Feature extraction: Face detection techniques and 3D object recognition based on local feature extraction

[1]Vighnesh Venkatakrishnan, [2]Ishan Shah, [3]Prof. Khushali Deulkar
[1]Student, [2]Student, [3]Assistant Professor,
[1]Computer Department,
[1]DJ Sanghvi College of Engineering, Mumbai, India

***Abstract* - Feature extraction and representation is an integral part of multimedia processing. How to extract ideal features which would reflect the intrinsic content of the images as complete as possible is still a very challenging problem in the field of computer science. However, very little progress has been achieved to find a solution to this problem in the last decades. So in this paper, we focus our review on the latest development in face detection and 3D object recognition and provide a comprehensive survey on the same. In this paper, we have discussed local feature based techniques that are used for 3D object recognition in cluttered scenes and the methodology that can be used for accurate face detection.**

***Index Terms* - keypoint detection, local surface feature, skin likelihood model, morphological operations**

I. INTRODUCTION

Feature extraction of images play an important role in image retrieval techniques. 3D object recognition methods can be classified into two broad categories: local feature based methods and global feature based methods. The object as a whole is used for processing in the global feature based recognition. A set of well-defined global features are described. This method is usually utilized in the context of 3D shape retrieval and classification. But this method ignores the shape details and requires a priori segmentation of the object from the scene. Hence, they are not capable of recognizing a partially visible object from cluttered scenes [1].

Human face detection is also a major scope in the field of research since there is no specific algorithm to detect faces from a given image. Also, the algorithms that exist are very much specific to the input that is given in order to detect the faces. Various methods like skin Segmentation, likely-hood model, Template matching and Morphological operation can be used for this purpose. Face detection is the 1st step in any automated system that solves problems such as: face tracking, face and facial expression recognition. Detection rate and the number of false positives are crucial factors for evaluating the face detection systems [2]. Detection rate is the ratio between the number of faces correctly detected by the system and the actual number of faces in the image. From many face detection algorithms, the method which is based on skin color model has been widely used because of its convenience, high detection speed and simple performance. Whenever there are a large no of objects similar to skin color, we need to utilize the other features of human face to further verify faces in different environmental settings

II. LITERATURE REVIEW

This paper is divided into two sections: First section presents 3D object recognition based on local feature extraction while the second section gives a detailed explanation of a specific technique that can be used for face detection.

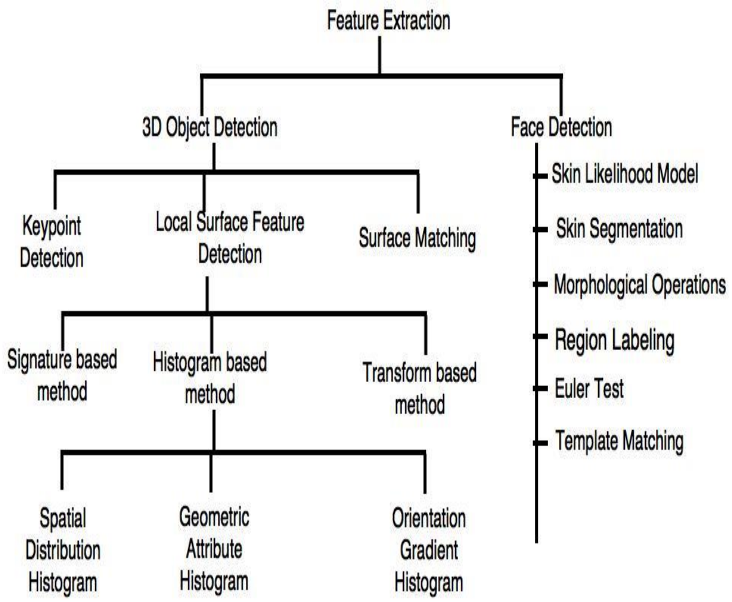

Fig 1: Classification of Feature Extraction

1) 3D OBJECT RECOGNITION

The local feature based methods extract only local surfaces around specific key-points. They are more capable of handling clutter than the global feature based methods [1]. This type is usually successful and hence common in the area of 2D object recognition. This method has also been successful for 3D object recognition. The primary aim of 3D object recognition is to correctly identify objects that are present in a 3D scene which is represented as a range image, and then to estimate the location or position and orientation of each object.) [3]. The range image is represented in three types- depth image, a polygon mesh or a point cloud.

