Glasses Frame Detection with 3D Hough Transform

Haiyuan WU*, Genki YOSHIKAWA, Tadayoshi SHIOYAMA, †Shihong LAO and †Masato KAWADE
Dept. of Mechanical and System Engineering, Kyoto Institute of Technology, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan.
†Information Technology Research Center, OMRON Corporation, Japan.
e-mail: wuhy@sys.wakayama-u.ac.jp

Abstract

This paper describes a method to detect glasses’ frames for robust facial image processing. This method makes use of the 3D features obtained by a trinocular stereo vision system. The glasses frame detection is base on a fact that the rims of a pair of glasses lie on the same plane in 3D space. We use a 3D Hough transform to obtain a plane in which 3D features are concentrated. Then, based on the obtained 3D plane and with some geometry constraint, we can detect a group of 3D features belonging to the frame of the glasses. Using this approach, we can separate the 3D features of a glasses frame from the ones of facial features. The characteristic of this approach is that it does not require any prior knowledge about the face pose, the eyes positions and the shape of the glasses.

1. Introduction

A robust facial image processing system ought to work correctly under a large variety of environments, such as different lighting conditions, changing of face pose or scale, and wearing glasses or not. In many cases, the performances of the facial image processing system are affected by the presence of the glasses. To make a facial image processing system more robust, it is helpful to detect the glasses. Several researches on the detection of glasses using the edge information from a monocular image have been reported[1] [2].

Assuming the face in the input image is in front pose and the face region has been detected, Jiang[1] et.al. have proposed a judgment method to decide whether or not a face wears glasses by checking six measures of edge information within several regions near the eyes. This method only gave information about the existence of the glasses, did not detect the position and the shape of the glasses.

Jing[2] et.al. have developed a glasses detection algorithm with a deformable contour based on both the edge features such as strength, orientation and the geometrical features including convexity, symmetry, smoothness and continuity. However, because this method requires that the face is frontal and the eyes’ position is known, the scope of this algorithm’s application is limited.

This paper presents a new method to detect glasses frames and separate it from the facial features using stereo images. We investigated the 3D shape of a various kinds of glasses, and discovered that the rims lie on a same plane approximately, while the other facial features do not. From this fact, we consider that the rims should be able to be detected by finding out a group of 3D features that lie on the same plane. The Hough transform is a powerful tool to detect the specified geometrical figures among a cluster of data points. In this paper, We use a segment-based trinocular stereo vision system VVV to detect the 3D features of objects (eyes, nose, mouth, glasses frames, etc.). And we use a 3D Hough transform to determine the plane containing the rims from the detected 3D data. Based on the determined 3D plane and with some constraints on the geometric relation, we first detect a group of 3D features belonging to the rims, and then separate the ones of a glasses’ frame. Since we do not make any assumption about the face pose, the shape of the glasses, and do not require any prior information about the position and the shape of the facial features, the proposed method would be very useful in a wide range of face image applications where the detection of glasses is required. Experimental results demonstrate the effectiveness of this method.

2. 3D Feature Reconstruction

Tomita et.al. are developing a versatile 3D vision system VVV (Versatile Volumetric Vision)[4] [5] [6] which consists of five subsystems: a trinocular stereo camera subsystem, a stereo vision subsystem that reconstructs 3D information of a scene, an object recognition subsystem, a motion tracking
subsystem, and a model building subsystem which generates object models from the all-around range data.

If we consider a three-dimensional space defined by the three parameters \( \rho, \theta, \) and \( \phi \), then any plane in \( X-Y-Z \) space corresponds to a point in this \( \rho-\theta-\phi \) space. Thus, the 3D Hough transform of a plane in \( X-Y-Z \) space is a point in \( \rho-\theta-\phi \) space.

Now considering a particular point \( B(x_b, y_b, z_b) \) in \( X-Y-Z \) space. There are many planes that pass through this point. All these planes can be expressed with the following equation from eq (1):

\[
\rho = x_b \sin \theta \cos \phi + y_b \sin \theta \sin \phi + z_b \cos \theta
\]  

(2)

Thus all planes that pass the point \( B(x_b, y_b, z_b) \) can expressed with a curved surface described by eq (2) in \( \rho-\theta-\phi \) space.

If we have a set of 3D data points \( (x_i, y_i, z_i) \) that lie on a plane having parameters \( \rho_0, \theta_0, \) and \( \phi_0, \) then for each 3D feature point we can plot a curved surface in \( \rho-\theta-\phi \) space that describes all planes that pass it. However, all these curved surfaces must intersect at the point \( P(\rho_0, \theta_0, \phi_0) \), because it corresponds the plane that all the 3D feature points fall on. (Figure 3(b))

Thus, to find the plane where a group of 3D data points fall on, we set up a three-dimensional histogram in \( \rho-\theta-\phi \) space. For each 3D data point, \( (x_i, y_i, z_i) \), we increment all the histogram bins that the curved surface describing all planes that pass \( (x_i, y_i, z_i) \) crosses. When we have done this for all of the 3D data points, the bin containing \( (\rho_0, \theta_0, \phi_0) \) will be a local maximum. Thus, we search the \( \rho-\theta-\phi \) space histogram to find the local maximum to obtain the parameters of the plane.

