2009

Face recognition using a time-of-flight camera

Simon Meers
University of Wollongong, meers@uow.edu.au

Koren Ward
University of Wollongong, koren@uow.edu.au

Face recognition using a time-of-flight camera

Abstract

This paper presents a novel three-dimensional (3D) method for detecting, tracking and recognising human faces using a time-of-flight camera. The system works by detecting a single central feature point, typically the nose tip, and by intersecting the 3D point data with spheres centred at the central feature point. The resulting spherical intersection profiles are used to perform face recognition and to track the position and orientation of the face. The main benefit of this method is that it is fast and efficient in terms or memory and computational expense. Furthermore, as the system utilises a time-of-flight camera and topographical information, it is not affected by variations in illumination, face orientation or partial occlusion of facial features. Experimental results are provided which show the potential of this method to exceed the real-time performance of existing head-pose tracking and face recognition systems.

Keywords

era2015, Face recognition, time-of-flight, SwissRanger

Disciplines

Physical Sciences and Mathematics

Publication Details

This conference paper was originally published as Meers, S and Ward, K, Face recognition using a time-of-flight camera, Proceedings of the 2009 Sixth International Conference on Computer Graphics, Imaging and Visualization. CGIV '09, Tianjin, China, 11-14 August 2009. Copyright The Institute of Electrical and Electronics Engineers, Inc. Original conference information available here

Face Recognition using a Time-of-Flight Camera

Simon Meers, Koren Ward
University of Wollongong, Australia
meers@uow.edu.au, koren@uow.edu.au

Abstract

This paper presents a novel three-dimensional (3D) method for detecting, tracking and recognising human faces using a time-of-flight camera. The system works by detecting a single central feature point, typically the nose tip, and by intersecting the 3D point data with spheres centred at the central feature point. The resulting spherical intersection profiles are used to perform face recognition and to track the position and orientation of the face. The main benefit of this method is that it is fast and efficient in terms or memory and computational expense. Furthermore, as the system utilises a time-of-flight camera and topographical information, it is not affected by variations in illumination, face orientation or partial occlusion of facial features. Experimental results are provided which show the potential of this method to exceed the real-time performance of existing head-pose tracking and face recognition systems.

1 Introduction

Systems that can recognise humans and transparently interact with them are finding increasing application in various fields. Face tracking and face recognition are key elements in natural human-computer interaction and have received considerable attention recently (e.g. [19, 7, 8, 17]). However, most existing face tracking and recognition systems are computationally expensive and susceptible to inaccuracies caused by variations in illumination, face orientation and partial occlusion of facial features. In this paper we present a three-dimensional (3D) method for detecting, tracking and recognising human faces in real-time using a time-of-flight camera. The main benefit of our system is that it is fast and efficient in terms of computational expense and memory usage. As the system utilises a time-of-flight camera and topographical information, it is not affected by variations in illumination, face orientation and partial occlusion of facial features.

In the following section we provide a brief summary of existing face recognition and tracking systems together with an overview of our system. This is followed by details of how we efficiently obtain a 3D 'faceprint' that can be used for face recognition and face tracking purposes. Experimental results are provided which prove the validity of our method and its potential to provide fast and efficient face tracking and face recognition in real-time.

2 Background

Due to the availability of the technology, face recognition research has predominantly centred around two-dimensional (2D) image processing, with a focus on facial feature detection. However, even the human brain can have difficulty recognising faces with varying orientation and illumination using 2D data alone [12]. The use of 3D data can eliminate viewpoint- and illumination-based limitations. Furthermore, 2D face recognition systems are unable to accurately determine the physical dimensions, location and orientation of the face relative to the sensor, whereas 3D sensors naturally integrate this information.

Research into face recognition using 3D sensors began in the 1980s. For example, Cartoux et al. [4] pioneered research into face profile matching using a laser range scanner. Although laser scanners are still used by many systems, they have the disadvantage that the user is required to remain motionless whilst the scan is completed. This limits the utility of these sensors in most practical situations. Stereo cameras are also widely used, and have the advantages of real-time data capture and integrated textural (2D) information. However, stereo camera systems require precise calibration, and rely upon surface features to determine depth. Consequently, surface regions possessing insufficient visual texture produce 'holes' in the depth map, which reduces the accuracy of the sensor. Some researchers overcome these issues by projecting structured light patterns onto the subject (e.g. [2]), however such solutions are too obtrusive and cumbersome for real-world applications.