For 3D object recognition based on local feature, the system needs to carry out three main phases: 3D keypoint detection, local surface feature description and surface matching. In the 3D keypoint detection phase, identification of the 3D points with rich information and its inherent scale is detected. The identified content are called as keypoints. A local surface is defined by the keypoint's location and its scale. This local surface is then used in the subsequent feature description phase. In the local surface feature description phase, encoding of the geometric information of the neighborhood surface of the keypoint is done into a representative feature descriptor. During the surface matching phase, the scene features are matched against all model features in the library. This results into a set of feature correspondences and hypotheses. These hypotheses are finally verified to conclude the identity and orientation of the object.

A. KEYPOINT DETECTION

The simplest keypoint detection methods are surface sparse sampling and mesh decimation. The problem with these methods is that they give very little consideration to the richness of discriminative information of these detected keypoints [4]. Hence, these methods do not give acceptable keypoints in terms of repeatability and informativeness. To overcome this, distinctiveness should be the main criteria for the detection of keypoints. Now, the scale can be predetermined or adaptively detected. Thus keypoint detection methods can be categorized into two categories: adaptive -scale keypoint detection methods and fixed-scale keypoint detection methods.

B. LOCAL SURFACE FEATURE DESCRIPTION

After detecting the keypoint, geometric information of the local surface in the proximity of that keypoint can be extracted, which can be further encoded into a feature descriptor. These methods are divided into three broad categories: histogram based, signature based and transform based methods.

a. Histogram Based Methods

Histogram methods accumulate geometric or topological measurements (like point numbers, mesh areas etc.) into histograms with respect to a specific domain. That is how they describe the local neighborhood of a keypoint [5]. These methods can further be classified as spatial distribution histogram (SDH), geometric attribute histogram (GAH) and oriented gradient histogram (OGH) based methods.

i. Spatial Distribution Histogram (SDH)

To describe the local neighborhood of a keypoint, these methods generate histograms according to the spatial distributions (e.g., point coordinates) of the local surface. First, an axis for the keypoint (LRF: local reference axis) is defined, and then a partitioning of the 3D local neighborhood into bins is done according to the LRF. Then the spatial distribution measurements (e.g., point numbers, mesh areas) are counted up in each bin to form the feature descriptor.

One such method is described by Johnson and Hebert [6]. They used the normal n of a keypoint p as the local reference axis (LRF) and expressed each neighboring point q with two parameters: the radial distance α and the signed distance β. They then discretized the α-β space into a 2D array accumulator, and counted up the number of points that fell into the bin indexed by (α, β). The 2D array was further bilinearly interpolated to obtain a "spin image" descriptor. This method is illustrated in Fig.1a. The spin image is invariant to rigid transformations and is robust to clutter and occlusion [6]. Being employed in many applications, it has become a benchmark for the evaluation of local surface features [5]. But, the spin image has some disadvantages, e.g., (i) It is sensitive to non-uniform sampling and varying mesh resolutions and [4]; (ii) Since the cylindrical angular coordinate is omitted, its descriptive power is limited.

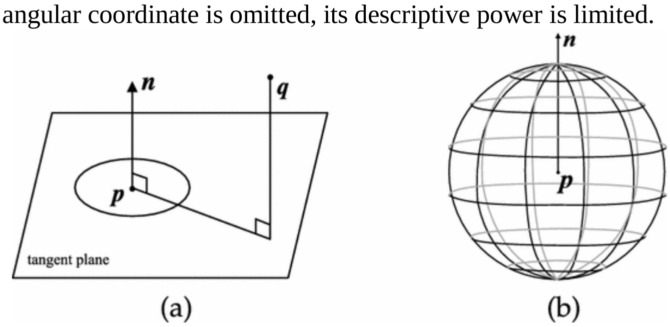

Fig 2: Illustration of spatial distribution histogram based methods. (a) Spin image. (b) 3D shape context.

ii. Geometric Attribute Histogram

By generating histograms according to the geometric attributes (like normals, curvatures) of the local surface, the methods present the local neighborhood of a keypoint.

