# Vision 3D artificielle - Final exam (duration: 2h30)

P. Monasse, M. Aubry, and R. Marlet

November 29th, 2016

You can choose to answer in French or English, at your convenience.

## 1 Fast P$n$P

The P$n$P (Perspective from $n$ Points) problem aims at finding the position and orientation of a camera knowing the camera calibration matrix $K$, a family of $n$ 3D points $X_i$ (expressed in a world coordinate system) and their 2D projection in the camera $x_i$. The method proposed by Lepetit, Moreno-Noguer and Fua in 2008, called EP$n$P, is remarkable in the sense that it is simple to implement and that its complexity depends linearly on $n$.

1. Show that given four 3D points $C_j^w$ ($j = 1, \dots, 4$), we can write in general

$$\forall i, \quad X_i = \sum_{j=1}^{4} \alpha_{ij} C_j^w \text{ with } \sum_{j=1}^{4} \alpha_{ij} = 1.$$

   What does the term "in general" mean precisely in the sentence above?

2. Suppose that the points $C_j^w$ are expressed in *the camera coordinate frame* as vectors $C_j = RC_j^w + T$, with $R$ and $T$ the (unknown) rotation and translation between the camera and world coordinate frames. Show that

$$\forall i, \exists \lambda_i \neq 0, \quad K^{-1} x_i = \lambda_i K^{-1} \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} = \sum_{j=1}^{4} \alpha_{ij} C_j.$$

3. Prove that the vector $x = \begin{pmatrix} C_1^T & C_2^T & C_3^T & C_4^T \end{pmatrix}^T$ satisfies a linear system of equations $Mx = 0$, the $2n \times 12$ matrix $M$ depending on the $u_i$, $v_i$, $\alpha_{ij}$ and $K$.

4. Deduce that $x$ is in the kernel of $M$ and that we can write $x = \sum_{k=1}^{N} \beta_k V^k$, $V^k$ being the rightmost $K$ columns of $V$ in the SVD of $M = U \Sigma V^T$. What does $N$ mean?

5. If $n \geq 6$, we have normally $N = 1$, but because of noise, it is better to test for different values of $N$, $1 \leq N \leq 4$, the one that minimizes the sum of square distances $\sum_i d^2(x_i, K \sum_j \alpha_{ij} C_j)$. Justify that this distance should ideally be null.

6. Show that (using Matlab notation) for $i, j \in \{1, \dots, 4\}$

$$\| \sum_{k=1}^{N} \beta_k \left( V^k(3i-2:3i) - V^k(3j-2:3j) \right) \| = \|C_i^w - C_j^w\|.$$

7. If $N = 1$, show that we can write

$$\beta_1 = \pm \frac{\sum_{i,j=1}^{4} \|V^1(3i-2:3i) - V^1(3j-2:3j)\| \, \|C_i^w - C_j^w\|}{\sum_{i,j=1}^{4} \|V^1(3i-2:3i) - V^1(3j-2:3j)\|^2}.$$

8. If $N = 2$, show that we have $L \begin{pmatrix} \beta_1^2 \\ \beta_1\beta_2 \\ \beta_2^2 \end{pmatrix} = \rho$ with $L$ a $6 \times 3$ matrix and $\rho$ a 6-vector depending on the $C_i^w$. Propose a solution to recover $(\beta_1, \beta_2)$.

9. If $N = 3$, show that we have some similar sytem but with $L$ $6 \times 6$.

10. If $N = 4$, show that the vector $\beta$ composed of the 10 unknowns $\beta_{ij} = \beta_i\beta_j$, with $1 \leq i \leq j \leq 4$, could be determined up to an unknown vector in the kernel of a $6 \times 10$ matrix $L$. Propose additional equations in the $\beta_{ij}$ that would permit to recover this unknown vector.

11. Recovering $R$ and $T$ from the data of $C_j^w$ and $C_j$ is a standard procedure. Assuming the pairs $(X_i, x_i)$ have some outliers, write in pseudo-code an algorithm of type RANSAC estimating $R$ and $T$.

## 2 Multiple view constraints with lines

We assume $n$ cameras in space, of calibration matrix $K_i$, oriented with rotation $R_i$ and translation $T_i$ with respect to some world coordinate frame.

1. Show that in general, the observation of a projection of a 3D line in a camera *does not* impose any constraint on its projection in another view.

2. What is the exception?

3. By a simple geometric consideration, show that the observation of the projections of a 3D line in *two* views determines the projection in a third view.

4. Let $l_i$ be the 3-D vectors representing the projections of a same 3D line in different views $i = 1, \dots, n$. Show that the $n \times 4$ matrix

$$N = \begin{pmatrix} l_1^T K_1 R_1 & l_1^T K_1 T_1 \\ \vdots & \vdots \\ l_n^T K_n R_n & l_n^T K_n T_n \end{pmatrix}$$

must have rank at most 2. Hint: consider the projections of two points on the 3D line.

5. Show that the $4 \times 5$ matrix $D$ below has rank 4.

$$D = \begin{pmatrix} l_1 & [l_1]_\times & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

6. We assume $R_1 = Id$ and $T_1 = 0$. Show that the matrix

$$\begin{pmatrix} l_1^T l_1 & 0 & 0 \\ l_2^T K_2 R_2 l_1 & l_2^T K_2 R_2 [l_1]_\times & l_2^T K_2 T_2 \\ \vdots & \vdots & \vdots \\ l_n^T K_n R_n l_1 & l_n^T K_n R_n [l_1]_\times & l_n^T K_n T_n \end{pmatrix}$$

has rank at most 2.