Time-of-flight cameras are a recent innovation which provide accurate 3D data in real-time from a single, compact solid-state camera. They are not affected by illumination or textureless regions, and provide integrated 2D amplitude and 3D range images [15]. Recently, researchers have developed gesture recognition systems which utilise time-of-flight cameras [13, 3]. Some effort has also been applied to head detection [9], nose detection [11] and head-

pose tracking [14]. However, no work on face recognition using time-of-flight cameras has been published at this point in time as far as we are aware.

Our research has utilised a Mesa Imaging SwissRanger SR3000 time-of-flight camera [15] (see Figure 1). The array of infrared LEDs shown on the front of the unit illuminates the scene with frequency-modulated light which is used for making time-of-flight measurements. The SR3000 captures simultaneous infrared amplitude (greyscale) and depth images at QCIF (176x144) resolution.

Despite the obvious advantages of time-of-flight cameras, they are not without limitations. When capturing range and image data in a stationary environment they perform well, providing adequate steps are taken to reduce any noise in the range data [5]. But any rapid motion in the environment causes significant noise problems. This is due to the fact that the SR3000 requires four samples per pixel to calculate the depth based on the phase, offset and amplitude of the incoming signal, but the hardware is only capable of storing one at a time [16]. Thus the four samples must be acquired consecutively for each frame captured, and therefore any motion during this process will cause noise artefacts. '4-tap' sensors capable of acquiring the four samples in parallel are currently available, and it is anticipated that these will be utilised in future models, thereby eliminating this problem.

Figure 1: SwissRanger SR3000 time-of-flight camera

3 Face Recognition with Spherical Intersection Profiles

Our system works by detecting a single central feature point, (the nose tip), and by intersecting the 3D data with spheres centred at this point. The tip of the nose is ideal for this purpose because it is the only facial feature which can be (almost) guaranteed to be visible to an observer, and not obscured by glasses, facial hair, orientation, etc. The nose tip is also an important feature because of its central position within the face, and relative ease of detection using both amplitude- and depth-based approaches [10, 11]. However detecting the nose tip alone is not sufficient to determine the 3D orientation of the face, since at least three points are required. To solve this problem without requiring the detection of additional features, we introduced the concept of *spherical intersection profiles*: paths of intersection of the facial surface with spheres centred at the detected feature (see Figure 2).

3.1 Spherical Intersection Profiles

The use of spherical intersection profiles for detection, tracking and recognition provides a number of advantages over alternative techniques. As each spherical intersection is comprised of 3D data points which have preset distance (radius) from the nose tip, they can be obtained with just one iteration through the 3D point data. It is therefore much faster than alternative rigid registration methods such as Iterative Closest Point (ICP) [1] used in other systems (e.g. [6]), and does not suffer from convergence problems. It also retains spatial information unlike approaches which utilise transformations such as Extended Gaussian Images (e.g. [18]).

3.2 Preprocessing

The raw data from the SR3000 is preprocessed to reduce noise as much as possible and to isolate the 3D points comprising the subject's head. Noise reduction is performed by median filtering. Erroneous data is also eliminated by applying correlations between the depth image and the amplitude image. The frame is then reduced to the 3D region of interest by first applying distance thresholds to eliminate unwanted foreground and background data. The horizontal and vertical limits are then determined by detecting the top of the head, sampling its distance from the depth map and applying anthropometric data to determine its maximum possible dimensions at that distance.

3.3 Locating the nose tip

The nose tip is detected within the 3D data by searching for specific illumination and geometric characteristics. Given that the tip of the human nose is an approximately spherical surface, and that the illumination angle is known (i.e. direct infrared from the SR3000), likely nose tip features can be found in the amplitude image data by testing the data for spherical reflective features. Any false nose tip features detected in the illumination data are eliminated by using the depth image to analyse the differential landscape to determine if the candidate region has the appropriate curvature. This technique has proven to be robust and produce accurate results [14].

