ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Regular Objects Identification in Aerial Images

K.Sai Divya
M.Tech Student,
Department of CSE,
Vasavi College of Engineering.

V.Punna Rao, M.Tech
Under The Guidance,
IIT kharagpur.

Dr.T.Adilakshmi
Professor & HOD,
Department of CSE,
Vasavi College of Engineering.

ABSTRACT:

In unmanned aerial vehicles or drones or missiles, etc the trajectory path is decided by the system with the help of camera images taken instantly by seeker part of vehicle, these images are matched with pre loaded or dataset images by using matching methods. For the matching purpose we extract features from images by feature extraction algorithms. One of the algorithm which is used for feature extraction is Scale Invariant Feature Transform (SIFT). The SIFT is a method for extracting interest point features from images, it detects interest point locations but also extracts features around the points that can be used to perform reliable matching between different views of an object or scene. The SIFT features is invariant to not only to image orientation but also image scale, and provide robust matching across a substantial range of affine distortion. This method can be used in the fields such as national development, military purposes, mapping and etc. Key points matching between two images are matched by identifying their nearest neighbours. The Euclidean distance method is used for matching images.

PROBLEM STATEMENT :

In unmanned aerial vehicles the object identification is important for choosing its right path. It is decided by the system with the help of camera images taken instantly by seeker part of vehicle, these images are matched with pre loaded images by using matching methods. For the matching purpose we extract features from images by feature extraction algorithm.

Introduction:

The progression of extraction and easy access to information, the processing of aerial images has been perceived as one of the first priorities by the researchers in the field of Pattern Recognition and Computer Vision. The important application of image analysis is the object recognition. It can be applied in various challenging fields like geology, natio development, military purposes, mapping, environmental monitoring, disaster response, and etc. The regular object is full of information but working with all the information associated with the object is time consuming and less efficient. Variability in scale, rotation, orientation, pose and illumination makes object detection a challenging task. Regular object has edges and corners. We can use them as features and extract those features by Feature extraction algorithm.Feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative, non redundant, facilitating the subsequent learning and generalization steps. When the input data to an algorithm is too large to be processed and it is suspected to be redundant, then it can be transformed into a reduced set of features.

This process is called feature extraction. Some of feature extraction algorithms are Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP) and Scale Invariant Feature Transform (SIFT). All of these approaches can be used in high-level feature extraction, where we find shapes in images. Common numerical programming environments such as MATLAB, SciLab and NumPy which is used for feature extraction techniques. In Histogram of oriented gradients, the gradient computation is the first step of calculation. The most common method is to simply apply the 1-D centered, point discrete derivative mask in both of the horizontal (DX) and vertical filter (D Y) kernels:

D_X= [-1 0 1] and D_Y= [-1 0 1] T; I_X= I * D_X and I_Y= I*D_Y. Where I is original image, IX and IY are derivatives. We can compute gradient magnitude $G = \sqrt{I_x^2 + I_y^2}$

and Gradient orientation $q = \tan^{-1}(\frac{I_y}{I_x})$

The second step is orientation binning which involves creating the cell histograms. Each pixel within the cell casts a weighted vote for an orientation-based histogram channel based on the values found in the gradient computation.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

The cells themselves can either be rectangular or radial in shape, and the histogram channels are evenly spread over 0 to 180 degrees or 0 to 360 degrees. For the vote weight, pixel contribution is the gradient magnitude.The Descriptor blocks account for changes in illumination and contrast, the gradient strengths must be locally normalized, which requires grouping the cells together into larger, spatially connected blocks. The HOG descriptor is the vector of the components of the normalized cell histograms from all of the block regions. These blocks typically overlap, meaning that each cell contributes more than once to the final descriptor. The Local Binary Pattern is other method for feature extraction algorithm. The LBP is created in the following manner:

•Divide the examined window into cells (e.g. 16x16 pixels for each cell).
•For each pixel in a cell, compare the pixel to each of its 8 neighbours (on its left-top, left-middle, left-bottom, right-top, etc.). Follow the pixels along a circle, i.e. clockwise or counter-clockwise.
•Where the center pixel's value is greater than the neighbour's value, write "1". Otherwise, write "0". This gives an 8-digit binary number (which is usually converted to decimal for convenience).
•Compute the histogram, over the cell, of the frequency of each "number" occurring (i.e., each combination of which pixels are smaller and which are greater than the center).
•Concatenate (normalized) histograms of all cells. This gives the feature vector for the window.

