Robust Correspondenceless 3-D Iris Location for Immersive Environments

Emanuele Trucco, Tom Anderson, Marco Razeto, and Spela Ivekovic

EECE-EPS, Heriot Watt University,
Riccarton, Edinburgh, EH14 4AS, UK
E.Trucco@hw.ac.uk

Abstract. We present a system locating the contour of an iris in space using robust active ellipse search and correspondenceless stereo. Robust iris location is the basis for gaze estimation and tracking, and, as such, an essential module for augmented and virtual reality environments. The system implements a robust active ellipse search based on a multi-scale contour detection model. The search is carried out by a simulated annealing algorithm, guaranteeing excellent performance in spite of heavy occlusions due to blinking, uncontrolled lighting, erratic target motion, and reflections of unpredictable scene elements. Stereo correspondence is avoided altogether by intersecting conjugate epipolar lines with the located ellipses. Experiments on synthetic and real images indicate very good performance of both location and reconstruction modules.

1 Introduction and Motivation

We present a system locating the contour of an iris in space using robust active ellipse search and correspondenceless stereo. Robust iris location is the basis for gaze estimation and tracking, and, as such, an essential module for augmented and virtual reality environments.

Context: immersive videoconferencing. The specific context is our work on the applications of computer vision to immersive videoconferencing [6,7,8,18]. Briefly, a station such as the one in Figure 1 (left) displays real-time, real-size videos of the two remote participants around a virtual table. Videos are acquired by four cameras surrounding the respective, remote plasma screens.

In order to create a visual impression of presence, the remote participants must appear as sitting around the virtual table and must be displayed from the local participant's viewpoint. To this purpose, we warp the incoming video by view synthesis, i.e., we synthesize the correct-viewpoint images [8,18]. This requires two components: real-time viewpoint tracking [19] and dense, accurate stereo disparity maps [6,8]. The latter are hard to achieve given the frequent occlusions created by arm movements, and the wide-baseline stereo geometry typical of immersive VC systems. Interpolation and model-based schemes to produce viable disparity maps have been reported, e.g., in [2,6,7].

Tracking the point between the observer's eyes is sufficient to provide the user's instantaneous viewpoint. However gaze, i.e., the direction in which one is looking with respect to the scene, is important for several purposes, including eye contact [3], metadata analysis (e.g., frequency of eye contact with other participants) and affective computing [9].

Monocular gaze is determined by the orientation of the eyeball in space, which is in turn given by the 3-D plane containing the iris contour. This paper concentrates on the problem of locating this plane robustly and accurately.

Fig. 1. Left: an immersive videoconferencing session. Right: active ellipse located around correct iris contour, and segments used to sample intensities.

Related work. Both *invasive* and *non-invasive* iris and pupil location techniques have been reported. Invasive techniques involve the use of devices to be worn or applied, e.g., electrodes, contact lenses and even head-mounted photo-diodes or cameras [5]. Non-invasive techniques avoid such solutions but often rely on structured illumination, e.g., Purkinje reflections [5,16,11]. Neither invasive devices nor structured illumination are admissible in our case. We choose not to restrict or control illumination, image quality and iris appearance, which precludes access to well-established techniques for people identification [10,4] relying on well-visible contours or limited eyelid occlusion.

Within immersive videoconferencing, studies have been reported, among others, on eye contact using stereo [3], eye tracking within a 6-camera setup [1] and Hausdorff tracking [19].

The location of the iris contour in space is linked to the problem of locating a conic in space from two perspective projections; closed-form solutions have been reported in [15]. Here, we prefer to exploit a simple model-based constraint to avoid completely stereo correspondence, and reconstruct the iris accurately by calibrated triangulation.

About this paper. In the remainder of this paper, Section 2 sketches the key technical challenges and summarizes our assumptions; Section 3 describes the robust iris detection based on active ellipses, Section 4 describes briefly the correspondenceless stereo module, Section 5 summarizes our experimental assessment of the system, and Section 6 offers some conclusions.

