\mu}_2^2$ - epipolar line corresponding to $\underset{\sim}{m}_1$, in the second image.

$\underset{\sim}{\mu}_1^2 = \underset{\sim}{F}_{12}\, \underset{\sim}{m}_1$ $\quad$ $\underset{\sim}{F}_{12}$ maps points in I_1 to epipolar lines in I_2.

Problem 4

a) TRUE. Ellipses are parametric: $\frac{(x-x_c)^2}{a} + \frac{(y-y_c)^2}{b} = 1$

with a, b representing the axes of the ellipse.
As a parametric curve the standard Hough Transform can be used to find ellipses.

b) NO. With local spatio-temporal information, we can only estimate the normal flow, the component of the optical flow that has the direction of the gradient (i.e. perpendicular to the contours). To estimate the complete optical flow we have to use regularization, or geometric assumptions.

c) NO. If we do not know the camera matrices, we can still use 8 pairs of point correspondences and calculate the fundamental matrix F. With F we can calculate the epipoles, epipolar lines, etc...