Head detection and localization from sparse 3D data

Abstract. Head detection is an important, but difficult task, if no restrictions such as static illumination, frontal face appearance or uniform background can be applied. We present a system that is able to perform head detection under very general conditions by employing a 3D measurement system namely a structured light distance measurement. An algorithm of head detection from sparse 3D data (19x19 data points) is developed that reconstructs a 3D surface over the image plane and detects head hypotheses of ellipsoidal shape. We demonstrate that detection and rough localization is possible in up to 90% of the images.

1. Introduction

Head detection is the starting point of several methods that are elaborated in the fields of face recognition, gesture recognition and man machine interaction. Usually the detection can be based on the robust detection of an outline (shape) [3] or based on general appearance [10]. The outline of the head can only be obtained reliably when a uniform background and non varying illumination is assumed Appearance based methods perform well in case of frontal face appearance but show high false detection rates, when the head of a person shall be detected in arbitrary pose i.e. [11].

Some authors have proposed to use 3D information for head detection [3,4,6,8]. All these approaches either use a pre-selected set of data points that are known to lie at the person, or they use a very dense data set coming from stereo/disparity imaging. In this paper we present a novel system that performs fast detection of head hypotheses using sparse 3D information derived by a simple structured light technique (19x19 points). The head search is performed by surface reconstruction and an ellipse search within equidistant planar cuts parallel to the image plane.

The paper is organized as follows: First the acquisition system similar to [2] is presented. Improvements of the technique are summarized and a new method to solve the correspondence problem is presented in section 2. Section 3 explains the head detection method that comprises surface reconstruction, ellipse search using planar surface cuts and ellipse hypothesis selection. Section 4 presents the results based on 2 real sequences, section 5 gives a summary and draws conclusions for future algorithms and applications.

2. Acquisition System

The acquisition system is based on the well known technique of structured light projection [7,12] and acquires a 3D point cloud of 361 points at maximum. It consists of the four steps: (i) projection of a dot matrix pattern (ii) dots detection (iii) dot labeling (iv) distance calculation.

2.1. Structured light technique to acquire distance points

Structured light is a well known technique for obtaining 3D information in various applications [13,14]. It is used to measure distances by projecting artificial feature points (patterns) onto a scene of interest. These points are emitted from a light projector and therefore lie on a line (epipolar line) in the image. The location of the point on the line (disparity) measures the position of the projected point in 3D space. The 3D coordinates can then be calculated via triangulation using the known positions of the calibrated camera and the projector.

Our implementation of the structured light system comprises two main devices: a camera and a pattern generation unit. The camera is a CCD-video camera with 752(H) x 582(V) picture elements and a chip size of 4.9mm (H) x 3.7mm (V). The pattern generation unit consists of a laser light source (LASIRIS-model: 670 nm, 10 mW), and a changeable beam shaping optics to generate the projection pattern (19x19 dot matrix). Both devices are mounted on a stable board at same height, without any horizontal tilt. From a top view the camera is pointed inwards at a distance of 12.5mm to the projector. The camera was calibrated by using the techniques presented by [5] and [17]. The projected dot pattern was rotated against the horizontal line to achieve a near-degenerate epipolar alignment [1].

2.2. Dot detection

To segment the projected dots from the rest of the image the simple thresholding procedure [2] was improved by applying a difference of Gaussian (DoG) filter operation to the Image I, which gives I_F. This filter checks the property that projected dots can be found as a small area of pixels that have higher intensity as the pixels in its surrounding. After filtering I_F is thresholded to get a binary output I_T that is used to determine the central point of each dot by calculating the center of gravity (CoG).

2.3. Dot labeling

Dot labeling is the task of assigning the detected dots to the corresponding epipolar line, defined by the projector and camera positions. This is necessary to allow a proper distance calculation and can be accomplished by using spatial and temporal constraints as intensively discussed in [2]. Here we add an alternative that can be used especially to improve the speed of the initialization procedure. It is based on the principle that the surfaces on which the dots are projected are locally approximately planar and therefore the projected dot pattern of rectangular shape is transformed by an affine transformation. The algorithm has two major steps:

First every four epipolar lines that belong to a rectangular part of the projected dot pattern are tested, whether their corresponding dots satisfy the affine transformation condition. This is tested by calculating the parameters of an affine transformation in a least square sense. If the error of the transformation is less than a limit ε, the dots are assigned to the corresponding epipolar lines.