One of the methods is described by Yamany and Farag [7]. They used simplex angles for estimating the curvature values on a free-form surface. By accumulating the simplex angles into a 2D histogram, they generated the "surface signature". One dimension of the histogram is the distance *d* from the keypoint **p** to a neighboring point **q**. Another dimension is the angle $\arccos\left(\frac{n.(q-p)}{||q-p||}\right)$, where ***n*** is the surface normal at ***p***. It was demonstrated that the surface signature was more descriptive compared to the splash, point signature and spin image [7].

iii. Orientation Gradient Histogram

By generating histograms according to the oriented gradients of the local surface, this methodology describes the local neighborhood of a keypoint.

One technique which utilizes this method is presented by Hua et al. [8]. First, it mapped a 3D surface to a 2D canonical domain. Then encoded the shape characteristics of the surface into a two-channel shape vector image. For each channel, a descriptor was generated using the same technique that was used in SIFT. That is, the 2D plane was divided into 16 subregions, and for each subregion an eight-element histogram was generated according to the orientations of the gradients, as shown in Fig. 2a. To form the overall descriptor, they concatenated all the histograms of the two channels. Experimental results solidified the fact that this descriptor indeed was very robust and well suited for surface matching purpose.

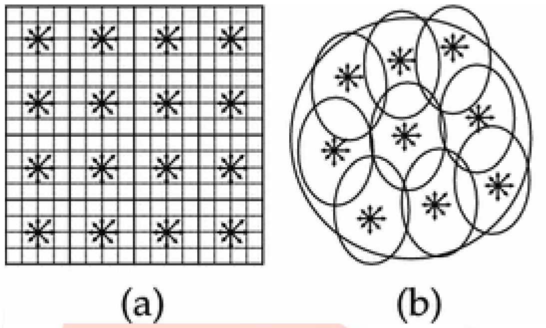

Fig. 3: Illustration of the oriented gradient histogram based methods. (a) The method in [8] and SI-SIFT. (b) 2.5D SIFT.

b. Signature Based Method

By encoding one or more geometric measures computed individually at each point of a subset of the neighborhood, these methods describe the local neighborhood of a keypoint [5]. One such method is given by Stein and Medioni [9]. Using the geodesic radius *r*, they first obtained a circular slice around the keypoint ***p***. By using the normal ***n*** and the tangent plane at the point ***p***, they constructed a local reference frame (LRF). The relationship (angular distance) between the normal at the point ***p*** and the normal at each point on the circular slice was encoded into a 3D vector (ϕ, ψ, θ) using this frame. Next, a straight line segment was fitted to this 3D curve, and the curvatures and torsion angles of the 3D segment were encoded as the "splash" descriptor. An illustrative example of the method is given in Fig. 3a. Experimental results proposed that this method is robust to noise.

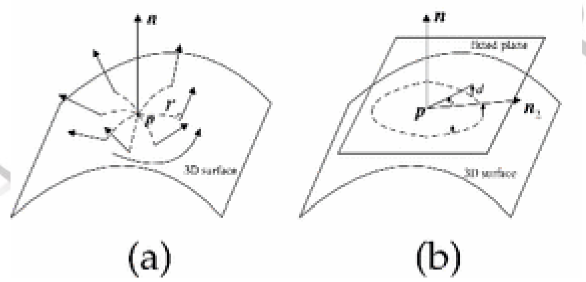

Fig. 4: Illustration of signature based methods. (a) Splash. (b) Point signature.

c. Transform Based Methods

First a range image is transformed from the spatial domain to another domain (e.g., spectral domain), and then the 3D surface neighborhood of a given point is described by encoding the information in that transformed domain.

Hu and Hua [10] presented one of the many methods. First, a Laplace-Beltrami transform is applied on the local surface to obtain the spectrum of the local surface. Then a histogram, which was used as the feature descriptor, was generated according to the spectral values of the local surface. Experimental results proposed that this descriptor was very powerful in matching similar shapes. The descriptor is also invariant to rigid transformations, isometric deformations and scaling.

C. SURFACE MATCHING

A majority of the existing surface matching methods comprise of three modules- feature matching, hypothesis generation and verification. First, by matching the scene features against the model features, a set of feature correspondences between the scene and the model are established. It then generates transformation hypotheses by using these feature correspondences to vote for candidate models. Next, these candidates are verified to obtain the identities and poses of the objects present in the scene.