3.4 Performing Spherical Intersections

The spherical intersection profiles are obtained by traversing the 3D point data to find the intersection paths of spheres with the desired radii (see Figure 2). The traversal of the 3D point data is performed by interpolating between the points to obtain sub-pixel accuracy. Supersampling is also applied to reduce the effects of noise on the intersection profiles.

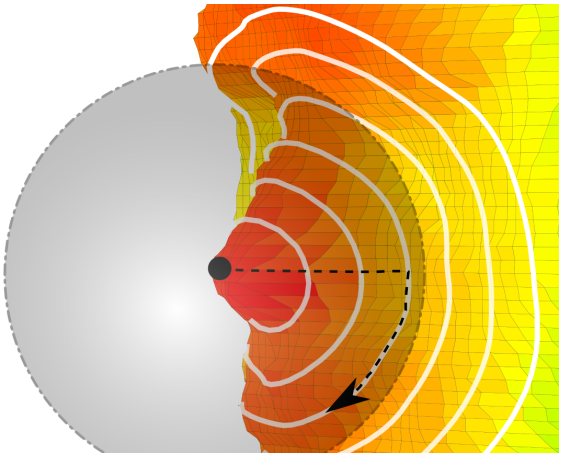

Figure 2: Illustration of tracing a spherical intersection profile starting from the nose tip

3.5 Orientation calculation

Once a set of spherical intersection profiles (faceprint) has been found in the 3D data, it must be normalised by a rotation transformation so that it is 'facing' a set direction (i.e. down the z-axis in our system). Incidentally, this transformation provides a measure of the head-pose or 'gaze' direction of the subject and can be used to implement a head-pose tracking system. The following subsections describe this process.

3.5.1 Initial Alignment

The most obvious method of calculating the orientation is to find the centroid of each spherical intersection profile, and then project a line through the centroids from the centre of the spheres using a least-squares fit. This provides a reasonable approximation in most cases but performs poorly when the face orientation occludes considerable regions of the spherical intersection profiles from the camera. A spherical intersection profile which is 40% occluded will produce a poor approximation of the true centroid and orientation vector using this method.

Instead, we find the average of the latitudinal extrema of each spherical intersection profile (i.e. the topmost and bottommost points). This proved effective over varying face orientations for two reasons. Firstly, these points are unlikely to be occluded due to head rotation. Secondly, in most applications the subject's head is unlikely to 'roll' much (as opposed to 'pan' and 'tilt'), so these points are likely to be the true vertical extremities of the face. If the head is rotated so far that these points are not visible on the spherical intersection profile, the system detects that the spherical intersection profile is still rising/falling at the upper/lower terminal points and therefore dismisses it as insufficient. A least-squares fit of a vector from the nose tip passing through the latitudinal extrema midpoints provides a good initial estimate of the face orientation. Several further optimisations are subsequently performed by utilising additional information, as discussed in the following sections.

3.5.2 Symmetry Optimisation

Given that human faces tend to be highly symmetric, the orientation of a faceprint can be optimised by detecting the plane of bilateral symmetry. Note that this does not require the subject's face to be perfectly symmetric in order to be recognised. This is performed by first transforming the faceprint using the technique described above, to produce good orientation estimation. Given the observation of limited roll discussed in Section 3.5.1, it is reasonable to assume that the orientation plane will intersect the faceprint approximately vertically. The symmetry of the faceprint is then measured using Algorithm 1. This algorithm returns a set of midpoints for each spherical intersection profile, which can be used to measure the symmetry of the current faceprint orientation. Algorithm 1 also provides an appropriate transformation which can be used to optimise the symmetry by performing a least-squares fit to align the plane of symmetry approximated by the midpoints. Figure 3 illustrates an example faceprint with symmetry midpoints visible.

