Title

Vision-based Body Tracking: Turning Kinect into a Clinical Tool

Authors

**Dr Cecily Morrison
Research Scientist
Microsoft Research
21 Station Rd
Cambridge, UK, CB4 1BS
cecilym@microsoft.com
+44 1223 479 951

Dr Peter Culmer
Senior Translational Research Fellow
School of Mechanical Engineering
University of Leeds
Leeds, UK, LS2 9JT
P.R.Culmer@leeds.ac.uk

Dr Helena Mentis
Assistant Professor
Department of Information Systems
University of Maryland, Baltimore County
Baltimore, MD 21250 USA
mentis@umbc.edu

Dr Tamar Pincus
Professor of Health Psychology
Department of Psychology
Royal Holloway University of London
Egham, UK, TW20 0EX
T.Pincus@rhul.ac.uk

Abstract

Purpose: Vision-based body tracking technologies, originally developed for the consumer gaming market, are being repurposed to form the core of a range of innovative healthcare applications in the clinical assessment and rehabilitation of movement ability. Vision-based body tracking has substantial potential, but there are technical limitations.

Method: We use our "stories from the field" to articulate the challenges and offer examples of how these can be overcome.

Results: We illustrate that: (1) substantial effort is needed to determine the measures and feedback vision-based body tracking should provide, accounting for the practicalities of the technology (e.g. range) as well as new environments (e.g. home). (2) Practical considerations are important when planning data capture so that data is analysable, whether finding ways to support a patient or ensuring everyone does the exercise in the same manner. (3) Home is a place of opportunity for vision-based body tracking, but what we do now in the clinic (e.g. balance tests) or in the home (e.g. play games) will require modifications to achieve capturable, clinically relevant measures.

Conclusions: This article articulates how vision-based body tracking works and when it does not to continue to inspire our clinical colleagues to imagine new applications.

VISION-BASED BODY TRACKING

Vision-based body tracking uses computer vision to track human motion in three dimensions. It relies on a depth sensing camera, such as the Kinect, to provide data which are then processed with advanced computer algorithms. Unlike predecessor technologies – gait tracking systems, inertial body sensors, and markered motion capture systems – vision-based body tracking is neither expensive nor cumbersome [1]. Indeed many people use this technology as part of their home entertainment systems every day.

As a reliable and low-cost way to track human motion, vision-based body tracking has the potential to dramatically change the way clinical assessment and rehabilitation of movement ability is carried out. It can, for example, be used to calculate an objective, quantified metric of movement ability to provide an alternative to standardised but subjective movement assessment rating scales widely used in clinical practice today. This would address common problems of subjective measures, such as low inter– and intra– rater reliability, poor sensitivity, or ceiling effects that can reduce the validity of clinical trials [2].

Available to the consumer market, vision-based body tracking could facilitate regular movement assessment at home. More frequent assessment could support reactive treatment regimes, or longitudinal clinical trial measure collection which would be too expensive to do in the clinic. Rehabilitation could also be supported, either synchronously or asynchronously, at home. This could decrease the substantial costs of face to face rehabilitation for providers and shorten long waiting lists and travel times for patients.

The gaming origins of vision-based body tracking (ie. xBox) could be utilised to motivate prescribed exercise regimes done at home through a serious gaming approach. Low uptake of home exercise regimes has traditionally been a problem [3], but current evidence suggests that off the shelf fitness games and virtual games (e.g. 'exergames') are acceptable and uptake is good [4]. Additionally, bespoke systems that address specific therapeutic exercises are also being developed supported by objective quantification of the physical demands needed for treatment [5].

