Computer Vision (600.461/600.661)
Exam 1

Instructor: René Vidal

October 14, 2014

Part I (20 points) Answer these questions in 1-4 lines.

1. **(1 point)** Name two different color representations. Are they linearly related?
2. **(1 point)** Why is a Gaussian filter preferred to a box filter?
3. **(1 point)** What do you do to sharpen an image?
4. **(1 point)** What is the difference between the derivative of a Gaussian filter and the difference of Gaussians filter?
5. **(1 point)** What is template matching?
6. **(1 point)** What is a Gaussian pyramid? Name two applications of it.
7. **(2 points)** List the main 3-5 steps of the Canny edge detector.
8. **(1 point)** Name any method described in class that requires image interpolation.
9. **(2 points)** List the main 3-5 steps of RANSAC as applied to line fitting with outliers.
10. **(2 points)** List the main 3-5 steps of the Harris corner detector.
11. **(2 points)** How do you make a patch descriptor rotationally invariant?
12. **(1 point)** Name two criteria for deciding whether two feature descriptors match or not.
13. **(1 point)** How many point correspondences are needed to fit a 2D translational model?
14. **(1 point)** How many parameters (or degrees of freedom) are there in a 2D affine model?
15. **(1 point)** What is the aperture problem?
16. **(1 point)** What is the direction in the image along which optical flow cannot be reliably estimated?

Part II (30 points) Solve the following problems.

1. **(15 points)** Suppose you have a large collection of photos from your trips, including photos of yourself alone, photos of yourself with other people, photos of other people without you, as well as photos without people. Suppose you also have a template of your face, which consists of a small image of your face together with the 2D coordinates for the center of your eyes, tip of your nose, and ears. Describe a fully automatic algorithm that uses what you have learned in class to find the subset of the images that contains unoccluded frontal faces of you as well as the location, orientation and scale of a bounding box containing your face.

2. **(15 points) 3D affine registration.** Let $\{\boldsymbol{X}_i \in \mathbb{R}^3\}_{i=1}^N$ be a set of points in $\mathbb{R}^3$ that are transformed by a 3D affine transformation (A, T), where $A \in \mathbb{R}^{3\times 3}$ and $T \in \mathbb{R}^3$, to produce another set of points $\{\boldsymbol{Y}_i \in \mathbb{R}^3\}_{i=1}^N$. Suppose that the transformed points $\boldsymbol{Y}_i$ are corrupted by noise $\boldsymbol{E}_i$, i.e., $\boldsymbol{Y}_i = A\boldsymbol{X}_i + T + \boldsymbol{E}_i$ for all $i = 1, \ldots, N$. Show that the transformation (A, T) that minimizes the sum of the squared errors

$$E(A, T) = \sum_{i=1}^{N} \|\boldsymbol{Y}_i - A\boldsymbol{X}_i - T\|_2^2 \tag{1}$$

is given by $T^* = \bar{Y} - A^*\bar{X}$, $A^* = (YX^\top)(XX^\top)^{-1}$, where $\bar{X} = \sum \boldsymbol{X}_i/N$, $X = \left[\boldsymbol{X}_1 - \bar{X} \cdots \boldsymbol{X}_N - \bar{X}\right]$ and similarly for $\bar{Y}$ and Y. Show that 4 is the minimum number of points needed to find the transformation.