3.6 Temporal Optimisation

Utilising data from more than one frame provides opportunities to increase the quality and accuracy of the faceprint tracking and recognition. This is achieved by maintaining a faceprint model over time, where each new frame contributes to the model by an amount weighted by the quality of the current frame compared to its predecessors. The quality of a frame is assessed using two parameters. Firstly, the noise in the data is measured during the median filtering process mentioned in Section 3.2. This is an important parameter, as frames captured during fast motions will be of substantially lower quality due to the sensor issues discussed in Section 2. In addition, the measure of symmetry of the faceprint (see Section 3.5.2) provides a good assessment of the quality of the frame. These pa-

Algorithm 1 Calculate the symmetry of a set of $SIPs$

```
midpoints ⇐ new Vector[length(SIPs)]
for s ⇐ 0 to length(SIPs) − 1 do
  for p ⇐ 0 to length(SIPs[s]) − 1 do
    other ⇐ ∅
    Find other point on SIPs[s] at same 'height' as p:
    for q ⇐ 0 to length(SIPs[s]) − 1 do
      next ⇐ q + 1
      if next >= length(SIPs[s] then
        next ⇐ 0 {Wrap}
      end if
      if q = p ∨ next = p then
        continue
      end if
      if (q.y < p.y < next.y) ∨ (q.y > p.y > next.y)
      then
        other ⇐ interpolate(q, next)
        break
      end if
    end for
    if other ≠ ∅ then
      midpoints[s].append((p + other)/2)
    end if
  end for
end for
return midpoints
```

rameters are combined to create an estimate of the overall quality of the frame using Equation 1.

$$quality = \sqrt{noiseFactor \times symmetryFactor} \quad (1)$$

Thus for each new frame captured, the accuracy of the stored faceprint model can be improved. Our system maintains a collection of the n highest quality faceprints captured up to the current frame, and calculates the optimal average faceprint based on the quality-weighted average of those n frames. This average faceprint quickly evolves as the camera captures the subject in real-time, and forms a robust and symmetric representation of the subject's 3D facial profile (see Figure 4).

Temporal information can also be gained by analysing the motion of the subject's head-pose over a sequence of frames. The direction, speed and acceleration of the motion can be analysed, and used to predict the nose position, head-pose and the resultant 'gaze' position for the next frame. This is particularly useful when the system is used as a 'gaze'-tracker as the system can utilise this data to smooth the 'gaze' path and further eliminate noise.

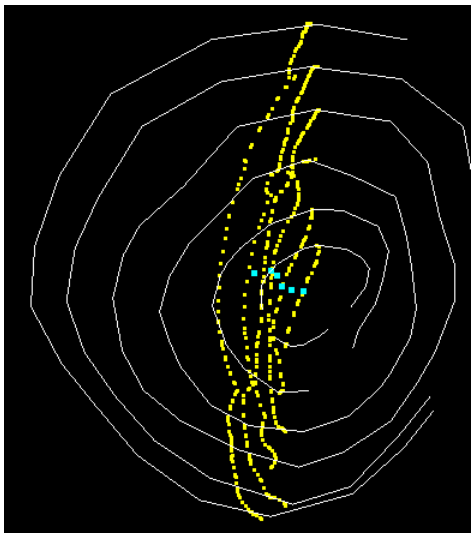

Figure 3: Example faceprint with symmetry midpoints (in yellow), and per-sphere midpoint averages (cyan)

3.7 Comparing Faceprints

Using the averaging technique discussed in the previous section, a single faceprint can be maintained for each subject. These are stored by the system and accessed when identifying a subject, which requires efficient and accurate comparison of faceprints. In order to standardise the comparison of faceprints and increase the efficiency, quantisation is performed prior to storage and comparison. Each spherical intersection profile is sampled at a set number of angular divisions, and the resultant interpolated points form the basis of the quantised faceprint. Thus each faceprint point has a direct relationship to the corresponding point in every other faceprint. Comparison of faceprints can then be simplified to the average Euclidean distance between corresponding point pairs.

4 Results

Figure 5 shows some actual 2D transforms of faceprints taken from different subjects. Although our system is yet to be tested with a large sample of subjects, our preliminary results suggest that faceprints derived from spherical intersection profiles vary considerably between individuals.