2   Assumptions and Challenges

We intend to estimate the normal to the iris plane and the 3-D iris location in space robustly, repeatably, and without restrictive assumptions or invasive gaze tracking equipment. We assume a stereo pair of cameras imaging a single eye, a setup not atypical in medical environments, biometrics and security. We do not assume special illumination and work with normal room lighting. We do not restrict the position of the iris in the image, nor require that the iris is completely or mostly visible, as assumed in [4].

The challenges are several and not insignificant. We face potentially extensive occlusions by eyelids or eyelashes, regular disappearance of the target due to blinking, frequent erratic target motion, and uncontrolled reflections of unpredictable scene elements and lights (see figures in Section 5). Our solution consists of two modules: robust location of the iris contour (limbus) in each image via active ellipse search, followed by correspondenceless stereo reconstruction of the iris in space. We describe each module in turn.

## 3   Robust Limbus Detection via Active Ellipses

The input is a monochrome image of a single eye; the output is an ellipse tracing the contour of the iris, illustrated in Figure 2. We suppress corneal reflections and other artefacts introducing distracting, strong contours, with a 10×10 median filter.

Modelling the iris contour. We find the limbus via an optimization in the parameter space of an active ellipse model. The unoccluded portion of the limbus is characterized by a noisy bright (sclera) to dark (iris) intensity transition of varying extent (3 to 12 pixels approximately in our application). We model this transition with two Petrou-Kittler ramp edges [12] at two different spatial scales.

The ellipse is parametrized by its semiaxes, a, b, and centre co-ordinates, O_x, O_y. The axes are assumed aligned with the image axes, as tilt is generally negligible. The cost function extracts intensity profiles along 30 normals to the candidate ellipse, distributed uniformly, as shown in Figure 1 (right) for the correct ellipse. These profiles are, ideally, convolved with two optimal ramp detection filter masks [12] at two different spatial scales. In practice, we are interested only in the filter output at the centre of the normal segments (i.e., at a control point on the ellipse perimeter), so we compute only *one* filtered value per segment. Filtered values are summed over all normals and over both filter sizes to obtain the criterion to optimize, c:

$$c = -\sum_{i=1}^{N} (\int_{-w}^{w} S_i(x) f_1(x) dx + \int_{-w}^{w} S_i(x) f_2(x) dx), \qquad (1)$$

where N is the number of control points, S_i is the intensity profile extracted at the control point i, f_1 and f_2 are the filters at the two different scales, and w is the filter's half-width.
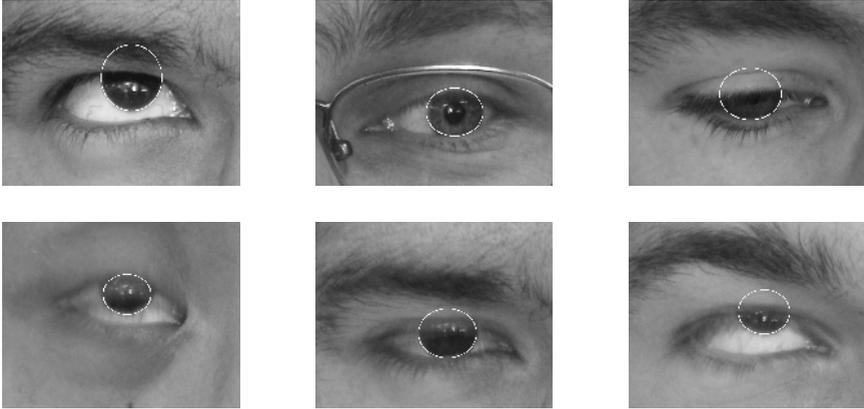
Fig. 2. Examples of iris detection results

Extensive testing identified masks optimal for ramps of width 4 and 10 pixels as responding well to limbus edges in our target images, and poorly to most transitions related to non-iris features (e.g., eyebrows, eyelashes).

Optimization scheme. Deterministic search proved inadequate for our problem, so we analysed various non-deterministic optimizers. We considered standard simulated annealing (henceforth) SA, two SA variations (great deluge, thresholded annealing), and the Girosi-Caprile stochastic optimizer, all reviewed in [14]. We recorded estimation errors in the four ellipse parameters over extensive ranges of variation of the algorithms' parameters, taking care to keep algorithms working in comparable conditions. This work is detailed in [14]. The result indicated standard SA as marginal winner over thresholded annealing.