The second step is performed by region growing. Stepwise all lines are labeled that are close to already assigned lines. The condition that neighboring dots satisfy the affine transformation condition is also applied. The algorithm increases the allowed deviation limit ε, if the number of assigned dots per iteration step is low (<5). This guarantees that the region grows in planar region first. After that discontinuities are overleaped due to the increase of allowed deviation limit ε. The two main assumptions of rectangular shaped pattern and projection on a planar surface may be violated to a limited degree which is given by the size of the deviation limit ε.

If the dots are detected and labeled correctly and if the camera is calibrated, distance calculation can be done by triangulation using the distance between camera and projector.

3. Head detection module

The structured light system measures surface points in 3D. The described system yields 361 (maximal) or fewer points (usually 20 to 70 points are occluded). The critical information that should be extracted from the point cloud is:

- Foreground/background separation: The background can be an arbitrary scene, the foreground is the head/thorax region of a human. This separation is simple, if the background can be modeled very accurate. All points that do not lie on the background model belong to the foreground.
- Localization of the head region: To localize the head, a model is fitted to the point cloud. This can be done either directly or by determining several parameters of the surface that is spanned by the given points.

In the following it is assumed that the background model is not available. Therefore the foreground/background separation can either be determined by searching for the relevant depth discontinuities in the point cloud which is a non-trivial problem for sparse (19x19) depth measurements, or by fitting a foreground (head/thorax) model onto the data. Trying to fit i.e. an ellipsoid model to the data points was not robust, because of the usually small amount of points that lie on the head surface.

3.1. Depth surface reconstruction

The depth of the surface is a function that arises over the image plane (*x,y*). The given data points *P(x,y,z)* define this surface *z=f(x,y)*. The surface reconstruction can therefore be seen as an approximation problem. The data points define a function *z* that shall be approximated. The approximation shall exhibit following properties:

- Robust approximation: outliers should have no influence on the approximation.
- Strong separation: different objects should appear separated in the representation. Therefore depth discontinuities should be preserved.
- Correct extrapolation: as data points are defined in a certain area, the behavior of the function at the borders should not influence its behavior in the defined area.

This surface can be approximated by a sum of basis functions *B(x,y)*

$$z = \sum_{n=1}^{m} c_n B_n(x, y) \tag{1}$$

The coefficients c_n are determined by solving the system of linear equations given by the coordinates of the data points *P* in a least square sense. The choice of the basis function B_n is critical to achieve the mentioned properties. Different multiquadrics and Gaussians were tested. It turned out that head and shoulder region contrasts from the background by applying

$$B(x, y) = \sqrt{x^2 + y^2 + 1}\,. \tag{2}$$

The advantage of a basis function that tends to ∞ lies in the fact, that head regions that lies at the edge of the working area stays a convex region, while it appears to be concave for functions that tend to 0.

3.2. Planar surface cuts (isodistance lines)

To analyze the resulting surface *z*, different representations can be used. For the specific task of finding the head region, the surface was represented by its planar cuts at equal distances. Therefore the surface was cut by parallel planes of equal distances and the resulting polygons were analyzed.

By taking the polygons stepwise from the nearest cut to the furthest, the center of gravity of *closed* polygons mark the extremas of the function. Growing confocal polygons are minimas (convex, from camera perspective), shrinking polygons (concave) are maximas. Minimas can be marked as head/thorax-hypotheses. Every closed polygon that surrounds a minimum is marked as supporting the corresponding hypothesis. Therefore each hypothesis consists of a sequence of growing polygons.

Open polygons must be handled differently (Fig.1). As they cannot be assigned to a certain hypothesis, they are cut into parts, where at least one polygon part has the following properties: (i) it is of elliptical shape (ii) it surrounds and therefore supports an existing hypothesis. If such a cut is not possible, the open polygon is not used. After considering all polygons usually a small number (~5) of elliptical hypotheses remain.

