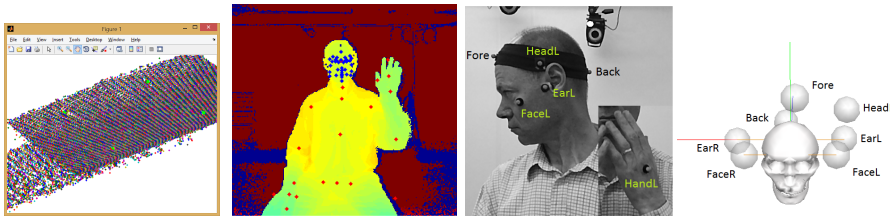
Supplementary Material

1. Sensor calibration



(a) (b) (c) (d)

Figure 1: Sensor calibration: (a) depth data from the calibration scene rotated to show the table surface; (b) Kinect body pose estimation during a wave (red dots); (c) the extra Vicon marker placed on participant's left hand (inlay); (d) creation of a head segment defined by the ear markers (proximal end), and face markers (distal end) using Visual 3D software (C-Motion, US).

1.1. Estimation of the spatial transformation

Kinect was placed on a tripod and angled to frame the upper body of participants when seated in the centre of the Vicon capture volume, see Fig. 1b. Vicon wand calibration was then performed with the Kinect powered on so that its active infrared light source could be masked. Four markers were then placed on the surface of a table in the centre of the capture volume, and imaged using both the Vicon system and the Kinect depth sensor.

A spatial transformation between the two sensors was estimated using a custom-written MATLAB (Mathworks, USA) graphical user interface (see Fig. 1a) that allowed the visualisation and clicking of Kinect depth data from the calibration scene. A datapoint on each of the four Vicon markers was manually

selected, and a rigid transformation estimated [1] between these four coordinates and the four Vicon marker estimates for the same scene.

1.2. Estimation of time synchronisation

During data capture each participant was asked to raise and then lower their left hand at the start of each recording. This allowed synchronisation in time between the two sensors by computing cross correlations between the height of the hand, as estimated by each sensor system. Skeletal tracking events were used to record the location of the left hand from the perspective of Kinect (see Fig. 1b), and an extra marker was attached for tracking by the Vicon system (see Fig. 1c). An offset was then calculated from the maximum of the cross-correlation of the two measurements, and the Vicon data downsampled (using a simple nearest neighbour interpolation) to provide a record of ground truth corresponding with every HDFT event, see Fig. 2 in the main paper.

1.3. Definition of a Vicon head segment

Head rotations were extracted from the Vicon marker data by using Visual 3D software (C-Motion, US) to create a virtual head segment defined by the two ear markers (proximal end), and two face markers (distal end), see Fig. 1d. Rotations of the segment coordinate system relative to the lab coordinate system were then exported as yaw, pitch and roll angles.

2. Subscribing to the HDFT stream

Listing 1 shows how the HDFT event handler from the HD Face Basics-WPF SDK example can be extended to write head rotations to disk. A `StreamWriter` can be used to open a text file in the `Window_Loaded` method before making calls to `WriteLineAsync` to record the variables in the listing (before closing the text file in `MainWindow_Closing`). The Face Basics-WPF example provides code for converting quaternions to yaw, pitch and roll rotations. If facial feature point locations are also required then the `for` loop in the `UpdateMesh` method

can be extended to additionally write the held in the `HighDetailFacePoints` enumeration.

Listing 1: Extending of the HDFaceBasics SDK example to retain head pose estimates.

```
/// <summary>
/// This event is fired when a new HDFace frame is ready for consumption
/// </summary>
/// <param name="sender">object sending the event</param>
/// <param name="e">event arguments</param>
private void HdFaceReader_FrameArrived(object sender,
        HighDefinitionFaceFrameArrivedEventArgs e)
{
        using (var frame = e.FrameReference.AcquireFrame())
        {
        // We might miss the chance to acquire the frame
        if (frame == null)
        {
                return;
        }

        // Also ignore this frame if face tracking failed.
        if (!frame.IsFaceTracked)
        {
                // START write to file
                // 1. the timestamp for this event:
                // frame.RelativeTime.TotalMilliseconds.ToString()
                // 2. the fact no rotation could be estimated:
                // "missing"
                // STOP write to file

                return;
        }

        frame.GetAndRefreshFaceAlignmentResult(this.currentFaceAlignment);

        // START write to file
        // 1. the timestamp for this event:
        // frame.RelativeTime.TotalMilliseconds.ToString()
        // 2. the estimated quaternion:
        // this.currentFaceAlignment.FaceOrientation.W
        // this.currentFaceAlignment.FaceOrientation.X
        // this.currentFaceAlignment.FaceOrientation.Y
        // this.currentFaceAlignment.FaceOrientation.Z
        // STOP write to file

        this.UpdateMesh();
        }
}
```

3. Movements

The following sections describe in detail each of the head movements tested in this study.