For reasons of space we can only sketch the SA module. We refer the reader to [13] for details of our implementation, and to Salamon et al. [17] for a full treatment of SA and its practicalities. The active ellipse (i.e., the state vector (a, b, O_x, O_y)) is initialised at the image centre with a default size. The number of ellipses tested is progressively reduced with temperature, from $T_{start} = 500$ to $T_{end} = 1$. These temperature values were decided by sampling the cost function over several images and calculating the relative acceptance ratio, whose desirable value at high temperature is around 50%.

New candidate values for each parameter are generated from a Gaussian distribution centered in the previous value, with standard deviation $\sigma_{new} = R\sigma_{old}$, where R controls the search range, starting from 2 pixels for ellipse centre and 1 pixel for axes lengths and decreasing with an independent annealing schedule. The acceptance rule for new states is the standard Metropolis rule. The annealing schedule affects the move class via the range parameter R:

$$R_{new} = (\frac{1}{\sqrt{t+1}} + 0.3)R_{old}$$

where t is the annealing iteration index (time).
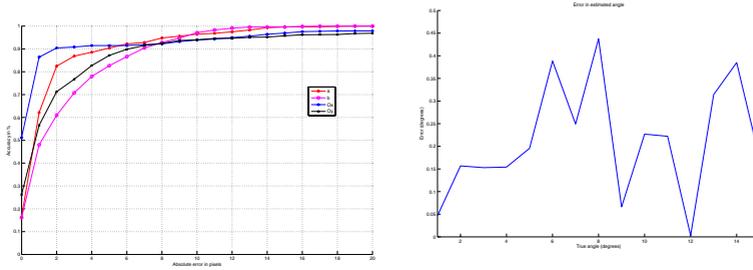
Fig. 3. Top: Monte-Carlo estimates of error probability of absolute error (pixels) for the four ellipse parameters. X axis: absolute error in pixels. Y axis: relative frequencies (probability estimates). Bottom: observed angular difference (degrees) between projections onto the XZ (ground) plane of estimated and true normals to the iris plane.

4 Correspondenceless Stereo

We find corresponding ellipse points without any search by locating the ellipse in both image, then intersecting the ellipses with conjugate epipolar lines. As a single ellipse is located in each image, no ambiguity exists. A circle (modelling the iris) is then fitted to the triangulated 3-D points.

We obtain the epipolar geometry from full calibration, but of course weak calibration (only image correspondences known) would suffice, at least for estimating the orientation of the iris plane in space. As the size of the human iris is very stable across individuals and even races [5], reasonable distance estimates could be achieved even with weakly calibrated cameras.

Figure 4 shows two pairs of images (with no occlusion for clarity), the detected irises, and the bundles of conjugate epipolar lines used for correspondence. The epipolar bundle must be chosen so to guarantee accurate intersections, i.e., the epipolar lines must be as normal to the ellipse as possible at the intersection points. We choose 20 points on the left-image ellipse avoiding the top and bottom arcs, where epipolar lines may approach the ellipse tangent. The points are spaced by $10°$ intervals along the ellipse, and grouped in two sets symmetric with respect to the vertical ellipse axis.

The 3-D plane best fitting the reconstructed points is found by linear least squares via singular value decomposition. Robust fitting is unnecessary as surely no outliers are present: correspondences are drawn from pre-fitted parametric curves.

5 Experimental Results

Iris detection accuracy. To test the accuracy of iris detection, we used a database of 327 monochrome test images with varying iris occlusion and blur, gaze directions, skin colours and eye shapes, and with and without spectacles. The images were 350×270, captured by a digital camera or camcorder with

uncontrolled room lighting. Ground truth was established manually by tracing ellipses following the limbus in each image. We performed 50 runs on each image ($50 \times 327 = 16,350$ runs). The ellipse is initialised always at the image centre, with semiaxes of 40 pixels each (the initial position is immaterial for SA). We computed the difference between estimates of ellipse parameters and the corresponding ground truth values. Examples of detections are shown in Figure 2.

