**Makerere University**

School of Computing and Informatics Technology

Department of Computer Science

MCS 7224: Computer Vision Project Examination

Name: Masongo Umar

Student No: 1800738510

Registration No: 2018/HD05/1967U

**Abstract**

This project explores the techniques of creating local feature matching algorithms and object detection algorithms which are fundamental concepts in computer vision. First, a SIFT descriptor pipeline is used to create a local feature instant-level matching algorithm for sugarcane images extracted from a video file containing crops in a garden. Brute-Force and FLANN based algorithms are then used to match the two objects. Secondly, an object detection pipeline based on Histogram of Oriented Gradient (HOG) and a Linear Support Vector Machines (SVM) baseline classifier is used to create and train an algorithm that detects cassava images extracted from the same video file.

**Key Words**: SIFT, Brute-Force Matcher, FLANN Based Matcher, HOG, SVM, Feature

Detection, Object Detection

1. Introduction

Image matching is an important concept in computer vision and object recognition because different images of the same object can be taken from different angles, with different lightings, and different scales [3]. These factors coupled with occlusion may cause challenges in recognizing similar images. Scale-Invariant Feature Transform (SIFT) is a kay-point detection technique that helps solve some of these challenges. SIFT is a rotation and scale invariant pipeline used to find unique key points, or the locations of the most distinctive features, on each image in order to match different images from the same object [1]. Content is then normalized around the key-points and the local feature descriptor is computed. Object detection is computer vision is the process by which an algorithm generates bounding boxes around the object of interest to determine its location in an image and categorise it [5][6].

Object detection can practically be applied in autonomous driving, surveillance, human behaviour analysis, and video/image indexing among others. Histogram of Oriented Gradient (HOG) is an object detection pipeline in which a detector window is tiled with a grid of overlapping blocks in which Histogram of Oriented Gradient feature vectors are extracted [1]. The combined vectors are fed to a linear SVM for object or non-object classification. The detection window is scanned across the image at all positions and scales, and conventional non-maximum suppression is run on the output pyramid to detect object instances.

The rest of this paper is organized as follow. In Section 2, the methods and materials used for

experimenting SIFT and HOG+SVM are discussed. In Section 3, the results obtained from running the SIFT and HOG+SVM algorithms are presented. recommendations are made in Section 4.

2. Literature Review

Y. Meng, and B. Tiddeman’s research focuses on deriving SIFT features from an image and using the features to perform face identification [3]. The SIFT features are computed at the edges and they are invariant to image scaling, rotation, addition of noise. The major issue with this research is that edges are poorly defined and hard to detect. At times, the images are too smooth to find that many features for a matching and a small face could be unrecognized from the training images.

E. Karami, M. Shehata, A. Smith evaluate the performance of the SIFT detection pipeline against image distortions parameter such as rotation, scaling, fisheye, and motion distortion [4]. The true and false positive rates for many image pairs are calculated and presented. The distribution of the matched key is also evaluated. The challenge with this work is that it does not use the results obtained to optimize the SIFT matching accuracy.

N. Dalal, B. Triggs show that the use of locally normalized HOG descriptors provides better performance compared to other existing feature sets including wavelets [2]. The HOG descriptors are reminiscent of edge orientation histograms, SIFT descriptors, and shape contexts, but they are computed on a dense grid of uniformly spaced cells, and they use overlapping local contrast normalizations for improved performance.

F. Suard, A. Rakotomamonjy, A. Broggi first introduce a single frame pedestrian detection system based on HOG+SVM and applied on infrared images. The initial system is based on the principals of N. Dalal, B. Triggs and detects small size images. The researchers then propose a complete detection system based on a focus of attention approach. This complete system is then able to detect any scale of pedestrians in a large size image.

3. Methods and Materials

**Datasets**. The two sugarcane images matched using the SIFT detection pipeline are obtained from the garden video provided in the question paper. These images are extracted from the same object but at a different scale and orientation. About 220 positive cassava images are obtained from Makerere University AI Lab and used to train the linear SVM classifier to detect cassava images. These images unlike those in the video are clean and free from forward and background noise which could affect the quality of the linear SVM classifier. About 195 negative images with a lot of noise in foreground and background are extracted from the video and used to train the SVM classifier to detect non cassava images. Image frames extracted from the video contain a mixture of both positive (cassava) images and negative (non-cassava) mages which are used to test the detection quality of the HOG detection pipeline. Both SIFT and HOG+SVM detection pipeline are executed using OpenCV 4.5.5 image processing library, Jupiter Notebook 4.8, and Python 3.9 all running in Anaconda Package Manager.

**SIFT Methodology**. Two sugarcane images with different scales and orientations are extracted from the same object in the garden video provided. SIFT pipeline first detects the key-points of each of the images using by maximising the Difference of Gaussians (DoG) in both the scale and the space to find same key points independently in each of the sugarcane image. SIFT pipeline then creates a local feature descriptor by using the Gaussian blurred image associated with the key point's scale, taking the image gradients over 16X16 array, rotating the gradient directions and locations relative to the key-point orientation, creating an array of orientation histograms for a 4X4 array within the 16X16 array, and adding the rotated gradients into their local orientation histograms with 8 orientation bins. SIFT pipeline finally calculates the Euclidian distance in order to compute the difference between two key-point descriptors. SIFT vectors are used to compare key points from the first sugarcane image to the sugarcane image B to find matching key-points by using Euclidean "distance" between descript or vectors. The matching is then done using the Brute-Force matcher and/or the Fast Library for Approximate Nearest Neighbours (FLANN) based matcher.