3.1. Static torso: range of motion tests

In the static torso conditions, participants were asked to keep their torso and shoulders still while completing each of the following tasks.

3.1.1. Up-down

Starting from their resting pose, participants were asked to:

1. look up as far as possible;
2. return to their resting pose;
3. look down as far as possible;
4. return to their resting pose. See also Fig. 2a.

3.1.2. Left-right

Starting from their resting pose, participants were asked to:

1. look left as far as possible;
2. return to their resting pose;
3. look right as far as possible;
4. return to their resting pose. See also Fig. 2b.

3.1.3. Side-to-side

Starting from their resting pose, participants were asked to:

1. tilt their head as far to the left as possible;
2. return to their resting pose;
3. tilt their head as far to the right as possible;
4. return to their resting pose. See also Fig. 2c.

3.1.4. 4-corners

Starting from their resting pose, participants were asked to:

1. look up to their top left as far as possible;
2. return to their resting pose;
3. look up to their top right as far as possible;

4. return to their resting pose;
5. look down to their bottom left as far as possible;
6. return to their resting pose;
7. look down to their bottom right as far as possible;
8. return to their resting pose. See also Fig. 2d.

3.2. Free torso: range of motion tests

Participants were asked to repeat the *up-down* (see Section 3.1.1), *left-right* (see Section 3.1.2) and *side-to-side* (see Section 3.1.3) movements with their torso and shoulders free to move. This had the effect of increasing their range of head motion, see Figs. 3a, 3b and 3c, respectively.

3.2.1. I-spy

Participants were asked to play a short game of "I spy with my little eye", in which they were challenged to find an object beginning with a particular letter only by looking around the room. In fact, there was no corresponding object, and the aim was simply to induce a wide but natural range of head poses as quickly as possible, see Fig. 3d for examples.

3.3. Occlusion

Participants were asked to repeat the static torso movements *up-down* (see Section 3.1.1) and *left-right* (see Section 3.1.2) using their hands to cover their mouth, see Figs. 4a and 4b, respectively. Fig. 4c shows the distribution of missed frames as a function of the angle measured by Vicon for these two movements. Compared with Fig. 3 in the main paper, missed frames are now spread across the whole range of motion. Tracked frames were concentrated at the start and end of the movements, while hands were raised to and then lowered from the face with the head held still at approximately zero rotation. This accounts for the small reduction in average angular errors for this condition (see Table 2 in the main paper).

3.4. Standing: range of motion tests

Participants were asked to repeat the static range of motion tests (see Section 3.1), whilst standing far enough from the sensor to enable full body pose estimation. Fig. 5 shows examples from each movement. Only the portion of the luminance image containing participants' heads was retained during capture, and so the full scene (including the body pose estimate) is shown from the perspective of the depth camera. Kinect joint estimates are plotted in cyan.

3.5. Calibrated: range of motion tests

Participants were asked to repeat the static range of motion tests (see Section 3.1) after performing interactive face shape calibration. Face shape calibration allows Kinect to learn the shape of a participant's face in order to increase the quality of facial feature tracking [2]. (Whether it improves head pose estimation is an open question.) The procedure is interactive and the participant must respond to requests from Kinect to:

- face to the left;
- face to the right;
- face forward and upwards.

Each of these requests may be repeated multiple times by the Kinect.

3.6. Rotated: range of motion tests

Participants were asked to repeat the static range of motion tests (see Section 3.1), with their chair rotated at a 45° angle to the Kinect. Fig. 6 shows examples from each movement. Although average errors were higher (see the main paper) the HDFT was able to return results reliably at the right-sided edge of range (large negative yaw rotations, e.g. Fig. 7).

3.7. Comparison with other approaches

Table 1 summarises results for two other approaches to RGB-D head pose estimation from the literature, both applied to the freely available Biwi Kinect Head Pose Database [3]. The Biwi dataset is very challenging, containing over 15K RGB-D images of 20 different participants striking a range of different head poses ($\pm 75°$ yaw and $\pm 60°$ pitch). Long hair is not tied back and a number of the participants are wearing glasses. In drawing any conclusions about performance the following points should be borne in mind:

1. The approach of Fanelli et al. [3] is a discriminative one, capable of processing RGB-D images in isolation (rather than recovering small inter-frame changes) and therefore does not suffer from the many gaps between recordings in the Biwi database.
2. The approach of Baltrušaitis [4] is a frame-to-frame tracker that expects small inter-frame changes, but to facilitate cross comparison the authors have nevertheless applied it to the Biwi database.

