SCC.366 Media Coding and Processing – CW2 – Report

Name: **Matthew Kenely** Student ID: **38834561**

Abstract

This project compares the performance of Raw Pixel Values and Histogram of Oriented Gradients (HoG) features for image matching, as well as testing the robustness of the Harris interest point detector to intensity shift, intensity scale and image translation. A mix of experimental and theoretical results are presented. Experiments were carried out using MATLAB. Image matching using HoG features was found to be more accurate than using Raw Pixel Values. The Harris interest point detector was found not to be robust to intensity shift, intensity scale or image translation.

1 Raw Pixel Values vs. HoG features

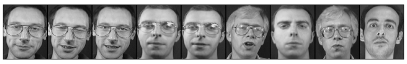

Figure 1: Image matching results using Raw Pixel Values

Figure 2: Image matching results using HoG Features

Taking accuracy as:

$$\frac{\text{Number of correct matches}}{\text{Total number of matches}}$$

Image matching using Raw Pixel Values has an accuracy of $\frac{3}{9} = 33.\dot{3}\%$, whereas image matching using HoG Features has an accuracy of $\frac{4}{5} = 80\%$. Image matching using HoG Features worked better than image matching using Raw Pixel Values in this experiment, which is to be expected as pixel values are susceptible to intensity shifts, changes in illumination/colour and noise. HoG Features mitigate these issues by extracting more generalised feature descriptors of the image which contain useful information (specifically, histograms of gradient directions of sections of the image) and ignoring extraneous information [1].

2 Is Harris interest point detector robust to intensity shift?

2.1 Experiments

Figure 3: Harris interest point detector applied to original image [686 corners detected], original image with all pixel values shifted by +1 [686 corners detected] and original image with all pixel values shifted by -20 [639 corners detected]

According to the above observations, **the Harris interest point detector is not robust to intensity shift as if the magnitude of the intensity shift is large enough to cause values to clip (values == 0 highlighted in blue) the corners detected may be affected.**

2.2 Theory

$$E(u,v) = \sum_{(x,y)\in W} [I(x+u, y+v) - I(x,y)]^2$$

Error score for pixel $p \in I$ is calculated by summing up the squared differences between the pixel values in a window W centred about p and the corresponding pixels in a window shifted by (u,v) [2].

Assuming a small window motion, the error score can be approximated by using the Taylor Series expansion:

$$E(u,v) \approx \sum_{(x,y)\in W} \left[\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v\right]^2$$

Considering $\frac{\partial I}{\partial x}$ as I', pixel $p \in I$ and the coordinates of its left and right neighbouring pixels, (l_x, l_y) and (r_x, r_y):

The partial derivative of I wrt. x at pixel p is calculated by taking the difference between the pixel values of I at (r_x, r_y) and (l_x, l_y). i.e. $I'(p) = I(r_x, r_y) - I(l_x, l_y)$. Taking the pixel values of $I(r_x, r_y)$ and $I(l_x, l_y)$ as r and l respectively, the partial derivative of I wrt. x at pixel p can be written as $I'(p) = r - l$. Assuming an intensity shift of s is applied to the entire image, in general, $r - l = (r+s) - (l+s)$, i.e. $\frac{\partial I}{\partial x}$ would be unaffected by the intensity shift. However, if the intensity shift causes a pixel value to clip, assuming an 8-bit image representation, one or both of $r+s$ or $l+s$ may be equal to 0 or 255, i.e.:

Clip 0	**Clip 255**
$r - l = 0 - (l+s)$	$r - l = 255 - (l+s)$
$r - l = (r+s) - 0$	$r - l = (r+s) - 255$
$r - l = 0 - 0$	$r - l = 255 - 255$

Simplifying:

$I'(p) = -l - s$	$I'(p) = 255 - l - s$
$I'(p) = r + s$	$I'(p) = r + s - 255$
$I'(p) = 0$	$I'(p) = 0$

Therefore, if the intensity shift s causes one or more of the values of the neighbouring pixels of p to clip, the partial derivative of I wrt. x at pixel p would be either 0 or subject to the variable s and is no longer relative to the difference between r and l, thus the error score produced in the original image will not necessarily be the same as the error score produced in the image with the intensity shift applied due to the change in value of $\frac{\partial I}{\partial x}$ at p.

