Problem Session #6

EE368/CS232

Digital Image Processing
1. Robustness of SIFT Keypoints to Rotation and Scaling

Part A:
Apply a SIFT keypoint detector and adjust the peak and edge thresholds so that about 450-850 SIFT keypoints are detected. Please submit a result showing the detected SIFT keypoints superimposed on the original image, and report the thresholds you choose. Describe which objects or regions in the image seem to generate large numbers of SIFT keypoints.

```matlab
% Set up MATLAB interface to VLFeat library
run('/path/to/library/vlfeat-0.9.20/toolbox/vl_setup.m');

% Extract SIFT keypoints on grayscale image
peakThresh = 4;
edgeThresh = 10;
[fc, dc] = vl_sift(single(gray_img), ...
  'PeakThresh', peakThresh, ...
  'EdgeThresh', edgeThresh);
```
We run the detector with peak threshold $\texttt{peak\_thresh}$ by

$$f = \texttt{vl\_sift}(I, \text{'PeakThresh'}, \text{peak\_thresh}) ;$$

obtaining fewer features as $\texttt{peak\_thresh}$ is increased.

Detected frames for increasing peak threshold.  
From top: $\texttt{peak\_thresh} = \{0, 10, 20, 30\}$.

The edge threshold eliminates peaks of the DoG scale space whose curvature is too small (such peaks yield badly localized frames). For instance, consider the test image

$$I = \texttt{zeros}(100,500) ;$$
$$\texttt{for } i=[10 \quad 20 \quad 30 \quad 40 \quad 50 \quad 60 \quad 70 \quad 80 \quad 90]$$
$$I(50-\text{round}(i/3):50+\text{round}(i/3),i*5) = 1 ;$$
$$\texttt{end}$$
$$I = 2\pi \times 8^{2} \times \texttt{vl\_immooth}(I,8) ;$$
$$I = \texttt{single}(255 \times I) ;$$

A test image for the edge threshold parameter.

We run the detector with edge threshold $\texttt{edge\_thresh}$ by

$$f = \texttt{vl\_sift}(I, \text{'edgethresh'}, \text{edge\_thresh}) ;$$
Eliminating low contrast responses

Peaks which are too short may have been generated by noise and are discarded. This is done by comparing the absolute value of the DoG scale space at the peak with the peak threshold $t_p$ and discarding the peak its value is below the threshold.

Eliminating edge responses

Peaks which are too flat are often generated by edges and do not yield stable features. These peaks are detected and removed as follows. Given a peak $x, y, \sigma$, the algorithm evaluates the $x, y$ Hessian of the DoG scale space at the scale $\sigma$. Then the following score (similar to the Harris function) is computed:

$$
\frac{(\text{tr} D(x, y, \sigma))^2}{\det D(x, y, \sigma)}, \quad D = \begin{bmatrix}
\frac{\partial^2 \text{DoG}}{\partial x^2} & \frac{\partial^2 \text{DoG}}{\partial x \partial y} \\
\frac{\partial^2 \text{DoG}}{\partial x \partial y} & \frac{\partial^2 \text{DoG}}{\partial y^2}
\end{bmatrix}.
$$

This score has a minimum (equal to 4) when both eigenvalues of the Jacobian are equal (curved peak) and increases as one of the eigenvalues grows and the other stays small. Peaks are retained if the score is below the quantity $(t_e + 1)(t_e + 1)/t_e$, where $t_e$ is the edge threshold. Notice that this quantity has a minimum equal to 4 when $t_e = 1$ and grows thereafter. Therefore the range of the edge threshold is $[1, \infty)$.
Part B:
Using the procedure defined in Problem 4 of HW5, plot repeatability versus rotation angle (in increments of 15 degrees, from 0 degrees to 360 degrees). Comment on the SIFT keypoint detector’s robustness against rotation and compare against the corresponding result for the Harris keypoint detector.
SIFT Keypoints

Harris Keypoints
Part C:
Using the same procedure defined in Problem 4 of HW5, plot repeatability versus scaling factor (with the scaling factors $m^0, m^1, m^2, \ldots, m^8$, where $m = 1.2$). Comment on the SIFT keypoint detector’s robustness against scale changes and compare against the corresponding result for the Harris keypoint detector.
SIFT Keypoints

Harris Keypoints

Robustness to Scaling: Cover-1.jpg

Robustness to Scaling: Cover-2.jpg
2. Recognition of Posters with Local Image Features

When you visit a poster at a conference/meeting, it would be useful to be able to snap a picture of the poster and automatically retrieve the authors’ contact information and the corresponding publication for later review. Please download the images `hw6_poster_1.jpg`, `hw6_poster_2.jpg`, and `hw6_poster_3.jpg` from the handouts webpage. These query images show 3 different posters during a previous EE368/CS232 poster session. Also download `hw6_poster_database.zip`, which contains clean database images of all posters shown during that poster session.
For each query image, use the following algorithm to match to the best database image:

• Extract SIFT features from the query image using vl_sift in the VLFeat library.