2) FACE DETECTION

Whenever there are a large no of objects similar to skin colour, we need to utilize the other features of human face to further verify faces in different environmental settings. Colour images with skin colour in the chromatic and pure colour space YCrCb, which separates luminance and chrominance components. A Gaussian probability density is estimated from skin samples which are collected from different ethnic groups via the maximum-likelihood criterion. Then, mathematical morphological operators are used to remove noisy regions and fill holes in the skin-colour region, so we can extract candidate human face regions. . This system is used to achieve high detection speed and accuracy and reduce the false detecting rate. Let us study face detection using the methods we talked about earlier:

A. Skin Likelihood Model

For segmenting human skin regions from non-skin regions based on colour, we need a skin colour model that is reliable and adaptable to people of different skin colours at different lighting conditions at the same time. The common RGB representation of colour images is not suitable for characterizing skin-colour because in the RGB space, the triple component of r, g and b represents colour as well as luminance. Luminance is not a reliable measure in separating skin from non-skin region since it may vary across a person's face due to the ambient lighting. However, luminance can be removed from the colour representation in the chromatic colour space. Chromatic colours, also known as "pure" colours, in the absence of luminance, are defined by a normalization process shown below:

$$r=R/(R+G+B)$$
$$g=G/(R+G+B)$$
$$b=B/(R+G+B)$$

Since r+g+b=1, the green colour is redundant after normalization. Chromatic colours have been used in many applications in order to segment colour images effectively. Even though skin colours of different people appear to vary over a wide range, they differ much less in colour than in brightness. A colour distribution of skin samples can be obtained by extracting the samples from colour images and passing them through a low pass filter in order to reduce the noise. A Gaussian distribution model can be obtained with the help of the impulse response from the low pass filter with the help of which, the likelihood of skin on any type of pixel of an image can be obtained. Thus, this skin color model can transform a normal colour image into a grayscale image in such a way that the gray value at each pixel shows the likelihood of the pixel being that of a skin.

B. Skin Segmentation

The process of partitioning an image into various segments is called segmentation. There are various methods available for segmentation based on colour. Precise segmentation of the image is the most important step contributing to localization and efficient detection of multiple faces in skin tone colour images. Skin regions can be separated from the remaining regions in an image through thresholding since they are much brighter than other parts of the image. It is not possible to find a fixed threshold value for images of people with different skin tones. It is necessary to find an adaptive threshold for different skin likelihoods. The adaptive thresholding is based on the observation that stepping the threshold value down may intuitively increase the segmented region. If the threshold value is too small, non-skin regions are also included. The threshold value at which the non-skin region is minimum would be taken as the optimum threshold. Using this method of adaptive thresholding, the skin regions are accurately segmented from normal regions.

C. Morphological Operations

Morphological operation simplifies image data and preserves their essential shape characteristics while eliminating irrelevant regions. It helps in giving a more accurate contour of the skin segment while using it. Morphological process includes erosion and dilation in order to investigate binary image where ’1‘ represents skin pixels and ’0‘ represents non-skin pixels and morphological operations for separating the skin areas that are closely connected. Morphological erosion is applied with the use of structuring elements of a particular disk size. With the application of dilation, the binary skin areas which are lost due to aggressive erosion step are regrown. The dilated binary image is then multiplied with the binary image from segmentation to maintain the holes.

D. Region Labeling

The binary images obtained after morphological operations has to be labelled in such a way that every cluster in the image can be identified as a single group or region and then later deteremine if it is a face region or not. Thus, instead of ones and zeroes, the regions are labelled as 1,2,3,4… Pixels having values 0 are left unchanged. A function can be used to show different regions in different colours.

E. Euler Test

Human face consists of distinct parts like nose, mouth, eyes, eyebrows, etc so its variance is much greater than other body parts like hands and arms. The number of holes in a region is computed by the Euler number

$$E=C-H$$

Where C is the number of candidate regions and H is the number of holes in a region.

The regions having Euler number greater than 15 or less than 1 are discarded. After this, the aspect ratio of every region is calculated in order to determine whether the face regions that have been detected are acceptable or not. This is mainly to distinguish between the face and the neck.