4. Glasses Frames Detection

4.1. Obtaining a 3D plane of the rims

With the 3D Hough transform, we can find the global maximum frequency \( \text{Freq}_{\max} \) in the \( \rho-\theta-\phi \) space from the extracted 3D data points \( (x_i, y_i, z_i) i = 1 \ldots n \). Then we choose \( N \) points of the local maximum that the frequency is bigger than \( \text{Freq}_{\max} \times 0.9 \). Let \( \text{freq}(\rho_n, \theta_n, \phi_n); n = 1, \cdots, N \) be the frequency of each point. We select the 3D plane as the \textit{rim plane} from the maximum of frequency accumulation value of a \( n(=1, \cdots, N) \) point and its vicinities as following.

\[
\max_{n=1, \cdots, N} \left\{ \sum_{\Delta \rho=-4}^{4} \sum_{\Delta \theta=-1}^{1} \sum_{\Delta \phi=-1}^{1} \text{freq}(\rho_n + \Delta \rho, \theta_n + \Delta \theta, \phi_n + \Delta \phi) \right\}
\]

Here, \( \Delta \rho \) is done at intervals of every 1 mm. \( \Delta \theta \) and \( \Delta \phi \) are done 4 degree notch, respectively.

4.2. The algorithm of rims detection

Based on the \textit{rim plane}, we determine some restricting conditions that considered both the geometric characteris-
tics of the rims, and the error included in the 3D data points. The 3D segments that met all the following conditions are selected as candidates of rim.
(1) The 3D segment is circular or straight line.
(2) If the 3D segment is circular, then the radius of the circular segment is more than 10 mm.
(3) If the 3D segment is straight line, then the length of the straight-line segment is more than 5 mm.
(4) The mean value of the distance between all of the 3D data points of the 3D segment and the rim plane is less than 5 mm.

From the selected candidates of 3D segments, we use the following conditions to detect the parts of rim.
(1) The 3D segment is a long segment (length ≥ 20 mm), and the scalar product between the normal vector of the rim plane and vector of the 3D segment is more than 0.9.
(2) If it is a short segment, first, we connect it with the nearest 3D segment, which has similar tangent vector, until the length of connected 3D segments is more than 20 mm. Then, we calculate the scalar product between the normal vector of the rim plane and vector of the connected 3D segment as a long segment.

4.3. The algorithm for glasses’ frames detection

To detect segments that compose glasses’s frames includes all the rims, bridge and earpieces. We first project all 3D Data points on the glasses rim onto the rim plane. Next we determine the rim region, which is defined as the smallest rectangular region on the rim plane that just contains all the projected 3D Data points of the rim. Then, from all of the 3D segments that were obtained with the VVV system, the 3D segments that met all the following conditions are detected as parts of glasses frames.
(1) At least one of the 3D data points of a 3D segment is lies within the rim region when projected onto the rim plane.
(2) At least one of the 3D data points of a 3D segment that the distance between the rim plane and it is less than 3 mm or, if it is one the camera side relative to the rim plane, the distance between it and the rim plane is less than 10 mm.

5. Experiments and Results

We have applied our glasses frames detection algorithm to some stereo face images to test its performance. The images contain faces of 19 people. Each person wore 3 kinds of glasses and took face images under 9 kinds of pose. The trinocular stereo camera system of the VVV was calibrated before taking stereo images. The baseline between the upper and the lower camera is about 30 centimeters and the distance between the baseline and central camera is about 16 centimeters. The distance between the cameras and the face is about 1 meter.

Figure 4 shows some example results. In figure 4, (a) shows the central image of the stereo image set, and the face in different pose, and wearing different glasses, (b) shows the reconstructed 3D features, and (c) shows the extracted feature points of the glasses.

The successful rate of the glasses detection only the frames were 80%, and the detection rate of glasses frames including some eyelids or irises was about 90%. The processing time is about 3 seconds on a PC with a Celeron533 processor.

6. Conclusion

This paper has described a new method to detect the glasses’ frames from stereo face images using a 3D Hough transform. The experiment results have shown that the proposed method is able to separate the glasses frames from the face images without the restriction of the face pose, the position of eyes, and the shape of glasses.

Acknowledgement: We would like to thank Dr. Y. SUMI and F. TOMITA at 3D Vision System Research Group, Intelligent Systems Institute, National Institute of Advanced Industrial Science and Technology, Japan.

References

(a) Polar coordinate expression of a 3D plane

(b) $\rho$-$\theta$-$\phi$ space

Figure 3. 3D Hough Transform.

(a) The central image of the stereo image set

(b) The reconstructed 3D features

(c) The extracted feature points of the glasses

Figure 4. Some experimental results.