Table 1: Performance of two other approaches from the literature on the Biwi Kinect Head Pose Database [3].

	Nose (mm)	Yaw (°)	Pitch (°)	Roll (°)	Missed (%)
[3]	12.2 ± 22.8	3.8 ± 6.5	3.5 ± 5.8	5.4 ± 6.0	6.6
[4]	-	6.3	5.1	11.3	0

Although it is not possible to process pre-recorded RGB-D data with the HDFT algorithm for a direct comparison on the Biwi database, results for all $9,126$ frames in the static range of motion tests studied here (see Section 3.1) are given in table 2, with means and standard deviations computed across all frames rather than between participants' averages, following [3].

Table 2: Range of motion tests with mean and standard deviation over all $9,126$ frames.

	Cheekbone (mm)	Yaw (°)	Pitch (°)	Roll (°)	Missed (%)
Range of motion	9.7 ± 4.5	2.1 ± 2.1	7.1 ± 4.1	2.7 ± 4.0	10.8

References

[1] P. J. Besl, N. D. McKay, Method for registration of 3-d shapes, in: Robotics-DL tentative, International Society for Optics and Photonics, 1992, pp. 586–606.

[2] Kinect for Windows SDK 2.0, `https://msdn.microsoft.com/en-us/library/dn785525.aspx`, accessed: 2015-07-05.

[3] G. Fanelli, M. Dantone, J. Gall, A. Fossati, L. Van Gool, Random forests for real time 3D face analysis, International Journal of Computer Vision 101 (3) (2013) 437–458.

[4] T. Baltrušaitis, P. Robinson, L.-P. Morency, 3D constrained local model for rigid and non-rigid facial tracking, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 2610–2617.

(a) *Up-down*

(b) *Left-right*

(c) *Side-to-side*

(d) *4-corners*

Figure 2: Static torso: range of motion tests.

(a) *Up-down* with free torso.

(b) *Left-right* with free torso.

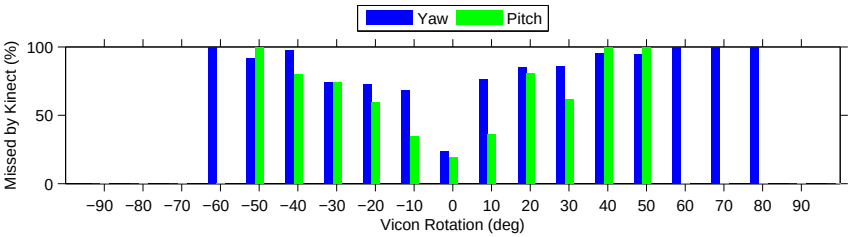
(c) *Side-to-side* with free torso.

(d) Example poses from a game of *I-spy*.

Figure 3: Free torso: range of motion tests.

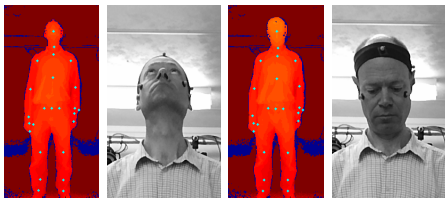
(a) *Up-down* with occlusion.

(b) *Left-right* with occlusion.

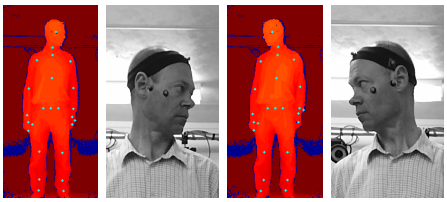


(c) Distribution of missed frames (see also Fig. 3 in main paper for comparison).

Figure 4: Movements with facial features occluded.



(a) *up-down*



(b) *left-right*

(c) *side-to-side*

(d) *4-corners* (top left, bottom right shown)

Figure 5: Example poses from the static range of motion tests performed standing. Kinect joint estimates are shown in cyan.

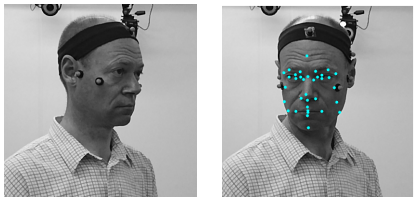
(a) *up-down*

(b) *left-right*

(c) *side-to-side*

(d) *4-corners* (top left, bottom right shown)

Figure 6: Example poses from the static range of motion tests performed with a 45° seating rotation.



(a) (b)

Figure 7: Rotated configuration: (a) participant sitting at rest; (b) successful pose estimation at full yaw rotation.