• Match the query image’s SIFT features to every database image’s SIFT features using nearest-neighbor search with a distance ratio test as implemented in vl_ubcmatch.

• From the feature correspondences that pass the distance ratio test, find the inliers using RANSAC with a homography as the geometric mapping.

• Report the database image with the largest number of inliers after RANSAC as the best matching database image.
Please submit the following results for each query image:

• A side-by-side view of the query image and the best matching database image.

• A side-by-side view of the query image and best matching database image, with SIFT keypoints overlaid on each image.

• A side-by-side view of the query image and the best matching database image with feature correspondences after the distance ratio test overlaid and connected by lines.

• A side-by-side view of the query image and the best matching database image with feature correspondences after RANSAC overlaid and connected by lines.
SIFT_MATCH Match two images using SIFT and RANSAC

SIFT_MATCH demonstrates matching two images based on SIFT features and RANSAC.

SIFT_MATCH by itself runs the algorithm on two standard test images. Use `SIFT_MATCH(IM1,IM2)` to compute the matches of two custom images IM1 and IM2.

SIFT_MATCH can also run on two pre-computed sets of features. Use `SIFT_MATCH(IM1, IM2, FEAT1, FEAT2)`, where FEAT1.f and FEAT1.d represent the SIFT frames and descriptors of the first image.

SIFT_MATCH returns MATCHRESULT, where MATCHRESULT.RATIO_TEST reports the number of correspondences after the distance ratio test, MATCHRESULT.RANSAC reports the number of correspondences after the distance ratio test + RANSAC with a homography, and MATCHRESULT.MODEL contains the best homography found.
3. Scale Selection for Determinant-of-Hessian

Consider a continuous-space version of the determinant-of-Hessian (DoH) keypoint detector. For an input image, the DoH response at scale $t$ is defined as follows:

$$\det H^t(x, y) = f_{xx}^t(x, y) f_{yy}^t(x, y) - \left( f_{xy}^t(x, y) \right)^2$$

$$f_{xx}^t(x, y) = \frac{\partial^2}{\partial x^2} g^t(x, y) * f(x, y)$$

$$f_{yy}^t(x, y) = \frac{\partial^2}{\partial y^2} g^t(x, y) * f(x, y)$$

$$f_{xy}^t(x, y) = \frac{\partial^2}{\partial y \partial x} g^t(x, y) * f(x, y)$$

$$g^t(x, y) = \frac{1}{2\pi t} \exp \left( -\frac{x^2 + y^2}{2t} \right)$$
Part A:
Assume the input image is a Gaussian blob of scale \( t_0 > 0 \): \( f(x, y) = 2\pi t_0 \cdot g^{t_0}(x, y) \). From the above definitions, derive and report an expression for the following objective functions:

\[
H_A(t) = \det H^t(0, 0)
\]

\[
H_B(t) = t \det H^t(0, 0)
\]

\[
H_C(t) = t^2 \det H^t(0, 0)
\]

\[
g^t(x, y) \ast f(x, y) = 2\pi t_0 \cdot g^t(x, y) \ast g^{t_0}(x, y) = 2\pi t_0 \cdot g^{t+t_0}(x, y)
\]

\[
f^t_{xx}(x, y) = \frac{\partial^2}{\partial x^2}\left(g^t(x, y) \ast f(x, y)\right) = 2\pi t_0 \frac{\partial^2}{\partial x^2} g^{t+t_0}(x, y)
\]

\[
f^t_{yy}(x, y) = 2\pi t_0 \frac{\partial^2}{\partial y^2} g^{t+t_0}(x, y)
\]

\[
f^t_{xy}(x, y) = f^t_{yx}(x, y) = 2\pi t_0 \frac{\partial^2}{\partial x \partial y} g^{t+t_0}(x, y)
\]

\[
\det H^t(0, 0) = f^t_{xx}(0, 0) f^t_{yy}(0, 0) - \left(f^t_{xy}(0, 0)\right)^2
\]
Part B:
Derive and report the scale values $t_A^*$, $t_B^*$, and $t_C^*$ that maximize $H_A(t)$, $H_B(t)$, and $H_C(t)$, respectively. Also report the maximal values. Please interpret the results.

$t_A^* = \ldots$
$t_B^* = \ldots$
$t_C^* = \ldots$

Are $H_A$, $H_B$, and $H_c$ suitable for finding the scale $t_0$?

$H_A(t_A^*) = \ldots$
$H_B(t_B^*) = \ldots$
$H_C(t_C^*) = \ldots$

Do maximum values of $H_A$, $H_B$, and $H_c$ depend on the scale $t_0$?