7. Deduce that the following $(n-1) \times 4$ matrix $M$ has rank at most 1.

$$M = \begin{pmatrix} l_2^T K_2 R_2 [l_1]_\times & l_2^T K_2 T_2 \\ \vdots & \vdots \\ l_n^T K_n R_n [l_1]_\times & l_n^T K_n T_n \end{pmatrix}.$$

8. Show that for any $i, j \in \{2, \dots, n\}$

$$l_i^T K_i R_i [l_1]_\times R_j^T K_j^T l_j = 0.$$
$$(l_j^T K_j T_j l_i^T K_i R_i - l_i^T K_i T_i l_j^T K_j R_j)[l_1]_\times = 0.$$

9. Show that any algebraic constraints between $m > 3$ projections $l_i$ are combinations of trilinear constraints involving 3 projections.

# 3 Feature detection and matching

1. Does the Harris Corner detector detect the same points in an image and its rotated version? Discuss and explain why.

2. Imagine you have a feature descriptor sensitive to rotations. Describe a simple strategy to build a rotation invariant descriptor.

3. One possible issue when trying to match images are repetitive elements, that can be found at multiple locations in the image and generate ambiguous matches. Describe a simple strategy to remove these ambiguous matches.

# 4 Graph cuts for disparity map estimation

We consider the estimation of disparity maps using graph cuts, based on an energy with the following form:

$$E(f) = \sum_{p \in \mathcal{P}} D_p(f_p) + \sum_{(p,q) \in \mathcal{N}} V_{p,q}(f_p, f_q)$$

where

- $f : \mathcal{P} \to \mathcal{L}$ is a labeling assigning a disparity $d \in \mathcal{L} = \{d_{\min}, \ldots, d_{\max}\}$ to each pixel $p \in \mathcal{P}$,
- $D_p : \mathcal{L} \to \mathbb{R}$ is a data term expressing the penalty for pixel $p$ to be assigned disparity $f_p$, and
- $V_{p,q} : \mathcal{L} \times \mathcal{L} \to \mathbb{R}$ is a regularization term expressing the penalty for neighboring pixels $p, q$ according to $\mathcal{N}$ to have potentially different disparities $f_p, f_q$.

We want to study the rotation sensitivity of the regularization term.

**Question 4.1.** We consider the case where $V_{p,q} = \lambda_{p,q} \min(K, |f_p - f_q|)$. Can $E(f)$ be minimized efficiently and exactly? If so, with what method(s)? If not, what method(s) can you however suggest?

We now consider the scene in Figure 1 with two fronto-parallel planar objects: an axis-aligned square (i.e., whose sides are horizontal or vertical in the picture) and an identical square rotated 45° (diamond).

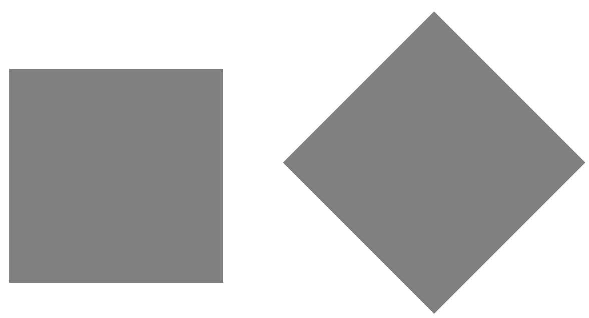

Figure 1: Axis-aligned square (left) and identifcal square rotated 45° (right).

We consider two types of pixel neighborhoods, pictured in Figure 2: $\mathcal{N}_4$ where a pixel has 4 neighbors and $\mathcal{N}_8$ where a pixel has 8 neighbors.

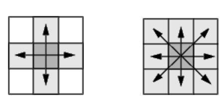

Figure 2: 4-pixel neighborhood (left) and 8-pixel neighborhood (right).

**Question 4.2.** We assume the difference of disparities is very large (larger than $K$) at the boundary between the background and the objects (square and diamond). Ignoring the situation near the vertices of the objects, i.e., looking only at their sides (edges) as shown in Figure 3, can a value be found for $(\lambda_{p,q})_{(p,q)\in\mathcal{N}_4}$ with a 4-pixel neighborhood $\mathcal{N}_4$ so that the penalty of the square boundary is identical to the penalty of the diamond boundary?

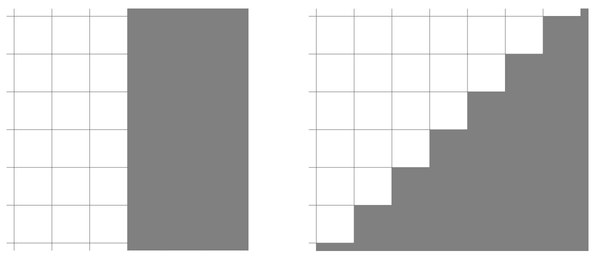

Figure 3: axis-aligned boundary (left) and diagonal boundary (right).

**Question 4.3.** Same question as Question 4.2 with neighborhood $\mathcal{N} = \mathcal{N}_8$.

**Question 4.4.** Are the answers to questions 4.2 and 4.3 different when the square is rotated with angle $\theta$ such that $\tan\theta = \frac{1}{2}$?

**Question 4.5.** Conclude. What is the impact in practice?