3.3. Ellipse search and hypothesis analysis

When a hypothesis is found, this represents a convex surface part of the overall surface. Every hypothesis is tested, whether it is a head region or not. Following tests are applied:

(i) Area limits: A head must have a certain minimum and maximum size. Hypotheses that violate those limits can be removed.

(ii) Curvature limits: The head can be approximated by an ellipsoid. Therefore planar cuts are ellipses that have same focus and have increasing area. The increase from one slice of the ellipsoid to the next can be used to calculate the curvature of the ellipsoid. If this curvature is beyond a certain interval, the corresponding hypothesis can be removed.

(iii) Shape limits: The polygon must be of elliptical shape. Additionally the ratio of the axis of the ellipse must be in a certain range, as very narrow ellipses are not allowed. Hypotheses that violate this condition are removed.

All remaining hypotheses are assumed to be head regions.

4. Results

We have tested the reconstruction and ellipsoid finding algorithm on 2 sequences of 50 image frames each. The sequences show one sitting person (46 frames each), some frames show an empty seat (4 each). Sequence 1 had no hand movements, whereas in Sequence 2 the hands were moved strongly, partly occluding the head or regions close to the head. The results can be summarized as follows: Sequ.1.: In all images showing a person the head location is found as elliptical hypothesis (46 out of 46). In 37 cases the head was found without any ambiguity. Sequ.2: From the 46 images showing a person the head location is found as elliptical hypothesis in 41 cases. In 13 cases the head was found correctly, in 18 cases more than one hypothesis was selected and in 3 cases no hypothesis was selected.

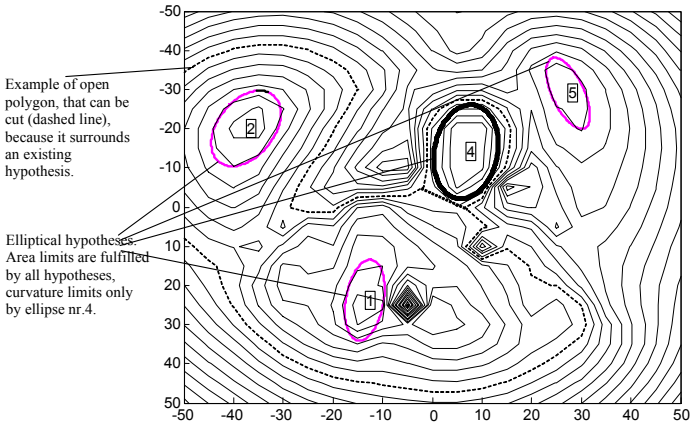

Fig.1.: Contour plot of reconstructed surface with ellipse hypotheses. Darker lines are contours that are close to the camera. Dashed line shows an example of an open polygon and a cut of an elliptical part of the polygon. Ellipses show hypothesis, dark ellipse fulfils all head conditions, bright ellipses violates curvature condition.

Table 1.: Result of head detection process of two sequences. In most frames the head was detected as elliptical hypotheses. The selection of the correct hypothesis as head is ambiguous or fails, if hands are occluding the head or are in regions close to the head.

	total	showing		head detected as ell. hyp.		head selection from ell. hyp.			
				yes	no	correct without amb.	correct with amb.	wrong	not done
Sequ 1	50	head	46	46		37	1	3	5
		no head	4						4
Sequ 2	50	head	46	41	5	13	18	7	3
		no head	4						4

As demonstrated in Fig. 2 the head is found robustly as ellipsoid hypothesis. Two errors occur in frames 4 and 7, due to the facts that the surface reconstruction is not perfect in small parts of the image (frame 4) and the curvature constraint is violated (frame 7). However seven frames show exact head ellipse detection from which the distance to the camera can be calculated correctly.

The head is not detected in the following cases: (i) The head is not fully within the dot projection area or rather close to the margin of that area.(ii) The head does not separate enough from its background. This is only the case if the background is concave in the way that the head fits well into it and forms a rather planar surface. Other possible errors occur, if close objects influence the surface in the way that it does not separate the head. This can occur i.e. if hands are close to the head. Objects are detected as heads if they form surfaces that are similar to that of heads. These cases cannot be detected from the 3D information of the region alone.

