

Assignment-4

Complex Object Recognition

Jatin Bhikadiya
bhikadiy@cs.colostate.edu

May 8, 2013

Contents

1	Introduction	2
2	Objective	2
3	Techniques for feature extraction	2
3.1	Line Extraction based on Hough Transform	2
3.2	Grabcut	3
3.3	Circular Hough Transform	4
3.4	Scale-invariant feature transform	4
3.5	Histogram of Oriented gradients	5

1 Introduction

Object recognition is the task of finding the presence of a given object in an image. In previous assignment, different feature detectors were studied and analyzed the performance of each so that they can be used in this assignment to recognize the complex objects. Feature based Image matching techniques are most efficient way for image search. Next step is to classify the objects, each having some specific property or features, so that they can be separated from each other. One of the successful classification technique used in computer vision is Bag of Features. Bag of Features approach creates code-book containing code-words(clusters), which can be considered as representative of similar patches or features. Lastly, image histograms based on code-book can be trained using classifier for different object and prediction can be made on the test images.

2 Objective

The objective of this assignment is to recognition the complex objects from the given data-set. There are two data-sets given, one for training and other from testing, containing 400 images each. Every image contains one the for objects: airplane, automobile, leopards, zebra. All the objects can be any scale or pose or just part of the object is available in image. Training images can be analyzes in any ways but test images can not be analyzed until the evaluation of the system. To tackle the problem image features can be combined with classifier using bag of feature or constellation techniques. As the result confusion matrix, for all the 400 test images of four classes, should be generated showing true positive, true negative, false positive and false negative outputs.

3 Techniques for feature extraction

3.1 Line Extraction based on Hough Transform

Line detection is one of the most basic algorithms in computer vision technology. It has variety of applications like pattern recognition,object comparison,Lane detection,Simultaneous localization and mapping.

The idea involved in Hough Transform is that any point in a binary image could be part of some set of possible lines. Each line is parameterized by a slope 'a' and an intercept 'b'. Algorithm transforms image from image coordinates to parameter space. Firstly, it initialize table of buckets, indexed by a and b. Now for each detected edge pixel (x,y) it determines all (a,b) such that $y=ax+b$ and increments bucket (a,b). Lastly, buckets with many votes indicates probable lines.

Before applying Hough transform, Canny edge detection is used to get binary image. User need to provide number of intersection to consider the bucket as line. Here by varying the number of intersection to detect the line, the number of line detected can be varied. The effect of changing the number of intersections can be seen in figure ??.

Advantages:

- The advantage of the Hough transform is that the pixels lying on one line need not all be contiguous. This can be very useful when trying to detect lines with short breaks in them due to noise, or when objects are partially occluded.
- By detecting the lines, object with different shapes can be detected by finding the intersection of lines.

Disadvantages:

- It can give misleading results when objects happen to be aligned by chance.
- Another disadvantage is that the detected lines are infinite lines, rather than finite lines with defined end points.
- The Hough transform is not a fast algorithm for finding infinite lines in images of a certain size. Since additional analysis is required to detect finite lines, this is even slower.

There is a variant of standard hough transform called Probabilistic hough transform. It takes in the extent of the lines than just the orientation of the lines. The reason its called Probabilistic is that accumulates only a fraction of the points in the accumulator plane and not all of them which makes it faster than standard hough transform. Figure ?? shows the difference between standard and probabilistic hough transform.

3.2 Grabcut

GrabCut is an iterative image segmentation technique based upon the Graph Cut algorithm. Grabcut is just extended version of graphcut to incomplete trimaps and color images. In grabcut user is asked to draw a rectangle around the desire background. GrabCut Algorithm is as [3]

1. User selects a rectangle from the image and trimap is generated with that. Pixel outside rectangle are marked as background and inside the rectangle are marked as unknown, which creates initial segmentation. Gaussian Mixture Models (GMMs) are created for initial foreground and background classes.
2. Each pixel are assigned to most likely component in foreground or background GMM respectively.
3. The old GMMs are thrown away and new GMMs are learned from the pixel sets created in the previous set.
4. A graph is built and Graph Cut is run to find a new tentative foreground and background classification of pixels.
5. Steps 2-4 are repeated until the classification converges.

Advantages:

- Grabcut is more robust as it is iterative version of Graphcut and includes color information.
- For images where the foreground and background are cleanly separated, GrabCut can robustly segment the image given only a rectangular region as input. It can be seen in figure ??.

Disadvantages:

- Significant amount of user interaction is required to achieve good segmentation.
- If foreground and background share many same color than Grabcut fails. For example in case of camouflage it is hard to distinguish between background and foreground. Figure ?? is example of camouflage. Here the part of the rock is shown as the foreground because of the color information.