Figure 3 (top) summarizes our analysis, showing, for each ellipse parameter, Monte-Carlo estimates of the cumulative probability of a given error value in pixels (relative frequencies). The graph is obtained by integrating the error histograms plotted for each parameter. For instance, 91.5% of the O_x histogram falls within a 5-pixel tolerance interval, suggesting an *indicative* probability of 91.5% for this accuracy level of the horizontal component. For O_y, this figure is 88%, due to frequent eyelid occlusion.

Correspondenceless stereo accuracy. All stereo tests were run with a MAT-LAB implementation on a Pentium III PC under Windows. Monochrome, PAL-resolution stereo pairs were acquired with PULNIX PEC3010 cameras and a Matrox Meteor II frame grabber. The stereo pair was calibrated using Tsai's classic procedure [20].

Controlled tests. To establish quantitative ground truth for the iris plane, we fixed a picture of a real iris onto a planar support. The support was rotated through an interval of 15 degrees around a vertical axis in steps of 1 degree. The interval was centered around the head-on direction (iris normal along the Z axis, pointing towards the cameras). For each angle, we estimated the orientation of the iris plane. The cameras were calibrated so that the axis of rotation was the X axis of the world reference frame, allowing consistent comparisons of estimates and ground truth. The interocular distance was 90mm, the focal lengths 12.48 and 10.74mm, and the stand-off distance (from left camera) about 200mm.

Figure 3 (bottom) shows, for each orientation, the angular error on the XZ (ground) plane, defined as the angular difference in degrees between the normals to true and estimated iris plane after projection on the XZ plane. The mean is $0.21°$, the standard deviation $0.13°$, both below the accuracy with which we could measure ground truth quantities. The full error, i.e., the angular difference between full (not XZ-projected) normals, is larger, in part because our manual positioning system did not guarantee repeatable orientations nor perfectly vertical iris planes, in part because estimated normals did include a small Y component. The mean of the full error was $1.5°$ and the standard deviation $0.4°$, ostensibly still very good results.

Real-eye tests. The above camera setup was used to acquire 20 stereo pairs of real eyes. Examples of images with a superimposed bundle of epipolar lines intersecting the detected ellipse are shown in Figure 4, together with the ellipse arcs fitted in 3-D space. The mean deviation from best-fit planes was 0.1mm, with average standard deviation 0.13mm, and maximum deviation of less than 1mm, suggesting accurate planar reconstruction. We could not measure the accuracy of absolute orientation, for which we rely on the controlled tests.
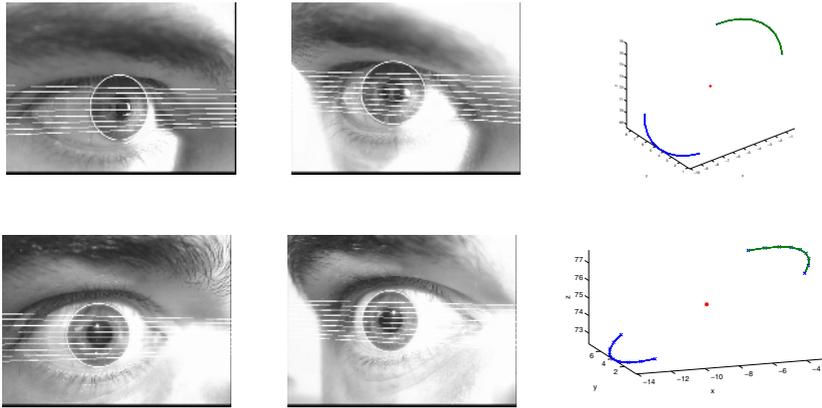
Fig. 4. Located limbus, bundles of conjugate epipolar lines and 3-D reconstruction of 3-D circle arcs for two stereo pairs (one per row)

6   Conclusions

Gaze estimation is an important capability for immersive and collaborative environments, and a crucial component of gaze estimation is 3-D iris location. We have presented a system performing this task reliably. The system implements a robust active ellipse search based on a multi-scale contour detection model. The search is carried out by a simulated annealing algorithm, guaranteeing excellent performance in spite of frequent occlusions due to blinking, uncontrolled lighting, erratic target motion, and reflections of unpredictable scene elements. Stereo correspondence is avoided altogether by intersecting conjugate epipolar lines with the located ellipses. Experiments indicate very good performance of both location and reconstruction modules. Current work is addressing the integration of stereo constraints in the ellipse location.