The number of spherical intersection profiles comprising faceprints, their respective radii and the number of quantised points in each spherical intersection profile are important parameters that define how a faceprint is stored and compared. Our experiments have shown that increasing the number of spherical intersection profiles beyond six and the number of quantised points in each spherical intersection profile beyond twenty did not significantly improve the system's ability to differentiate between faceprints. We believe this is due mainly to the continuous nature of

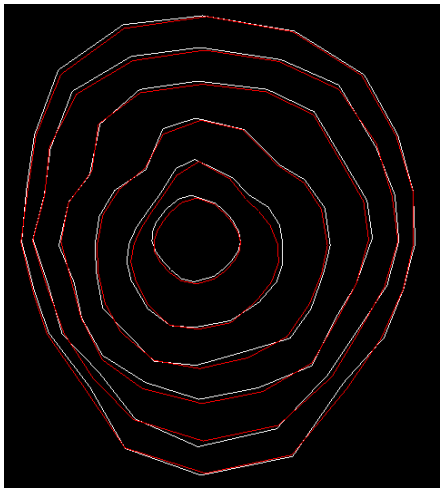

Figure 4: Faceprint from noisy sample frame (white) with running-average faceprint (red) superimposed

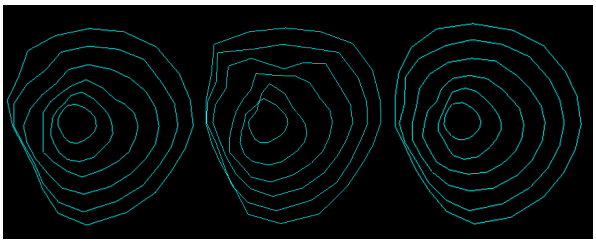

Figure 5: Example faceprints

the facial landscape and residue noise in the processed data. Consequently, our experiments were conducted with faceprints comprised of six spherical intersection profiles. Each spherical intersection profile was divided into twenty points. All faceprints were taken from ten human subjects positioned approximately one metre from the camera and at various angles.

Our experiments showed that 120-point faceprints provided accurate recognition of the test subjects from most view points. We found it takes an average of 6.7μs to process and compare two faceprints with an Intel dual-core 1.8GHz Centrino processor. This equates to faceprint comparison rate of almost 150,000 per second which clearly demonstrates the potential search speed of the system. The speed of the each procedure (running on the same processor) involved in capturing faceprints from the SR3000 camera is outlined in Table 1.

Our experiments also showed that tracking the head-pose (face orientation) of subjects was tracked relatively accurately. Figure 6 shows the 'gaze' of a subject being projected onto a 'Virtual Screen'. The head-pose tracking was found to be accurate to within approximately three to five degrees, even when significant portions of the faceprint were occluded by the camera angle.

Processing Stage	**Execution Time**
Distance Thresholding	210μs
Brightness Thresholding	148μs
Noise Filtering	6,506μs
Region of Interest	122μs
Locate Nose	17,385μs
Trace Profiles	1,201μs
Quantise Faceprint	6,330μs
Face Recognition	91μs
Update Average Faceprint	694μs
Total per frame	32,687μs

Table 1: Average execution times for each processing stage

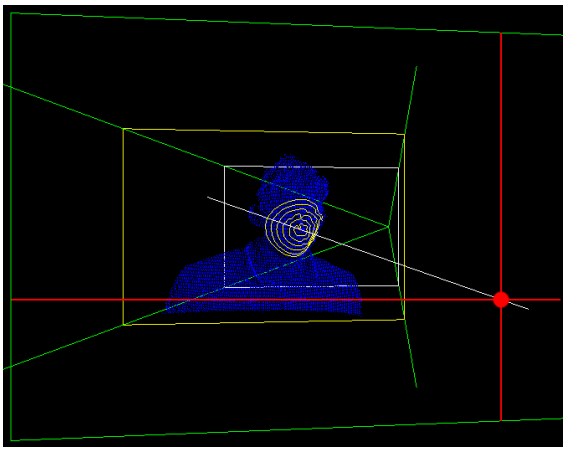

Figure 6: Example 'gaze' projected onto a 'Virtual Screen'

Conclusion

This paper presents the details and preliminary experimental results of a novel 3D method for detecting, tracking and recognising human faces by using a time-of-flight camera. Our experimental results show that 3D faceprints comprised of spherical intersection profiles are capable differentiating human faces with minimal processing and storage. Our results also show that faceprints can be captured quickly (at frame rates) and compared with minimal processing. Furthermore, the faceprint capturing procedure enables head-pose tracking to be performed in real-time with reasonable accuracy. Although our system is yet to be tested with a large sample of subjects, our preliminary results indicate that fast searching of large faceprint databases is achievable with reasonable accuracy using currently available time-of-flight cameras. Improvements to time-of-flight cameras, to reduce noise, would also reduce the processing required in our system and further improve accuracy.