F. Template Matching

This is the final step in face detection which performs cross correlation between template face and grayscale region. The template face is the average frontal face of a person. This final stage includes the following processes:

1) Firstly, height, width, orientation and centroid of the binary region under consideration are to be computed.
2) Then, the template face image is resized, rotated and its centroid placed on the centroid of the region in original grayscale image with only one region.
3) The rotated template needs to be cropped properly and the size has to be same as that of the region.
4) Then the obtained face is cropped, scaled and stored in a specified location using the digital image processing techniques, and converted into a gray image of size 92 x112 pixels and saved as individual images for further processing in face recognition.

III. CONCLUSION

This paper displays the feature extraction methodology for 3D object recognition based on local surface features and face detection techniques. A comprehensive study of the object recognition process which consists of keypoint detection, local surface feature detection, and surface matching is presented with one technique in each procedural step. A study on one of the many face detection algorithms is also done which consists of various steps- skin likelihood, skin segmentation, morphological operation, region labeling, Euler test and template matching.

IV. ACKNOWLEDGMENT

We would like to express our sincere gratitude to our guide Prof. Khushali Deulkar for her guidance, encouragement and gracious support throughout the course of our work, for her expertise in the field that motivated us to work in this area and for her faith in us in every stage of this research. We're grateful to Dr. Narendra Shekokar, Head of the Computer Engineering Department, for letting us use the department resources, labs and books. We're highly indebted to our Principal, Dr. Hari Vasudevan for availing us with library books and relevant materials and also the SVKM management for providing us with popular and resourceful journals like IEEE and other online material that helped us finish our project with ease and perfection. We would also like to thank all our fellow students and staff of Department of Computer Engineering for their help in the whole process leading to the conceptualization of the project.

REFERENCES

[1] Bayramoglu, Neslihan. "Shape index SIFT: Range image recognition using local features." *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010.

[2] Lamiaa, Mostafa and Sherif Abdelazeem, ―Face Detection Based on Skin Color Using Neural Networks‖ GVIP 05 Conference, 19-21 December 2005, CICC, Cairo, Egypt
Journal of Computer Theory and Engineering, Vol. 2, No. 4, August, 2010. pp. 552–558.

[3] Bariya, Prabin, and Ko Nishino. "Scale-hierarchical 3d object recognition in cluttered scenes." *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010.

[4] Mian, Ajmal, Mohammed Bennamoun, and Robyn Owens. "On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes." *International Journal of Computer Vision* 89.2-3 (2010): 348-361.

[5] Tombari, Federico, Samuele Salti, and Luigi Di Stefano. "Unique signatures of histograms for local surface description." *Computer Vision–ECCV 2010*. Springer Berlin Heidelberg, 2010. 356-369.

[6] Johnson, Andrew E., and Martial Hebert. "Using spin images for efficient object recognition in cluttered 3D scenes." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 21.5 (1999): 433-449.

[7] Yamany, Sameh M., and Aly Farag. "Surface signatures: an orientation independent free-form surface representation scheme for the purpose of objects registration and matching." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.8 (2002): 1105-1120.

[8] Hua, Jing, et al. "Geodesic distance-weighted shape vector image diffusion." *Visualization and Computer Graphics, IEEE Transactions on* 14.6 (2008): 1643-1650.

[9] Stein, Fridtjof, and Gérard Medioni. "Structural indexing: Efficient 3-D object recognition." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 2 (1992): 125-145.

[10] Hu, Jiaxi, and Jing Hua. "Salient spectral geometric features for shape matching and retrieval." *The visual computer* 25.5-7 (2009): 667-675.

[11] H C Vijay Lakshmi and S. PatilKulakarni. "Segmentation Algorithm for Multiple Face Detection in Color Images with Skin Tone Regions using Color Spaces and Edge DetectionTechniques ". International Journal of Computer Theory and Engineering, Vol. 2, No. 4, August, 2010. pp. 552–558.

[12] Upraity, H.; Arya, K.V. "Efficient face recognition using morphological operations", Industrial and Information Systems (ICIIS), 2014 9th International Conference on
Year: 2014.

[13] Pallabi Saikia , Gollo Janam , Margaret Kathing "Face Detection using Skin Colour Model and distance between Eyes" , International Journal of Computing, Communications and Networking, Volume 1, No.3, November – December 2012.

[14] S. Kherchaoui and A. Houacine , "Face Detection Based On A Model Of The Skin Color With Constraints And Template Matching"", 978-1-4244-8611-3/10/$26.00 ©2010 IEEE.