Face Detection and Recognition using Surf and Viola-Jones Algorithm

Jinal R. Thakkar¹

¹M.E Student

¹Department of Electronics & Communication Engineering

¹Kalol Institute Of Technology and Research Center, Kalol

Abstract—Face recognition has so many problems like to identify an image robustly of a human face. It also gives databases of some known faces. We use Viola-Jones algorithm for face detection and SURF based approach for face recognition. The principle behind the Viola-Jones algorithm is to scan a sub-window. To rescale the input image to different sizes and then after run the fixed size detector by using these images the standard image processing approach would be. It is time consuming approach because of the calculation of the different size images. On the Contrary by using the standard approach like Viola-Jones detector which is used for rescale instead of the input image and to run the detector many times by using the image – each time with a different size. Both approaches are equally time consuming, but Viola-Jones is a scale invariant detector which requires the same number of calculations of any size. This detector is constructed by using integral image and by using some simple rectangular features of Haar wavelets. This concept represents a scale-invariant and a novel rotation detector and descriptor known as SURF (Speeded-Up Robust Features). SURF is used to perform previously defined schemes with respect to repeatability, robustness and distinctiveness. It is more computing and comparing can be much faster. It is done through on integral images for image convolutions; by making the strengths of the leading existing detectors and descriptors (using a Hessian matrix-based measure for the use of detector, and a descriptor which is distribution-based). It gives result of combination of novel detection and description. This concept shows an overview of the detector and descriptor and then finds out the effects of the most important parameters. This paper uses SURF’s application to two challenging. It achieves such goals like camera calibration which is a special case of image registration and recognition of objects. By this experiment we can say that SURF is very fast and used for image recognition.

Key words: Viola-Jones face detector, the cascaded classifier, Face Normalization

I. INTRODUCTION

Biometrics is a subject that deals with the methods of recognizing humans uniquely based upon one or more intrinsic physical or behavioral traits. In Electronics & communication, it is used mainly as a form of identity access control and access management. It is divided into two main classes, based upon the traits. Behavioral: Also coined by some researchers as behavior metrics, these traits are related to the behavior of the person. Examples include typing rhythm, gait, voice etc. Generally a biometric system is operated in one of the below two modes:

1) Identification: This includes a one by one comparison of the unknown taken biometric, by using all the stored templates to identify or reject him as one of the person in the stored templates.

2) Verification: This process includes a one by one comparison of the taken biometric, with a saved template to verify whether the individual is who he claimed. It is done by conjunction by using username or ID number, a smart card.

3) Physical: Traits that related to the body structure. Palm print, DNA, hand geometry, face recognition, Iris recognition, fingerprint etc are examples.

II. PROPOSED SYSTEM

As face detection and recognition become very important today we are describing the whole process for face detection and recognition. First of all our aim is to detect the faces in the image. We have to extract the face from the image. We are using Viola-Jones algorithm for face detection purpose. Then we are using SURF algorithm for recognition purpose. We extract facial feature from that input image and save it. Then we compare that facial feature with the feature from the database. The contrast between that two feature extractions are match then face is recognized.

A. Numbers of Challenges in Face Recognition:

The major challenges in face recognition are the possibilities of when a face recognition system might fail. These include

- Speed, robustness and reliability of the system
- Illumination variation
- Head pose
- Change in the expression
- Change in Camera angle
- Growth of facial hair due to age or duplicated hair attached to fool the system
- Size of the scaled face image

Our system overcomes all of these problems using the SURF based features, which is Easier to calculate and are both rotation and scale invariant.

B. Viola-Jones Face Detector:

1) Scale Invariant Face Detector:

The first step is to find out input integral image of the input image in Viola-Jones face detection algorithm. This can be done by making the entire sum of all pixels and making each pixel like this and to the left of the concerned pixel. This
tables show how the input image is converted into integral image.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: Input Image

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 2: Integral image

This is used for calculation of sum of all the pixels which are given in the four values of rectangle. These values are the pixels of the integral image with the corners of the rectangle in the input image.

![Fig. 2: rectangle summations](Image 380x31 to 516x43)

\[ \text{Sum in grey} = D - (B + C) + A \]

The sum of A has to be added to the calculation as both rectangle B and C include rectangle A. So, it has been shown in figure that how it can be used arbitrary size for rectangle it can be calculated with respect to time. Viola Jones detector is used to analyze this sub window which contains features of two or more rectangles.

III. ADABOOST ALGORITHM

Modified AdaBoost algorithm at base resolution there can be 160000 feature values within a detector. On top of a face few feature are used to give almost high values. To find such features a modified version of the AdaBoost algorithm this was developed by Freund and Schapiro in 1996, used by Viola-Jones. It is a machine learning boosting algorithm. Through a weighted combination of weak classifiers AdaBoost is capable of constructing a strong classifier. (A weak classifier can do correction of only small bit more than half the cases.) Each feature is considered to a potential weak classifier. This algorithm is modified to select only features which are best, as only 160 feature values are expected to be potential weak classifiers. Determination of best feature, polarity and threshold is an important part of modified AdaBoost algorithm. There is no solution to this problem and brute force method is suggested by Viola-Jones as a simple solution. It is the most time consuming part of the training procedure which is used to determine of each new weak classifier involves evaluating each feature on all the training examples in order to find the best performing feature. Based on the weighted error it produces the best performing feature is chosen. This weighted error is a function of the weights belonging to the training examples. An alternative interpretation is that the second feature is forced to focus harder on the examples misclassified by the first. The point being that the weights are a vital part of the mechanics of the AdaBoost algorithm. The computationally efficient features and the modified AdaBoost algorithm in place it seems like the face detector is ready for implementation with the integral image, but Viola-Jones have one more solution.