The above also applies to $\frac{\partial I}{\partial y}$ taking neighbouring pixels above and below p.

Therefore, **the Harris interest point detector is not robust to intensity shift as if the magnitude of the intensity shift is large enough to cause values to clip the corners detected may be affected.**

3 Is Harris interest point detector robust to intensity scale?

3.1 Experiments

Figure 4: Harris interest point detector applied to original image [686 corners detected], original image with all pixel values scaled by 1.5 [590 corners detected] and original image with all pixel values scaled by 0.85 [683 corners detected]

According to the above observations, **the Harris interest point detector is not robust to intensity scale regardless of the magnitude of scaling.**

3.2 Theory

Continuing on 2.2:

Assuming an intensity scale of s is applied to the entire image:

$$\begin{aligned} r - l &= sr - sl \\ r - l &= s(r - l) \\ s &= \frac{r - l}{r - l} \\ s &= 1 \end{aligned}$$

Therefore, the partial derivative of I wrt. x at pixel p would be unaffected only in the case that $s = 1$, i.e. the pixel values of the image were not scaled at all. Thus the error score produced in the original image will not be the same as the error score produced in the image with the intensity scaling applied due to the change in value of $\frac{\partial I}{\partial x}$ at p.

The above also applies to $\frac{\partial I}{\partial y}$ taking neighbouring pixels above and below p.

Therefore, **the Harris interest point detector is not robust to intensity scale regardless of the magnitude of scaling.**

4 Is Harris interest point detector robust to image translation?

4.1 Experiments

Figure 5: Harris interest point detector applied to original image [686 corners detected], original image with all pixel values translated by (0, -100) [686 corners detected] and original image with all pixel values translated by (200, 0) [689 corners detected]

According to the above observations, **the Harris interest point detector is not robust to image translation as the translation may displace the image in such a way that new corners are detected due to the zero padding introduced.**

4.2 Theory

Continuing on 2.2:

Given $x \in Z^+$, assuming an image translation of $(x, 0)$ is applied to the entire image, for all pixels p in image column $x + 1$ (i.e. the column to the right of the zero padding introduced by the translation):

$$r - l = r - 0$$

Simplifying:

$$I'(p) = r$$

Given $x \in Z^-$, assuming an image translation of $(x, 0)$ is applied to the entire image, for all pixels p in image column $(\text{size}\,(I,2)^* - |x| - 1)$ (i.e. the column to the left of the zero padding introduced by the translation):

$$r - l = 0 - l$$

Simplifying:

$$I'(p) = -l$$

Therefore, at the border of the translation applied to the original image, the partial derivative of I wrt. x at pixel p is subject to either the variable r or l and is no longer relative to the difference between r and l, thus the error score produced in the original image will not necessarily be the same as the error score produced in the image with the translation applied due to the change in value of $\frac{\partial I}{\partial x}$ at p.

The above also applies to $\frac{\partial I}{\partial y}$ taking $y \in Z^+$ and $y \in Z^-$ respectively, a translation of $(0, y)$, and all pixels p at image row y + 1 (i.e. the row below the zero padding introduced by the translation) and $(\text{size}\,(I,1)^{**} - |y| - 1)$ (i.e. the row above the zero padding introduced by the translation) respectively.

Therefore, **the Harris interest point detector is not robust to image translation as the translation may displace the image in such a way that new corners are detected due to the zero padding introduced.**

*No. of columns in the **translated** image.

No. of rows in the **translated image.

References

[1] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," in *International workshop on automatic face and gesture recognition*, vol. 12. Zurich, Switzerland, 1995.

[2] C. G. Harris and M. J. Stephens, "A combined corner and edge detector," in *Alvey Vision Conference*, 1988.