5. Summary and Conclusions

We have presented a fast method based on a simple structured light system that performs head hypotheses detection in situations, where methods using shape or appearance fail. Using sparse 3D data up to 100% location detection and up to 80% correct hypothesis selection is possible. Ambiguities occur only in very specific cases (see section 4). To avoid these situations, the following improvements are proposed:

- using model information: in this work the underlying head model is very simple. It assumes that the head is an ellipsoid of a certain size, defined by all possible human head sizes. If specific properties of the person are known, this could lead to a more exact description of the person's head. Additionally information about the body size and position can be included to remove wrong hypothesis.
- using sequence information (tracking): as each frame is treated independently, we expect significant improvements, if frame-to-frame dependencies are considered.
- using 2D (gray value) information: each hypothesis can be tested by applying facial feature detection algorithms to the 2D part of the image that corresponds to the detected head area. If facial features are found, the hypothesis is confirmed.

The 3D range information is the main information that is used by our head detection method. As this is not limited to 3D information derived from a structured light

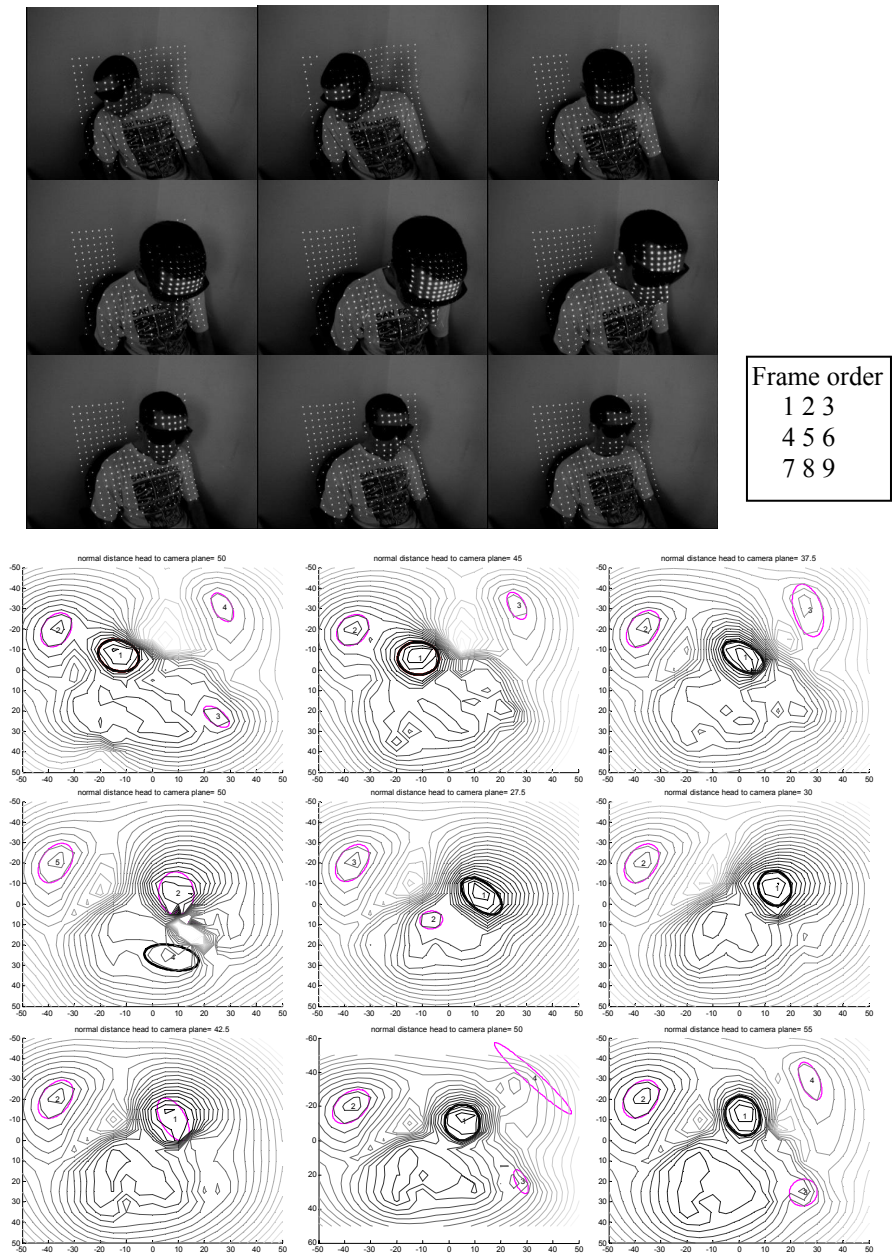

Fig.2.: Nine consecutive frames of a sequence showing a moving person. Upper part: control images. Lower part: contour plot of the reconstruction surface. The ellipse hypotheses and the selected head hypotheses are shown. The estimated distance to the camera is given. Seven (nr.1,2,3,5,6,8,9) show correct head detection, nr.4 shows a wrong head detection due to bad reconstruction in the area between ellipse 2 and 4, nr. 7 shows a non-detection due to the fact that ellipse 1 violates the curvature constraint.

system, other systems such as stereo or multiple camera views can be used alternatively. The algorithm of finding the suitable ellipsoids in the 3D view can be applied to any 3D range data that can be described as a function arising from the image plane.

References

1. Blake A., McCowen D., Lo H. R., Lindsey P. J.: Trinoculuar active range-sensing. IEEE Trans. on Pattern Analysis and Machine Intelligence 15, (1993) 477-483.
2. Clabian M., Rötzer H., Bischof H.: Tracking structured light pattern. SPIE – Conf. Intelligent Robots and Computer Vision XX: Algorithms, Techniques and Active Vision (2001) 183-192.