IV. THE CASCAD ED CLASSIFIER

The principle behind the Viola-Jones face detection algorithm is to scan through the same image with the different size detector many times. It is obvious that an excessive large amount of the evaluated sub-windows would still be negatives (non-faces) if an image should contain one or more faces. This leads to a new problem: The algorithm should discard non-faces, Instead of finding faces. It is faster to reject nonface than to find a face. Suddenly it looks inefficient as the evaluation time remains constant so it is no matter the input as a detector consisting of only one (strong) classifier. So there arises the need for a cascaded classifier. With stages each containing a strong classifier the cascaded classifier is composed of. A given sub-window is definitely not a face or maybe a face is to determine. It is immediately discarded when a sub-window is declared to be a nonface by a given stage. Face is passed on to the next stage in the cascade conversely a sub-window classified. The higher the chance the sub-window actually contains a face it follows that the more stages a given sub-window passes.

V. PROPOSED APPROACH

A. Preprocessing:

We have preprocessed the images taken from the color FERET database before applying SURF which are used for testing purpose as stated in the documentation of color FERET database.

B. Face detection - Eye Center Coordinates:

Normalize the images as explained in the next section the eye center coordinates of both the eyes are need to be determined in order to do the face detection. To perform the eye detection and use that data to calculate the eye center coordinates, we already have constructed a Haar classifier based program. It gave false positives for 631 images, Out of the 2722 images tested. For these images, to do manually hand mark in the eye center coordinates, we used GIMP photo manager.

C. Face Normalization:

We used the CSU face evaluation system by supplying the eye center coordinates calculated as discussed in the previous section for normalization. Schedule of the normalization is as follows:

- Conversion from Integer to float - It is used to convert 256 gray levels into floating points.
  1. Normalization based on Geometric – It is used to lines up human detected eye coordinates.
  2. Use of Masking – It is used to crop the detected image by using an elliptical shape mask and to make image borders likewise the face from forehead to chin and cheek to cheek which is visible to us.
3) Equalization of Histogram– Which is used to equalize the histogram of the unmasked part of the image.

4) Normalization of Pixel – It scales the pixel values to have a mean of zero and a standard deviation of one.

This completes the process of normalization and it gives images in three formats:
- PGM
- SFI
- NRM

Fig. 4: Original to Normalized image

D. SURF (Speeded-up Robust Features) Algorithm:

1) Interest Point Detection:
To find out the interesting points we are using Hessian matrix approximation. So as given by Viola and Jones it is obvious to use integral images. Computation time is very less by use of it. It is proposed by Simard et al. that Integral images fit in the more general framework of box lets.

2) Integral Images:
We discuss how and what the integral images are. At location \( x = (x, y)^T \) as these used for fast calculation and computation of box type convolution filters. The sum of all pixels in the input image \( I \) bounded by a rectangular region formed by the origin and \( x \) represented by it. Breadth and length of the rectangle will be like this.

If the point \( X \) having coordinate \( (x, y) \).And Length= \( x \); Breadth=\( y \); Here the top-left corner is taken as the origin and along right wards it is taken as X-axis and along top-down wards it is taken as Y-axis.

\[
I_{\Sigma}(x, y) = \sum_{i=x}^{x+y} \sum_{j=y}^{y+j} I(x, y)
\]

Where \( 0<i<x \) and \( 0<j<y \);

3) Hessian Matrix-Based Interest Points:

The performance of Hessian matrix detector is good in accuracy. Mostly speaking, blob-like structures are detected so it’s our basis. These are the locations having maximum determinant value. In opposite to the Hessian-Laplace detector by Mikolajczyk and Schmidt, we focus on the determinant of the Hessian. It is also for the scale selection, as done by Lindeberg. Let us say a point \( x = (x, y) \) in an image \( I \), the Hessian matrix in \( x \) at scale \( r \) is defined as follows-

The SURF detector is based up on the determinant of the Hessian matrix. In order to utilize the use of the Hessian, we considered a continuous function of two variables in such a way that the value of the function at \( (x, y) \) is given by \( f(x, y) \). The Hessian matrix, \( H \), is the matrix of partial derivatives of the function \( f \).

\[
H(f(x, y)) = \begin{bmatrix}
\frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\
\frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2}
\end{bmatrix}
\]

The determinant of this matrix is known as the discriminant and is calculated by:

\[
det(H) = \left( \frac{\partial^2 f}{\partial x \partial y} \right)^2 - \left( \frac{\partial^2 f}{\partial x^2} \right) \left( \frac{\partial^2 f}{\partial y^2} \right)
\]

By the second order derivative test the values of the discriminants are used to classify the maximum and minimum value of the function. Since the determinant is the product of eigenvalues of the Hessian, so we can classify the points based on the sign of the result. If the determinant is negative then the eigenvalues have different signs and hence the point is not a local extreme; if it is positive then either both eigenvalues are positive or both are negative and in either case the point is classified as an extreme.