References

1. Baker, H.H, Malzbender, T., Bhatti, N., Tanguay, D., Sobel, I, Gelb, D., Goss, M., MacCormick, J., Kuasa, K. Cultbertson, W.B.: Computation and performance issues in Coliseum, an immersive videoconferencing system, Proc. 11th ACM Int. Conf. Multimedia (2003), 470–479.
2. Chang, N.L., Zakhor, A.: Constructing a multivalued representation for view synthesis, Int. Journ. of Comp. Vis., **45**:2 (2001), 147–190.

3. Criminisi, A., Shotton, J, Blake, A., Torr, P.: Gaze manipulation for one-to-one teleconferencing, Proc. IEEE Int. Conf. on Comp. Vis., Nice, France, (2003).
4. Daugman, J.: How iris recognition works, IEEE Trans. Circ. Sys. Vid. Tech., **14**:1 (2004).
5. Duchowsky, A.T.: Eye tracking methodology, Springer Verlag (2003).
6. Ivekovic, S., Trucco, E.: Dense wide-baseline disparities from conventional stereo for immersive videoconferencing, Proc. IAPR Int. Conf. Patt. Rec., Cambridge, UK (2004)
7. Isgro', F., Trucco, E., Kauff, P., Schreer, O.: 3-D image processing in the future of immersive media, IEEE Trans. Circ. Sys. Vid. Tech., special issue on immersive telecommunications **14**:3 (2004), 288–303
8. Isgro', F., Trucco, E., Xu, L-Q.: Towards teleconferencing by view synthesis and large-baseline stereo, Proc. IEEE/IAPR Int. Conf. Image Anal. and Processing, Palermo, Italy (2001), 198–203.
9. Kepoor, R., Qi, Y, Picard, R.W.: Fully automatic upper facial action recognition, Int. Wksp. on Analysis and Modelling of Faces and Gestures, IEEE Int. Conf. on Comp. Vis., Nice, France, (2003).
10. Ma, L., Tan, T., Wang, Y., Zhang, D.: Personal identification based on iris texture analysis, IEEE Trans. Patt. Anal. Mach. Intell., **25**:12, (2003).
11. Morimoto, C.H., Koons, D., Amir, A., Flickner, M.: Pupil Detection and Tracking Using Multiple Light Sources, Image and Vision Computing, **18**:4 (2000), 331–335.
12. Petrou, M., Kittler, J.: Optimal edge Detectors for Ramp Edges, IEEE Trans. PAtt. Anal. Mach. Intell., **13**:5 (1991).
13. Razeto, M, Trucco, E.: Robust iris location in close-up images of the eye, Pattern Analysis and Application, to appear, 2005.
14. Richard, S.: Comparative experimental assessment of four optimisation algorithms applied to iris location, MSc Thesis, EECE-EPS, Heriot Watt University, 2004.
15. Schmid, C. Zisserman, A.: The geometry and matching of lines and curves over multiple views, Int. Journ. Comp. Vis. **40**:3 (2000), 199–233.
16. S-W Shih and J Liu, A novel approach to 3-D gaze tracking using stereo cameras, IEEE Trans. Sys. Man Cybern. (Part B), **34**:1 (2004), 234–245
17. Salamon, P., Sibani, P., Frost, R.: Facts, Conjectures and Improvements for Simulated Annealing, SIAM, 2002.
18. Trucco, E., Plakas, K., Brandenburg, N., Kauff, P., Karl, M., Schreer, O.: Real-time disparity analysis for immersive 3-D videoconferencing, Proc. IEEE ICCV Workshop on Video Registration, Vancouver, Canada (2001).
19. Trucco, E., Razeto, M.: Hausdorff iconic matching with applications to 3-D video-conferencing, Proc. WIAMIS, London, UK (2003), 417–422.
20. Trucco, E., Verri, A.: Introductory Techniques for 3-D Computer Vision, Prentice-Hall, 1998.