A local binary pattern is called uniform if the binary pattern contains at most two 0-1 or 1-0 transitions. The formula for Local Binary Pattern is shown below.

$$\mathbf{LBP}(x_c, y_c) = \sum_{n=0}^{7} 2^n \, s(i_n - i_c)$$

Where i_c corresponds to the value of centre pixel (x_c, y_c, i_n is the value of the eight surrounding pixels and function s(x) is defined as

$$s(x) = \begin{cases} 0 & if\ x < 0 \\ 1 & if\ x \geq 0 \end{cases}$$

The drawback of HOG and LBP are not invariant to scale & orientation.

Scale-Invariant Feature Transform

Scale-Invariant Feature Transform (SIFT) is a known local feature extraction method which detects and describes local features in images. The features extracted by Scale-Invariant Feature Transform are invariant to image scale, orientation, change in illumination and substantial range of affine distortion.

Scale-Invariant Feature Transform feature extraction method was initially developed for object recognition purposes. Lowe proposed to use Scale-Invariant Feature Transform features for object recognition. Though SIFT used for object recognition, issues like comparing similarities between tests and training image of two different objects. For efficient recognition, we propose to extract Scale-Invariant Feature Transform features from multiple object training images and set a threshold for maximum matching features. Since for a test image, there are many training images, thus image with the maximum matching features will be the recognized image. The objects are recognized by its corners points and we extract corners as features by using Harris corner detection algorithm. The steps followed by SIFT algorithm are:

1.Scale-space Extrema Detection
2.Key point localization
3.Orientation Assignment
4.Key point descriptor

1.Scale-space Extrema Detection:

This stage of the filtering attempts to identify those locations and scales that is identifiable from different views of the same object. This can be efficiently achieved using a "scale space" function. Further it has been shown under reasonable assumptions it must be based on the Gaussian function. The scale space is defined by the function.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

The first stage of computation searches over all scales and image locations. It is implemented efficiently by using a Laplacian-of-Gaussian function to identify potential interest points that are invariant to scale and orientation. The Laplacian L(x,y) of an image with pixel intensity values I(x,y) is given by

$$L(x, y) = \frac{\partial^2 I}{\partial^2 x} + \frac{\partial^2 I}{\partial^2 y}$$

Since the input image is represented as a set of discrete pixels, we have to find a discrete convolution kernel that can approximate the second derivatives in the definition of the Laplacian. This pre-processing step reduces the high frequency noise components prior to the differentiation step.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

In fact, since the convolution operation is associative, we can convolve the Gaussian smoothing filter with the Laplacian filter first of all, and then convolve this hybrid filter with the image to achieve the required result. Doing things this way has two advantages they are since both the Gaussian and the Laplacian kernels are usually much smaller than the image, this method usually requires far fewer arithmetic operations.The LoG ('Laplacian of Gaussian') kernel can be pre calculated in advance so only one convolution needs to be performed at run-time on the image. The 2-D LoG function centered on zero and with Gaussian standard deviation σ. Formula of the LoG

$$LoG = \left(\frac{-1}{\pi\sigma^4}\right) * \left[1 - \frac{x^2+y^2}{2\sigma^2}\right] e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Harris Corner Detection Algorithm:

i. Compute x and y derivatives of an image

$I_x = G_{x\sigma} * I$ and $I_y = G_{y\sigma} * I$

$$G_\sigma^x = \frac{1}{\sqrt{2\prod\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}$$

$$G_\sigma^y = \frac{1}{\sqrt{2\prod\sigma^2}} e^{\frac{-y^2}{2\sigma^2}}$$

ii. Computer products of derivatives at every pixel

$I_{x2} = I_x . I_x$, $I_{y2} = I_y . I_y$ and $I_{xy} = I_x . I_y$

iii. Compute the sums of the products of derivatives at each pixel

$S_{x2} = G_\sigma * I_{x2}$, $S_{y2} = G_\sigma * I_{y2}$ and $I_{xy} = G_\sigma * I_{xy}$

iv. Define at each pixel (x,y) the matrix

$$H(x,y) = \begin{bmatrix} S_{x2}(x,y) S_{xy}(x,y) \\ S_{xy}(x,y) S_{y2}(x,y) \end{bmatrix}$$

v. Compute the response of the detector at each pixel

1. $R = \text{Det}(H) - k(\text{Trace}(H))^2$

vi.Threshold on value of R. Compute non-max suppression

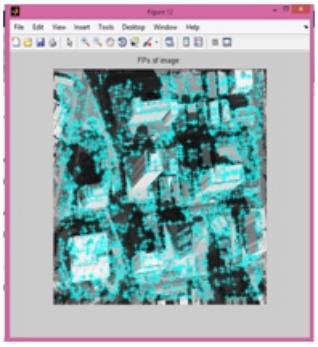


Figure: Harris corner detection

2.Key point Localization :

