PROCESSAMENTO DE IMAGEM E VISÃO

Academic Year 2018-2019

Dept Electrical and Computer Engineering – Instituto Superior Técnico

Test: #2 Duration: 2 hours

- Solve each problem on a separate sheet (may be stapled/bound together).
- Read all questions carefully and give concrete, structured answers. Use mathematical models when appropriate.
- You may use on A4 sheet of paper with formulas, but not with a "nano-transcript" of the slides.
- Only simple calculators are allowed.

---

1) **[5 pts, Features]** We want to design a computer vision system for the detection and recognition of traffic signs during autonomous driving. The detection/recognition of the traffic signs is done with image features/keypoints.

2 a) Describe the importance of invariance to <u>orientation and scale</u> for image keypoint detection and how such invariance can be achieved.

1.5 b) Describe the approach adopted by the SIFT features method for describing <u>local image patches</u> and justify the degrees of invariance provided by the different steps to build the descriptor.

1.5 c) Describe possible metrics for comparing SIFTs descriptors and their advantages and drawbacks.

2) **[5pts, Image Alignment, recognition]** A drone carrying an optical camera is used to inspect bridges and other critical infrastructures. The different structure parts are identified by unique colour-textured markers.

2 a) Explain how to identify the marker shown in the image based on colour information, e.g. how to build colour descriptors and compare them. Each marker has a unique colour pattern.

1.5 b) Having detected a (possible and approximate) position of the marker in the image, explain how to use the Lucas-Kanade algorithm for verification, by aligning the current (noisy, partially-occluded) view of the marker with a canonical, frontal-view, stored in advance.

1.5 c) After finding the predefined marker and orienting the camera optical axis to become perpendicular to the surface, explain how to use optical flow information to: (i) approach the surface to a prescribed distance and (ii) move laterally to scan the surface with the image plane parallel to the surface.

3) **[5pts, Stereo vision]** Consider a binocular stereo system, with two cameras described by projective matrices given by $\tilde{x}_i = \tilde{P}_i \tilde{X}$ and $i = 1,2$ and assuming the intrinsic parameters have been calibrated (i.e $K = I_{3\times3}$):

$$\tilde{P}_1 = \begin{bmatrix} 1 & 0 & -1 & 20 \\ 0 & \sqrt{2} & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}, \text{ and } \tilde{P}_2 = \begin{bmatrix} 1 & 0 & 1 & -20 \\ 0 & \sqrt{2} & 0 & 0 \\ -1 & 0 & 1 & 0 \end{bmatrix}$$

2 a) Calculate the relative position and orientation of the two cameras and draw a scheme showing their relative geometric configuration.

1.5 b) Calculate the epipoles and the fundamental matrix, F_{12}, that associates projections in image I_1 to epipolar lines in image I_2. Verify the properties of the epipolar matrix. → fundamental

1.5 c) Demonstrate analytically, that there is a one-to-one correspondence between epipolar lines in the two stereo images.

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

1 each

a) The Hough transform can only be used to detect circles in an image if the radius is fixed in advance.

b) All the epipolar lines of a stereo system intersect at the image principal point.

c) Each fundamental matrix corresponds to a unique pair of projection matrices, $\tilde{P}_1$ and $\tilde{P}_2$.

d) The *Snakes* approach is well suited to detect parallel lines in images.

e) The principal components analysis can be used to measure the presence of noise in a dataset.

TÉCNICO LISBOA

PIV - Image and Vision Processing

2018/19 Test #2

(draft resolution)

P1

a) Invariance to orientation and scale are important to deal with the different viewpoints when images

Orientation: it is difficult to deal with 3D rotations. Handling 2D rotations can be done with rotationally symmetric operators like the norm of the gradient or the Laplacian

Scale: can be done with a scale function that peaks at a specific local scale. Searching for maxima of the Laplacian of Gaussians, while varying the value of σ is a good option.

b) The SIFT patch descriptor takes the following steps

- From a 16x16 patch, create a histogram of the gradient orientation and define a canonical orientation for normalization (invariance to rotation)
- Scaling invariant results from the detection part as an estimate of the local scale can be used to normalize windows
- The descriptor is formed with 4x4 8-bin histograms of the gradient of the rotated patch. The use of gradient information provides invariance to illumination offsets.