First the image pixel intensities \( I(x, y) \) is replaced by the function values \( f(x, y) \). Next we need a method to calculate the second order partial derivatives of the images. By convolution with an appropriate kernel we can calculate the derivatives. In case of SURF the second order scale normalized Gaussian is the chosen filter as it allows for analysis over the scales as well as the space. We can now calculate the Hessian matrix \( H \), as function of both space \( X = (x, y) \) and scale \( \sigma \).

\[
H(x, \sigma) = \begin{bmatrix}
L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\
L_{xy}(x, \sigma) & L_{yy}(x, \sigma)
\end{bmatrix}
\]

Here \( L_{xx}(x, \sigma) \) is the convolution of second order derivative \( \frac{\partial^2 g}{\partial x^2} \) with the image at point \( X = (x, y) \) and similarly for \( L_{xy}(x, \sigma) \) and \( L_{yy}(x, \sigma) \).

4) Interest Point Description and Matching:
The SURF descriptor describes how the pixel intensities are distributed within a scale dependent of neighborhood of each interest point detected by the Fast-Hessian. This approach is similar to that of SIFT but integral images used in conjunction with filters known as Haar wavelets are used in order to increase robustness and decrease computation time. Haar wavelets are simple filters which can be used to find gradients in the x and y directions.

Fig. 5: to calculate the sum of intensities inside a rectangular region of any size. It takes only three additions and four memory accesses.
5) **Detector – Descriptor Scheme:**
This can be decomposed into three parts -
- Fixing a reproducible orientation based on information from a circular region around the interest point.
- Construct a square region aligned to the selected orientation and extract the SURF descriptor from it.
- Finally features are matched between the two images.

6) **Representation of Scale Space:**
Interest points are necessary to find out at different scales, not least because the search of correspondences which often requires their comparison in images where they show at different scales. Scale spaces are generally implemented just like a pyramid of an image. The images are constantly smoothed with a Gaussian and then subsampled in order to achieve a strong pyramid. Lowe subtracts these pyramid layers. The main reason is to get the DoG (Difference of Gaussians) images so we can find blobs and edges. As we are using box filters as well as integral images, we do not need to apply again and again the same filter to the output of a previously filtered layer. But instead of this we can apply box filters of any size at exactly the same speed directly and even in parallel.

![Fig. 7](image_url)

**Fig. 7:** Rather than using repetitively reducing the image sizes (left), the use of integral images allows the up-scaling of the filter at constant cost (right).

**E. Descriptor Component:**
The first step in extracting the SURF descriptor is to construct a square window around the interest point. This window contains the pixels which will form entries in the descriptor vector and is of size $20\sigma$, again where $\sigma$ refers to the detected scale. Furthermore the window is oriented along the direction found in orientation assignment such that all subsequent calculations are relative to this direction. The descriptor window is divided into $4 \times 4$ regular sub-regions. Within each of these sub-regions Haar wavelets of size $2\sigma$ are calculated for 25 regularly distributed sample points. If we refer to the $x$ and $y$ wavelet responses by $dx$ and $dy$ respectively then for these 25 sample points (i.e. each sub-region) we collect. Therefore each sub-region contributes four values to the descriptor vector leading to an overall vector of length $4 \times 4 \times 4 = 64$. The resulting SURF descriptor is invariant to rotation, scale, brightness and, after reduction to unit length.

$$v_{\text{subregion}} = \left[ \sum dx, \sum dy, \sum |dx|, \sum |dy| \right]$$

![Fig. 8](image_url)

**Fig. 8:** To build $4 \times 4$ square detectors

1) **Feature Matching:**
The sign of the trace of Hessian matrix for the underlying interest point is included for fast indexing during the matching stage. The candidate is not considered a valuable match if the contrast between two interest points is different.

2) **Storing Feature Vectors:**
We need to store the vectors of key points and descriptors related to each and every image of the testing color FERET data set. We used the above SURF based method to construct an algorithm. Then we constructed a program based on that algorithm to extract these features in the form of vectors and store them in plain text files.

![Fig. 9](image_url)

**Fig. 9:** If the contrast between two interest points is different the candidate is not matching

VI. **RESULTS**

![Fig. 10](image_url)

**Fig. 10:** Original image
This paper uses the concept of VIOLA-JONES face detector for face detection because it is robust, distinctive, and less time consuming compared to other algorithms. After detecting the face, it uses the SURF ALGORITHM for face recognition. It approximates or even outperforms previously proposed schemes with respect to repeatability, distinctiveness, and robustness, yet can be computed and compared much faster.

VII. CONCLUSION

By using this concept, three improvements are done (1) dynamically build image Pyramid; (2) feature point filtering; (3) fast matching algorithm.

REFERENCES