This stage attempts to eliminate more points from the list of key points by finding those that have low contrast or are poorly localised on an edge. If the function value at z is below a threshold value then this point is excluded. This removes extrema with low contrast. To eliminate extrema based on poor localisation it is noted that in these cases there is a large principle curvature across the edge but a small curvature in the perpendicular direction in the difference of Gaussian function. If this difference is below the ratio of largest to smallest eigenvector, from the 2x2 Hessian matrix at the location and scale of the key point, the key point is rejected.

3.Orientation Assignment :

One or more orientations are assigned to each key point location based on local image properties. All future operations are performed relative to the assigned orientation, scale and location for each feature, providing invariance to these transformations.The scale of the key point is used to select the Gaussian smoothed image L, with the closest scale, as all computations must be performed in a scale-invariant manner. For each image sample, Lx, y the gradient magnitude m, and orientation ɵ, is pre computed using pixel differences:

$$\text{Gradient magnitude } m = \sqrt{\left(L_{(x+1,y)} - L_{(x-1,y)}\right)^2 + \left(L_{(x,y+1)} + L_{(x,y-1)}\right)^2}$$

$$\text{Gradient orientation } \theta = \tan^{-1}\left(L_{(x+1,y)} + L_{(x-1,y)}\right) / \left(L_{(x,y+1)} + L_{(x,y-1)}\right)$$

4.Key point Descriptor :

The local gradient data, used above, is also used to create key point descriptors. The gradient information is rotated to line up with the orientation of the key point and then weighted by a Gaussian with variance of 1.5 * key point scale. This data is then used to create a set of histograms over a window centred on the key point.

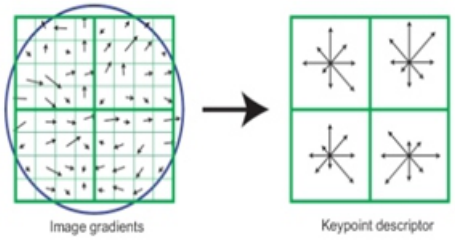


Figure: The computation of the key point descriptor

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Key point descriptors typically uses a set of 16 histograms, aligned in a 4x4 grid, each with 8 orientation bins, one for each of the main compass directions and one for each of the mid-points of these directions. These results in a feature vector containing 128 elements.

Key point matching:

The best candidate match for each key point is found by identifying its nearest neighbour in the database of key points from training images. The nearest neighbour is defined as the key point with minimum Euclidean distance for the invariant descriptor vector. It discard features that do not have any good match to the database. A global threshold on distance to the closest feature does not perform well, as some descriptors are much more discriminative than others.

A more effective measure is obtained by comparing the distance of the closest neighbour to that of the second-closest neighbour. If there are multiple training images of the same object, then we define the second-closest neighbour as being the closest neighbour that is known to come from a different object than the first, such as by only using images known to contain different objects. This measure performs well because correct matches need to have the closest neighbour significantly closer than the closest incorrect match to achieve reliable matching. For false matches, there will likely be a number of other false matches within similar distances due to the high dimensionality of the feature space. The Euclidean distance formula is $\sqrt{\sum(X-Y)^2}$.