**HOG+SVM Methodology.** Histogram of Oriented Gradient (HOG) uses edges to describe the content of cassava and other crops images. HOG quantifies the differences in edge shapes to differentiate one crop from the other. The edges of the crops are described in terms of the gradient of each pixel in the crops. The Sobel operator is used to compute the gradient in both the x and y directions for each pixel. A crop image is divided into groups of pixels known as cells and a histogram of each cell is created by dividing the range of orientations into bins. Each of the pixels in a cell vote towards these bins based on its magnitude and orientation. The cells are then grouped into blocks and each block is normalized based on the magnitude of all the values in that block. The linear SVM classifier is first trained using HOG features extracted from 220 positive images of different pure cassava plants. The SVM is also trained using HOG features extracted from 195 negative images of a mixture of several crops except cassava. The SVM then computes an optimal hyperplane that separates cassava images (true positives) from non-cassava images (false positives). After learning, the SVM was tasked to detect cassava images in the unlabelled test data containing image frames extracted from the garden video.

4.Results

SIFT Pipeline. Two different sugarcane images from the same object go through the same

process of key-point detection and local feature description creation. The two images are then matched using the Brute-Force matcher and FLANN Based matcher algorithms.



Figure 1: SIFT Pipeline using Brute-Force Matcher

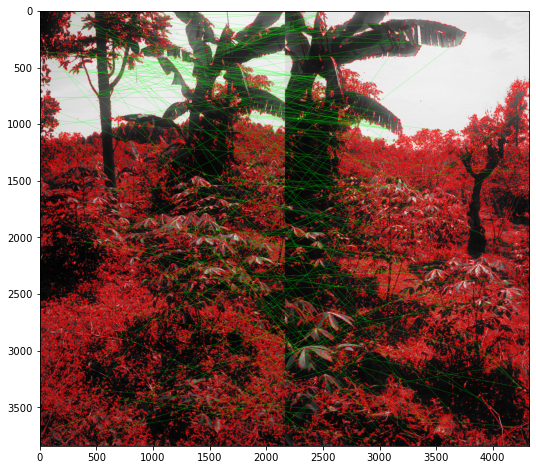


Figure 2: SIFT Pipeline using FLANN Based Matcher

Results in the Figure 1 show that Brute-Force matcher successfully matches most common

features in the upper parts of the two images and a few features in the middle and bottom of the images. Brute-Force matcher can either return the number best matches automatically detected or the number of matches specified by the user in the algorithm. Analysis of Figure 2 image shows that FLANN Based Matcher returns more optimized matches for fast nearest neighbour search and the matches are almost uniformly distributed in the two images. FLANN is faster and more accurate than BFMatcher for large datasets and high dimensional features like the sugarcane images in this case.

HOG+SVM. HOG features extracted from the and negative images were used to train the

SVM to be able to detect cassava images from a cluster of several different image frames from the garden video file. The statistics of the features are shown in the below:

On training the linear SVM with the positive and negative features, the following results were achieved.

5. Conclusion

This paper first covered the SIFT pipeline for local feature detection. The Brute-Force and

FLANN Based algorithms were used for matching the local features. The paper then looked at HOG object detector algorithm and linear SVM for training the object detector. Traditional feature and object detection pipelines are built on handcrafted features and shallow trainable architectures. The performance of these pipelines degrades with the construction of more complex algorithms which combine multiple low-level image features with high-level context from object detectors and scene classifiers. Advances in deep learning technologies have led to the development of more powerful algorithms such as YOLO, SDD, Faster-RCNN.

Since image frames of pure cassava images could not be extracted from the garden video, this paper relied on pure cassava images from Makerere AI Lab. However, these images were from plants infected with diseases which had damaged and changed the shapes of the cassava leaves. The paper therefore recommends the use of cassava images from healthy plants in order to achieve better results.

References

[1] D. Lowe, “Distinctive Image Features from Scale-Invariant Key-Points”, International

Journal of Computer Vision, Volume 60, No 2, Page 91-110, 2004.

[2] N. Dalal, B. Triggs, “Histogram of Oriented Gradient for Human Detection”

[3] Y. Meng, B. Tiddeman, “Implementing the Scale Invariant Feature Transform (SIFT)

Method”

[4] E. Karami, M. Shehata, A. Smith, “Image Identification Using SIFT Algorithm:

Performance Analysis against Different Image Deformations”

[5] N. Bodla, B. Singh, R. Chellappa, “Soft-NMS – Improving Object Detection with One Line

of Code”

[6] Z. Zhong-Qiu, P. Zheng, S. Xu, et al, “Object Detection with Deep Learning: A Review”,

IEEE Transactions on Neural Networks and Learning Systems

[7] F. Suard, A. Rakotomamonjy, A. Broggi, “Pedestrian Detection using Infrared images and Histograms of Oriented Gradients”, Intelligent Vehicles Symposium 2006