3. Gavrila D. M., Davis L.S.: 3-D model-based tracking of human upper body movement: a multi-view approach. (1995) 253-258.
4. Grammalidis N., Strintzis M.G.: Head detection and tracking by 2-D and 3-D ellipsoid fitting. IEEE Proc. Int. Conf. Computer Graphics (2000) 221-226.
5. Heikkilä J., Silven O.: A four-step camera calibration procedure with implicit image correction. Proc. of Int. Conf. on Computer Vision and Pattern Recognition (1997) 1106-1112.
6. Iwasawa S., Ohya J., Takahashi K., Sakaguchi T., Kawato S., Ebihara K., Morishima S.: Real-time, 3D-estimation of human body postures from trinocular images. Proc. Int. Workshop on Modelling People (1999) 3-10.
7. Jarvis R.: A perspective on range finding techniques for computer vision, IEEE Trans. on Pattern Analysis and Machine Intelligence 5 (1983) 122-139
8. Luo R., Guo Y.: Tracking of moving heads in cluttered scenes from stereo vision. Robot Vision (2001) 148-156.
9. Le Moigne J., Waxman, A.M.: Structured light patterns for robot mobility. IEEE J. of Robotics and Automation 4 (1988) 541-548.
10. Papageorgiou C., Poggio T.: A trainable system for object detection. Int. J. of Computer Vision 38(1) (2000) 15-33.
11. Reyna R., Giralt A., Esteve D.: Head detection inside vehicles with a modified SVM for safer airbags. IEEE Proc. Int. Transp. Systems (2001) 268-272.
12. Sansoni G., Carocci M., Rodella R.: Calibration and performance evaluation of a 3-D imaging sensor based on the projection of structured light. IEEE Trans. on Instrumentation and Measurement 49 (2000) 628-636.
13. Salvi J., Mouaddib E., Batlle J.: An overview of the advantages and constraints of coded pattern projection techniques for autonomous navigation. Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems , (1997) 1264-1271.
14. Stockman G. C., Chen S.-W., Gongzhu H., Shrikhande N.: Sensing and recognition of rigid objects using structured light. IEEE Control System Magazine 8 (1988) 14-22
15. Trobina M., Leonardis A.: An application of a structured light sensor system to robotics: Grasping arbitrarily shaped 3-D objects from a pile. Proc. IEEE Int. Conf. on Robotics and Automation 1 (1995) 241-246.
16. Zhang L., Lenders P.: A New Head Detection Method Based on the Region Shield Segmentation in Complex Background. Proc. Int. Symp. on Intell. Multimedia, Video and Speech Processing (2001) 328-331.
17. Zhang Z.: Flexible camera calibration by viewing a plane from unknown orientations. Proc. of 7th IEEE Int. Conf. on Computer Vision 1 (1999) 666-673.