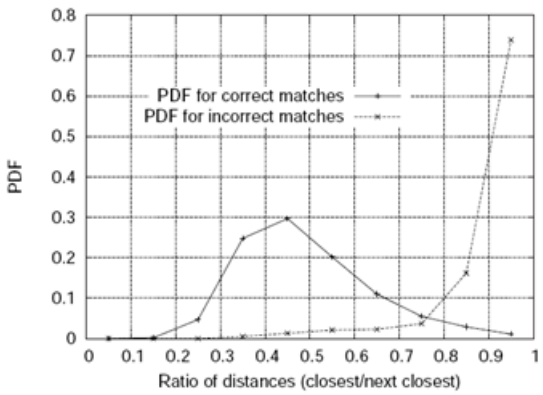
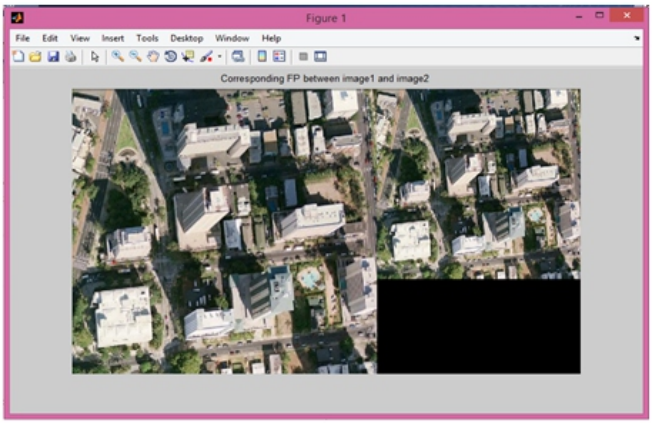


Figure: The probability that a match is correct can be determined by taking the ratio of distance from the closest neighbour to the distance of the second closest.

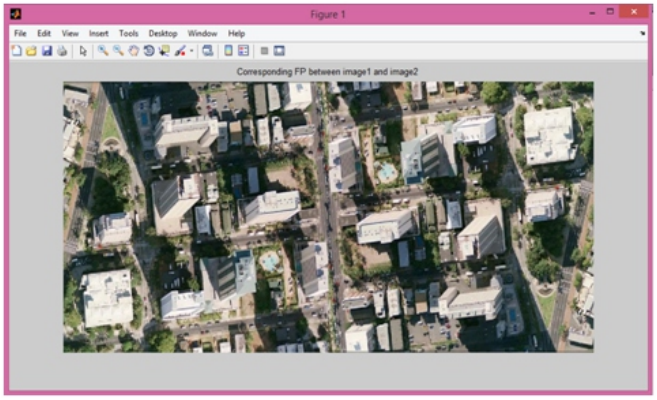
Figure shows probability density functions for correct and incorrect matches are shown in terms of the ratio of closest to second-closest neighbours of each key point. Matches for which the nearest neighbour was a correct match have a PDF that is centred at a much lower ratio than that for incorrect matches. For our object recognition implementation, we reject all matches in which the distance ratio is greater than 0.8, which eliminates 90% of the false matches while discarding less than 5% of the correct matches. This figure was generated by matching images following random scale and orientation change.

Results:

Output for scale invariant

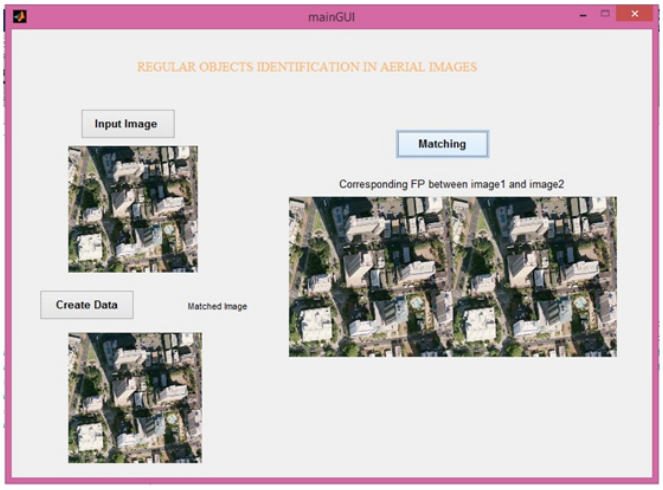


Output for invariant to rotation



ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Overall output in Gui representation



Conclusion:

The proposed system can presented a novel approach of detecting regular shape corners in aerial images. The intensity value for irregular shapes is a low or same value, but it is relatively high at corners in regular shapes. The algorithm has a good robustness against rotation and scale variation. Experimental results proved that the proposed technique could be very effectively applied for aerial images and it could be employed in the variety of significant applications and further processing such as detecting man-made or artificial structures. The SIFT key points are particularly useful due to their distinctiveness, which enables the correct match for a key point to be selected from a large database of test image key points. The key points have been shown to be invariant to image rotation and scale and robust across a substantial range of affine distortion, addition of noise, and change in illumination.