TÉCNICO LISBOA

c) The SIFT-descriptor is a 128-component long vector. Vectors can be compared with the Euclidean distance:

$d_{ij} = \| d_i - d_j \|$ but usually it is required

the distance to meet two criteria for absolute quality check.

a) $d_{ij} < \text{threshold}$

b) $d^{(1)} / d^{(2)} < \delta$ where $d^{(1)}$ is the best choice and $d^{(2)}$ the second best-choice

An alternative method could be the inner product.

$| d_i \cdot d_j | < \delta$.

The Euclidean distance is often used as it is easy to define thresholds. The other variations include reasoning about more points than the best choice only.

Problem 2

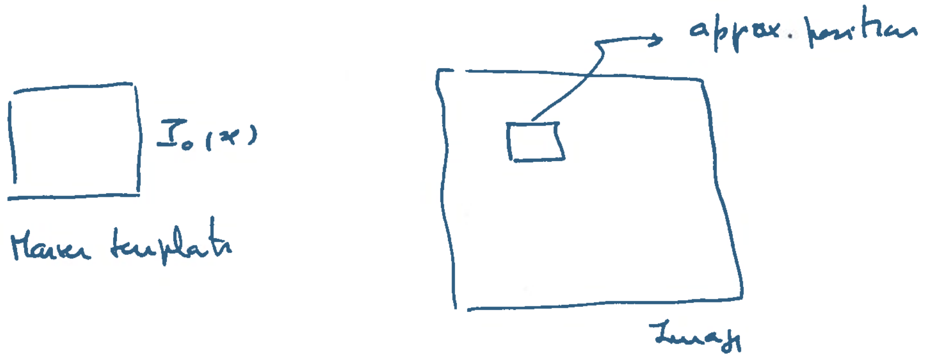
1) The target/marker has a unique texture / colour distribution. One possibility would be to segment the target based on colour but the colour is not necessarily unique (texture).

Use color histograms in R,G,B or, better, normalised R,G. Given the model image, we create the histogram P(k) and compare to colour histograms of candidate windows in the image $P_c(k)$

The metrics can be $\|P_1 - P_2\|$ (Euclidean norm), minimum distance, minimum, a Kullback divergence

$$\sum_k \min(P_1(k), P_2(k)) \quad , \quad \sum_k P_1(k) \log \frac{P_2(k)}{P_1(k)}$$

b) We assume we have found an initial position of the template in the image:



We can use the Lucas-Kanade with parametric motion

$$E(p) = \sum_i \left[I_0(x_i) - I_1(f(x_i; p)) \right]^2$$ with $f(x_i, p)$

representing a homographic transformation (the marker is on a flat region).

The cost is minimized iteratively (we already have the initial estimate).

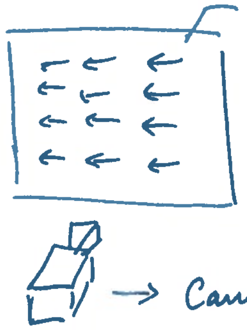
c) Assuming we have the camera optical axis perpendicular to the wall, we have to:

(i) Approach the wall at a constant speed until reaching a predefined value of the time-to-collision

$$u = \frac{t_z}{Z}(x - x_{FOE}) \quad \text{and} \quad \frac{Z}{T_z} = \frac{x - x_{FOE}}{u}$$ is the time to collision

TÉCNICO LISBOA

iii) Moving laterally can be done ensuring the optic flow vectors are parallel.



Problem P3

a) $P_1 = \sqrt{2}\left[\begin{array}{ccc|c} 1/\sqrt{2} & 0 & -1/\sqrt{2} & 20/\sqrt{2} \\ 0 & 1 & 0 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 \end{array}\right]$; $P_2 = \sqrt{2}\left[\begin{array}{ccc|c} 1/\sqrt{2} & 0 & 1/\sqrt{2} & -20/\sqrt{2} \\ 0 & 1 & 0 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 \end{array}\right]$

$R_1 = Rot(y, -\pi/4)$

$Rot(y, +\pi/4) = R_2$

$R_2 = R_1^T$

Optic center O_1, O_2

Let

$P_i = K_i [R_i \mid T_i]$

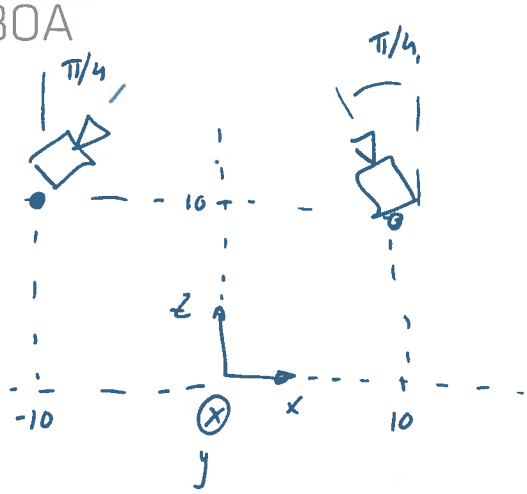
$O_i = -R_i^{-1} T_i$

"minus"

