Machine Perception Assignment 1

Reece Jones   
Student ID: 19476111

# Task 1

## Harris Corner Detector

The Harris corner detector [1] is variant under scaling and invariant under rotation. Scale invariance is not achieved since the size or standard deviation of the window function is a fixed parameter of the algorithm; with images of different scale, the window will cover a different size image patch. I expect that the higher the ratio of window size to feature size, the lower the corner response, since the window will include a higher proportion of non-corner pixels. The experimental results are shown in Fig. 1. Interestingly, as the image size increases, the number of corners detected decreases, contrary to my hypothesis. I believe that in the process of resampling the images, the increase in the size of the edges has reduced the gradient, reducing the corner response. This effect is likely dependent on the resampling method used (preliminary testing with nearest-neighbour resampling indicates different results).

Rotational invariance is achieved because the corner response is calculated based on the eigenvalues of the autocorrelation matrix at a point [2]. These eigenvalues give the magnitudes of curvature of the image surface in the directions of most and least curvature [1]. That is, any rotation present has no effect on the eigenvalues, and therefore no effect on the corner response. The experimental results are shown in Fig. 2; as expected, the same corners are detected regardless of rotation.

The experimental results for the dugong image are similar and may be reproduced using the code in Appendix 1.

## Intensity Histograms

Image histograms are variant under scaling and may be variant under rotation (depending on the rotation). Scaling an image changes its resolution, and since histograms are based on pixel counts, the histogram will be affected. However, for small or moderate changes in scale, the relative ratios of intensities will be retained, i.e. the histogram will have approximately the same “shape” (if normalised histograms are used, then scale invariance is achieved). This is demonstrated in Fig. 3. Notice that the histogram peaks are vastly different, but the overall shape is almost identical.

Histograms may be invariant under rotation, depending on whether the rotation affects the image bounds. Typically, histograms are calculated across an image or bounding box, the content of which will change under most rotations (consider the 45-degree rotation of an entire image – some regions are cut off, and empty regions are introduced). If we apply a rotation but adjust the bounds within which we calculate the histogram accordingly, then the histogram will be approximately unchanged. The experimental results are shown in Fig. 4. A fixed bounding box was selected, with the image rotates with respect to it. The image content inside the box did change, but since the background (ocean) is mostly homogeneous, the histograms are similar.

The experimental results for the playing card image are similar and may be reproduced using the code in Appendix 1.

|  |  |  |
| --- | --- | --- |
| **Rotation** | **Keypoints** | **Keypoints common with 0°** |
| 0° | 16 | 16 |
| 15° | 16 | 12 |
| 45° | 16 | 11 |
| 75° | 15 | 10 |

|  |  |  |
| --- | --- | --- |
| **Scale** | **Keypoints** | **Keypoints common with 1x** |
| 1x | 16 | 16 |
| 1.25x | 25 | 13 |
| 1.5x | 32 | 16 |
| 2x | 45 | 16 |

## Scale-invariant Feature Transform (SIFT) Keypoints

SIFT [3, 4] keypoints are moderately invariant under small to moderate scaling and rotation. Scale invariance is achieved by examining the image at many scales (via Gaussian filtering and image downsampling), transforming the image into “scale space”. Keypoints are detected within this scale space, yielding not only the keypoint’s location in the image, but also the scale at which the keypoint exists. Further processing of the keypoint is adjusted for this scale, producing keypoints which are largely invariant to scale. The experimental results are summarised in Fig. 5. As the image size increases, the number of keypoints detected increases (presumably since there are simply more pixels in the image), but a significant proportion of them match the keypoints in the original image region. The match proportions upwards of 80% for small scaling factors coincides with the results presented in [3].

Rotational invariance is (theoretically) achieved since keypoint detection only considers the magnitude of the difference of Gaussians (which is isotropic) [3, 4]. The experimental results are summarised in Fig. 6. Surprisingly, only about 65-75% of keypoints from the original image region are present in the rotated versions. My only explanation as to why such a large proportion of keypoints did not match is that it is as a result of the image resampling during rotation, especially given that the image region is small, however, I am unsure. It seems plausible that a more complex explanation exists, due to the many steps in the SIFT algorithm.

The keypoint-annotated images are omitted for brevity but may be reproduced using the code in Appendix 1. The experimental results for the playing card image are similar and may be reproduced using the same code.

# Task 2

## Local Binary Patterns (LBP) Features

LBP [5] generates an 8-bit integer feature for every pixel in a greyscale image (or image region). The feature is determined from the 8 pixels neighbouring the target location. These neighbouring pixels are traversed in a fixed order and compared to the target pixel. If the neighbouring pixel is less than the target pixel, a 1 is recorded; if the neighbouring pixel is greater than the target pixel, a 0 is recorded. The 1s and 0s are concatenated to form an 8-bit value, which is the feature descriptor for the target location. Further processing may involve dividing the image into cells, within which histograms of the feature values are calculated.

The main advantage of LBP is that it is quite effective compared to other texture descriptors, as found by [5] and [6]. Contributing to its effectiveness is its invariance under global changes in illumination. The algorithm may be extended to consider neighbouring pixels at any radius, allowing adjustment of the locality of detail that is extracted [7]. Finally, the algorithm is simple and computationally efficient.