3.3 Circular Hough Transform

The same approach as detecting lines can be used here. In contrast to a Hough Transform, a Circular Hough Transform relies on 3 parameters, which requires a larger computation time and memory for storage, increasing the complexity of extracting information from our image. The required parameters are coordinates of the center and radius of circle. For simplicity, most programs set the radius to a constant value (hard coded) or provide the user with the option of setting a range of radius between minimum to maximum.

For each edge point, a circle is drawn with that point as origin and radius r . It uses an array (3D) with the first two dimensions representing the coordinates of the circle and the last third specifying the radii. The values in the accumulator (array) are increased every time a circle is drawn with the desired radii over every edge point. The accumulator, which kept counts of how many circles pass through coordinates of each edge point, proceeds to a vote to find the highest count. The coordinates of the center of the circles in the images are the coordinates with the highest count.

Advantages:

- Intersecting and partial circles are can be detected as shown in figure ??.
- Detection of radii works on concentric circles.
- works greatly with very little noise in image. It can be noticed in figure ??.

Disadvantages:

- If the minimum and maximum radius is not provided than it takes lots of time to detect the circles. A Program that entirely depends on user inputs is not practical for general use.
- Performance depend on edge detection and Gaussian filter. Changing the sigma value for Gaussian filter changes the output. Effect of changing sigma can be seen in figure ??.
- Amount of noise is more than it not able to detect circles properly. Figure ?? shows that.

3.4 Scale-invariant feature transform

SIFT is method for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of an object or scene. SIFT is invariant to image scaling and image rotation, while partially invariant to Change in illumination, change in 3D camera viewpoint and occlusion, clutter or noise. The SIFT algorithm is composed of following stages [2] :

1. Detection of Scale-Space Extrema - Identify location and scales that can be assigned under different views of the same object.
2. Key point Localization - Rejects points with low contrast and the points that are localized along edge.
3. Orientation Assignment - Each Key point is assigned to one or more orientations, based on local image gradient direction. Data is transformed relative to the assigned orientation, scale and location hence providing invariance to these transformation.
4. Local Image Descriptor Computes descriptors for the local image regions which are highly distinctive and and invariant as possible to the illumination, 3D viewpoints and occlusion.

Advantages :

- Difference in contrast does not affect the performance of SIFT. Figure ?? shows that even if contrast value is different for both the input, the detected interest points are the same and can be matched easily.
- SIFT is Partially invariant to illumination. This effect can be seen in figure ?. In output with affine illumination interest points in the dark part of the image are not detected. Otherwise it worked nicely.

Disadvantage :

- SIFT returns multiple interest points at the same location with different orientations if there is not clearly a single dominant orientation. It can be seen in output of SIFT in figure ? and ?.

3.5 Histogram of Oriented gradients

Histogram of Oriented Gradients (HOG) are feature descriptors for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy. Initially it was developed for Human detection but later it is expended to detect animals and vehicle in static imagery.

The HOG method tiles the detector window with a dense grid of cells. Each cell contains a local histogram over orientation bins. Image gradient vector is calculated at each pixel. The angle of the vector is used as a vote into the corresponding orientation bin and the vote is weighted by the gradient magnitude. Votes are accumulated over the pixels of each cell. The cells are grouped into blocks and a robust normalization process is run on each block to provide strong illumination invariance. The normalized histograms of all blocks are concatenated to give the window-level visual descriptor vector for learning. Spatial and angular linear interpolation, and in some cases Gaussian windowing over the block, are used during voting to reduce aliasing . The blocks overlap spatially so that each cell appears several times with different normalizations, as this typically improves performance [1] . Figure ? shows the output of HOG.

Advantages:

- Capture edge or gradient structure that is very characteristic of local shape.
- Relatively invariant to local geometric and photometric transformations - Within cell rotations and translations do not affect the HOG values and illumination invariance achieved through normalization.
- The spatial and orientation sampling densities can be tuned for different applications like human detection, gesture recognition and vehicle detection.

Disadvantage :

- Histogram of Oriented Gradient technique is not much useful in real time application as it takes lots of memory and time. For figure ?, 3780 descriptor are calculated for the image dimension of 64×128 and the time taken for that is 1.285 seconds. There are more techniques like compressed Histogram of Oriented Gradients which is fast and requires less memory.

References

- [1] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, CVPR '05, pages 886–893, Washington, DC, USA, 2005. IEEE Computer Society.
- [2] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, November 2004.
- [3] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, August 2004.