$O_1 = -Rot(y, +\pi/4)\begin{bmatrix} 20/\sqrt{2} \\ 0 \\ 0 \end{bmatrix}$ $O_2 = -Rot(y, -\pi/4)\begin{bmatrix} -20/\sqrt{2} \\ 0 \\ 0 \end{bmatrix}$

$= \begin{bmatrix} -10 \\ 0 \\ +10 \end{bmatrix}$ $= \begin{bmatrix} 10 \\ 0 \\ 10 \end{bmatrix}$

TÉCNICO LISBOA



b) Epipolos

In camera $\underline{2}$

$$e_2 = \underset{\sim}{P_2}\begin{bmatrix} O_{c_1} \\ 1 \end{bmatrix} = -R_2 R_1^{-1} T_1 + T_2$$

$$= -Rot(y, \pi/4)\, Rot(y, \pi/4)\, T_1 + T_2$$

$$= -Rot(y, \tfrac{\pi}{2})\, T_1 + T_2$$

$$= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} -20/\sqrt{2} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -20/\sqrt{2} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -20/\sqrt{2} \\ 0 \\ 20/\sqrt{2} \end{bmatrix} \sim \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

$$e_1 = -R_1 R_2^{-1} T_2 + T_1$$

$$= -Rot(y, -\pi/2)\, T_2 + T_1$$

$$= \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 20/\sqrt{2} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 20/\sqrt{2} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 20/\sqrt{2} \\ 0 \\ 20/\sqrt{2} \end{bmatrix} \sim \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$[a]_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}$$

$$F_{12} = [e_2]_\times R_2 R_1^{-1}$$

$$= \underbrace{\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}}_{[e_2]_\times} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

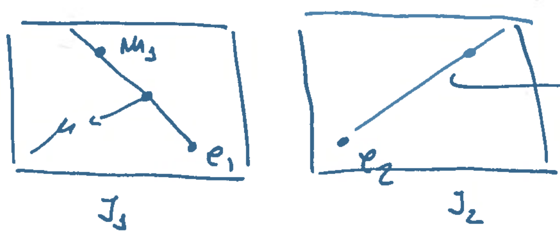
$$= \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

$\det(F_{12}) = 0$ since $F_{12}\, e_2 = \emptyset$.

$$F_{21} = F_{12}^T$$

c) From the fundamental matrix:

$$m_2^T F_{12}\, \underset{\sim}{m_1} = \emptyset$$



$\mu_e = \underset{\sim}{F_{12}}\, \underset{\sim}{m_1}$

We consider a line in I_1

$$\underset{\sim}{m} = \alpha\, \underset{\sim}{e_1} + \beta\, \underset{\sim}{m_1}$$

The epipolar line of $\underset{\sim}{m}$ is: $\mu_e = F_{12}(\underset{\sim}{m})$

$$= \alpha F_{12}\, \underset{\sim}{e_1} + \beta\, \underset{\sim}{F_{12}}\, m_1 \quad , \text{ and } F_{12}\, e_1 = \emptyset$$

$$= F_{12}\, \underset{\sim}{m_1} \text{ , the epipolar line of } m_1$$

This shows that lines μ & μ_e are mutually correspondent.

TÉCNICO LISBOA

Problem 4

a) FALSE. We can also incorporate variations in "r" by having a 3D accumulator / Hough space (x, y, r) or casting votes in circular sectors.

b) FALSE. They intersect at the epipole. It may be coincident with principal point if the cameras are translated along z, e.g.:

c) ~~they~~ FALSE. There are infinite solutions of pairs P_1, P_2 consistent with a given F matrix. They differ by a 4x4 projective transformation in $\mathbb{P}^3$.

d) FALSE. It is primarily used for curved contours without parametric description. Combines internal (smoothness) and external (gradient, potential) terms in the iterative optimization.

e) FALSE: PCA is used to find a linear subspace approximation of a larger dimensional space like images. (the eigenvalues of the covariance matrix express distance to the subspace ....)