Homework 4: Feature Matching and Optical Flow

Instructor: René Vidal

Due 10/16/2014, 11.59PM Eastern

1. **(15 Points) Corner localization via quadratic fit.** The second step of SIFT is to fit a quadratic function to the response of the Difference of Gaussian (DoG) filter applied to the image around each local maximum. Specifically, if $r(\boldsymbol{x})$ is the response at pixel $\boldsymbol{x} = (x, y)$, we seek a quadratic function $\frac{1}{2}\boldsymbol{x}^\top Q\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c$ that approximates $r(\boldsymbol{x})$ in a neighborhood of $\boldsymbol{x}$. We can do this by minimizing the sum of the squares of the fitting errors
$$\min_{Q,\boldsymbol{b},c} \sum_{\boldsymbol{u}} w(\boldsymbol{x}+\boldsymbol{u})(\frac{1}{2}(\boldsymbol{x}+\boldsymbol{u})^\top Q(\boldsymbol{x}+\boldsymbol{u}) + \boldsymbol{b}^\top(\boldsymbol{x}+\boldsymbol{u}) + c - r(\boldsymbol{x}+\boldsymbol{u}))^2, \tag{1}$$
where $\boldsymbol{u} = (u, v)$ is the displacement vector in a window around $\boldsymbol{x}$ and w is a weighting function inside the window (e.g., a Gaussian). Propose a least-squares like algorithm based on the SVD for computing the parameters Q, $\boldsymbol{b}$ and c. Recall that Q is a 2×2 symmetric negative definite matrix (to get a maximum).

2. **(20 Points) Feature point matching under a 2D rigid body motion.** Let I_1 and I_2 be two images related by an unknown 2D rotation $R \in SO(2)$ and an unknown 2D translation $\boldsymbol{t} \in \mathbb{R}^2$, i.e., $I_2(\boldsymbol{x}) = I_1(R\boldsymbol{x} + \boldsymbol{t})$. Let $\{\boldsymbol{x}_j\}_{j=1}^N$ be a set of image points (e.g., corners) extracted from I_1. Suppose you have run a feature matching algorithm and extracted a set of corresponding image points $\{\boldsymbol{y}_j\}_{j=1}^N$ in I_2, i.e., $\boldsymbol{y}_j \approx R\boldsymbol{x}_j + \boldsymbol{t}$. Propose an algorithm for computing the unknown transformation $(R, \boldsymbol{t}) \in SE(2)$ that minimizes the sum of squared errors:
$$\min_{R,\boldsymbol{t}} \sum_{j=1}^N \|\boldsymbol{y}_j - R\boldsymbol{x}_j - \boldsymbol{t}\|_2^2. \tag{2}$$
Specifically, show that the optimal translation is given by $\boldsymbol{t}^* = \bar{\boldsymbol{y}} - R^*\bar{\boldsymbol{x}}$, where $\bar{\boldsymbol{x}} = \sum \boldsymbol{x}_i/N$ and $\bar{\boldsymbol{y}} = \sum \boldsymbol{y}_i/N$, and that the optimal rotation is given by $R^* = \operatorname{argmin}_{R\in SO(2)} \|Y - RX\|_F^2$, where $X = \begin{bmatrix}\boldsymbol{x}_1 - \bar{\boldsymbol{x}} \cdots \boldsymbol{x}_N - \bar{\boldsymbol{x}}\end{bmatrix}$ and $Y = \begin{bmatrix}\boldsymbol{y}_1 - \bar{\boldsymbol{y}} \cdots \boldsymbol{y}_N - \bar{\boldsymbol{y}}\end{bmatrix}$. Show that $R^* = \operatorname{argmax}_R \langle Y, RX\rangle = \operatorname{argmax}_R \operatorname{trace}(Y^\top RX)$. Parametrize R in terms of the rotation angle θ and show that
$$\theta^* = \operatorname*{argmax}_\theta \operatorname{trace}(X^\top Y)\cos(\theta) + \operatorname{trace}(X^\top \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} Y)\sin(\theta), \tag{3}$$
Find the optimal θ and show that the optimal R is given by
$$R^* = \frac{\begin{bmatrix} \operatorname{trace}(X^\top Y) & -\operatorname{trace}(X^\top \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} Y) \\ \operatorname{trace}(X^\top \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} Y) & \operatorname{trace}(X^\top Y) \end{bmatrix}}{\sqrt{\operatorname{trace}(X^\top Y)^2 + \operatorname{trace}(X^\top \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} Y)^2}}. \tag{4}$$

3. **(15 Points) Optical flow with changes in illumination.** Let $I(x, y, t)$ be a video sequence taken by a moving camera observing a rigid, static and Lambertian scene. Assume that between two consecutive views there is an affine change in the image intensities, i.e., the brightness constancy constraint reads
$$I(x+u, y+v, t+1) = aI(x, y, t) + b, \tag{5}$$
where $u(x, y)$ and $v(x, y)$ are the optical flow and $a(x, y)$ and $b(x, y)$ represent photometric parameters. Propose a linear algorithm for estimating (u, v, a, b) from the image brightness I and its spatio-temporal derivatives I_x, I_y, I_t. What is the minimum size of a window around each pixel that allows one to solve the problem?

Submission instructions. Send email to vision14jhu@gmail.com with subject 600.461/600.661:HW4 and attachment `firstname-lastname-hw4-vision14.zip` or `firstname-lastname-hw4-vision14.tar.gz`. The attachment should have the following content:

1. A file called `hw4.pdf` containing your answers to each one of the analytical questions. If at all possible, you should generate this file using the latex template hw1-vision14.tex. If not possible, you may use another editor, or scan your handwritten solutions. But note that you must submit a single PDF file with all your answers.

2. For coding questions, submit a file called `README`, which contains instructions on how to run your code. Use separate directories for each coding problem. Each directory should contain all the functions and scripts you are asked to write in separate files. For example, for HW2 the structure of what you should submit could look like

   (a) `README`

   (b) `hw2.pdf`

   (c) `hw2q3`: `hw2q3c.m`, `hw2q3e.m`

   (d) `hw2q4`: `hw2q4b.m`, `hw2q4c.m`

   The TA will run your scripts to generate the results. Thus, your script should include all needed plotting commands so that figures pop up automatically. Please make sure that the figure numbers match those you describe in `hw2.pdf`. You do not need to submit input or output images. The output images should be automatically generated by your scripts so that the TA can see the results by just running the scripts. In writing your code, you should assume that the TA will place the input images in the directory that is relevant to the question solved by your script. Also, make sure to comment your code properly.