The disadvantages of LBP are that it is not (without extensions) scale nor rotationally invariant and does not consider colour information.

## Scale-invariant Feature Transform (SIFT) Features

SIFT [3, 4] generates feature descriptors from the local neighbourhoods around the detected keypoints. First, the 16x16 region of pixels centred at the keypoint is considered and divided into 16 4x4 pixel blocks. For each block, an 8-bin histogram of gradient orientations is created. The orientation of each pixel is calculated at the scale of the detected keypoint, achieving scale invariance, and is given relative to the orientation of the keypoint, achieving rotational invariance. Each pixel’s histogram vote is weighted by the magnitude of its gradient, and by a Gaussian window centred at the keypoint. The 16 histograms are concatenated into a single 128-element vector to form the feature descriptor. Finally, this vector is thresholded and normalised to achieve some invariance to illumination changes.

The main advantages of SIFT features is, of course, that they are invariant under scaling, rotation, and illumination changes. Another advantage, demonstrated in [4], is that the high distinctiveness of the descriptors enables reliable feature matching even within very large databases.

The disadvantages of SIFT features are that the algorithm is quite complex, and colour information is not considered. [4] also notes that the high dimensionality of the feature descriptors inhibited feature matching efficiency when used with nearest-neighbour classification.

## Histogram of Oriented Gradients (HOG) Features

HOG [8] generates a feature descriptor for an image (or image region), derived from 8x8 pixel cells. A gradient orientation histogram is created for each cell. The histogram uses 9 bins for 0° to 180°, with orientations taken modulo 180° (this was found to be most effective for human detection). Each pixel’s vote is linearly interpolated between the two nearest histogram bins and is additionally weighted by the magnitude of the gradient. The cells are then grouped into overlapping 2x2 cell (16x16 pixel) blocks and normalised, to account for local illumination changes. Finally, the blocks are flattened and concatenated into a single vector to form the feature descriptor.

The primary advantage of HOG features is their excellent performance in the detection of humans (significantly better than other techniques), in part due to their illumination invariance [8].

The disadvantages of HOG features are that they are not scale nor rotationally invariant. Another possible disadvantage is the high dimensionality of the descriptor, which may be a challenge for classification.

## Comparison of SIFT and HOG Descriptors

Fig. 7 and 8 compare the similarity of SIFT and HOG feature descriptors taken from a region of the dugong image, across different scales and rotations. The descriptors are taken at SIFT keypoints that match in location and orientation across all versions of the image. For each transformed version of the image, the Euclidean distance is calculated between its matrix of descriptors and the matrix of descriptors from the original, untransformed image.

As expected, the HOG descriptors exhibit less similarity as the degree of transformation increases, since HOG is variant under scaling and rotation. The SIFT descriptors also exhibit some dissimilarity; however, to a lesser amount. Critically, while the HOG descriptor similarity is strictly decreasing, the SIFT descriptor similarity is not particularly correlated to the amount of transformation. This result highlights the scale and rotational invariance of SIFT.

Interestingly, both SIFT and HOG descriptors exhibit a large jump in dissimilarity between the untransformed image versions and the first transformed version. From a theoretical standpoint, I would expect the smallest amount of transformation to produce the smallest descriptor difference. Additionally, the SIFT descriptor difference decreases from scales of 1.6x to 2x – I would expect that it stays the same or increases slightly. I believe these observations are an effect of the image resampling during scaling/rotation. Preliminary testing with the application of a transformation followed by the inverse transformation – theoretically producing no change – shows similar patterns. Further investigation may be required to determine the effects of resampling on the results presented here.

# Prepare Your Paper Before Styling

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

## Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

## Units

* Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
* Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
* Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter”, not “webers/m2”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
* Use a zero before decimal points: “0.25”, not “.25”. Use “cm3”, not “cc”. (*bullet list*)

## Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively”.
* In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
* Do not confuse “imply” and “infer”.
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”.
* The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

# Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Authors and Affiliations

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**

##### References

1. C. Harris and M. Stephens, “A combined corner and edge detector,” *Alvey Vision Conference*, vol. 15, no. 50, pp. 147-151, 1988.
2. R. Szeliski, *Computer Vision: Algorithms and Applications*, 2nd ed. Accessed Sep. 14, 2020. [Online]. Available: http://szeliski.org/Book.
3. D. G. Lowe, “Object recognition from local scale-invariant features,” in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, pp. 1150-1157 vol.2, doi: 10.1109/ICCV.1999.790410.
4. D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.
5. T. Ojala, M. Pietikäinen and D. Harwood, “A comparative study of texture measures with classification based on feature distributions,” *Pattern Recognition*, vol. 29, no. 1, pp 51-59, Jan. 1996, doi: 10.1016/0031-3203(95)00067-4.
6. T. Ahonen, A. Hadid and M. Pietikäinen, “Face description with local binary patterns: Application to face recognition,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, Dec. 2006, doi: 10.1109/TPAMI.2006.244.
7. T. Ojala, M. Pietikäinen and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, July 2002, doi: 10.1109/TPAMI.2002.1017623.
8. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.