References

[1] P. J. Besl and H. D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on pattern analysis and machine intelligence*, 14(2):239–256, 1992.

[2] C. Beumier and M. Acheroy. Automatic 3D face authentication. *Image and Vision Computing*, 18(4):315–321, 2000.

[3] P. Breuer, C. Eckes, and S. Muller. Hand gesture recognition with a novel IR time-of-flight range camera – a pilot study. *Lecture Notes in Computer Science*, 4418:247, 2007.

[4] J. Y. Cartoux, J. T. Lapreste, and M. Richetin. Face authentification or recognition by profile extraction from range images. pages 194–199, 1989.

[5] Derek Chan. Noise vs. Feature: Probabilistic Denoising of Time-of-Flight Range Data, 2008. `http://www.stanford.edu/class/cs229/proj2008/ChanD-ProbabilisticDenoisingOfRangeData.pdf`.

[6] J. Cook, V. Chandran, S. Sridharan, and C. Fookes. Face recognition from 3d data using iterative closest point algorithm and gaussian mixture models. In *3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings. 2nd International Symposium on*, pages 502–509, 2004.

[7] Kresimir Delac and Mislav Grgic, editors. *Face Recognition*. I-Tech Education and Publishing, July 2007.

[8] Kresimir Delac, Mislav Grgic, and Marian Stewart Bartlett, editors. *Recent Advances in Face Recognition*. I-Tech Education and Publishing, December 2008.

[9] S. B. Gokturk and C. Tomasi. 3D head tracking based on recognition and interpolation using a time-of-flight depth sensor. volume 2, pages 211–217, June 2004.

[10] Dmitry O. Gorodnichy. On importance of nose for face tracking. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 181–186, May 2002.

[11] M. Haker, M. Bohme, T. Martinetz, and E. Barth. Geometric invariants for facial feature tracking with 3D TOF cameras. *Signals, Circuits and Systems, 2007. ISSCS 2007. International Symposium on*, 1:1–4, July 2007.

[12] H. Hill, P. G. Schyns, and S. Akamatsu. Information and viewpoint dependence in face recognition. *Cognition*, 62(2):201–222, 1997.

[13] M. B. Holte, T. B. Moeslund, and P. Fihl. View invariant gesture recognition using the CSEM SwissRanger SR-2 camera. *International Journal of Intelligent Systems Technologies and Applications*, 5(3):295–303, 2008.

[14] Simon Meers and Koren Ward. Head-pose tracking with a time-of-flight camera. In *Australasian Conference on Robotics & Automation*, December 2008.

[15] MESA Imaging. SwissRanger SR3000 - miniature 3D time-of-flight range camera, 2006. `http://www.mesa-imaging.ch/prodview3k.php`.

[16] Thierry Oggier, Michael Lehmann, Rolf Kaufmann, Matthias Schweizer, Michael Richter, Peter Metzler, Graham Lang, Felix Lustenberger, and Nicolas Blanc. An all-solid-state optical range camera for 3d real-time imaging with sub-centimeter depth resolution (swissranger). volume 5249, pages 534–545. SPIE, 2004. `http://link.aip.org/link/?PSI/5249/534/1`.

[17] Julio Ponce and Adem Karahoca, editors. *State of the Art in Face Recognition*. I-Tech Education and Publishing, January 2009.

[18] H. T. Tanaka, M. Ikeda, and H. Chiaki. Curvature-based face surface recognition using spherical correlation. principal directions for curved object recognition. pages 372–377, 1998.

[19] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys (CSUR)*, 35(4):399–458, 2003.