As researchers working to develop innovative healthcare applications that use vision-based body tracking, we embrace the potential. However, weathered by experience, we are also aware of the limitations. We are frequently exposed to unrealistic expectations from our clinical partners as to what vision-based body tracking could do. These seem to derive from misconceptions of how the technology works. In this article, we provide a short explanation of the technology and then offer "stories from the field" of how to overcome its limitations to achieve workable systems that deliver the potential highlighted above.
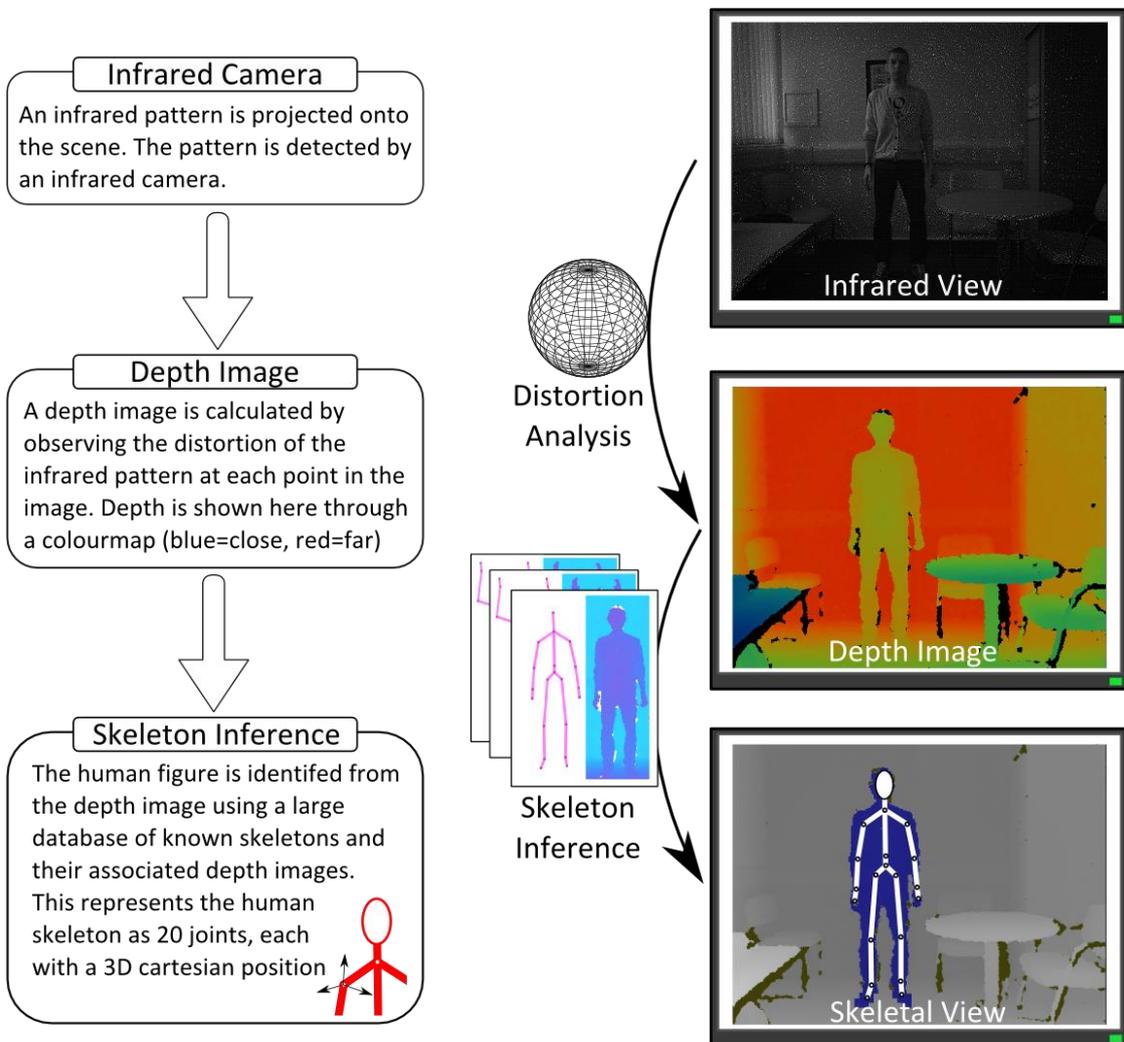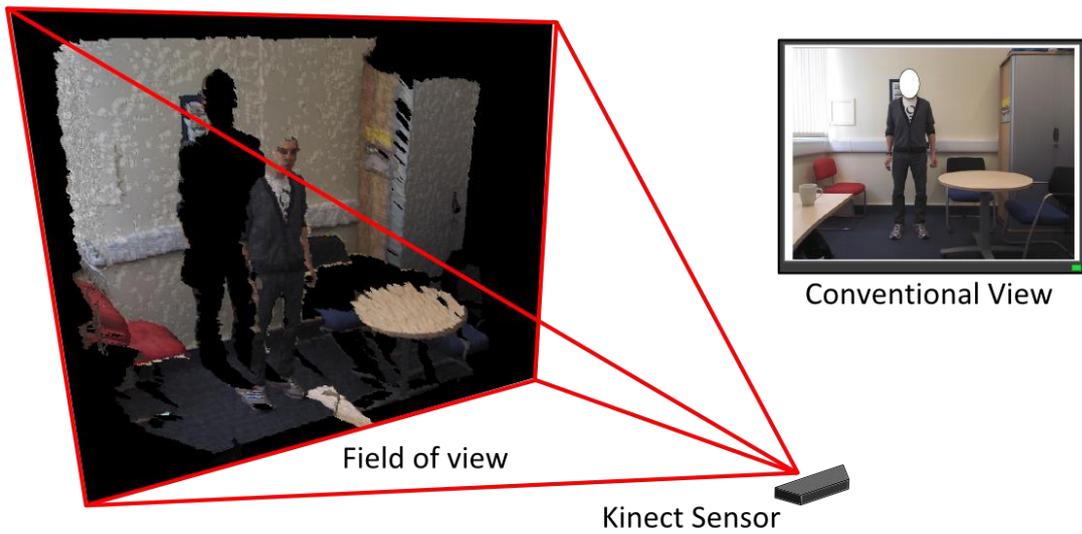
HOW DOES VISION-BASED BODY TRACKING WORK?

Vision-based body tracking relies on depth sensing camera systems, such as the Kinect, which record an 'image' in which each pixel has a 3-dimensional position rather than a colour. The depth camera produces this image by projecting a known infrared light pattern and then analysing any distortion when it bounces back to the camera from the environment. The distortions are used to mathematically calculate the 3D shape of objects in the camera view. This approach provides a way to gather 3D data regardless of colour, texture and lighting. However, anything that distorts the infrared light in unpredictable ways, such as shiny wheelchairs, can introduce incorrect measures, or "noise," into the depth image. See Figure 1 for visual description.

To capture human motion, the "skeleton" of the person moving is inferred from a series of depth images over time. A statistical inference algorithm is used to estimate where important parts of the body are, such as the head and hand, and connect them into a skeleton. The algorithm that comes with the Kinect, for example, can track and fit skeletons comprised of 20 joints to two people [6]. Joints include the head, shoulder, elbow, and wrist position, but not finer details such as individual fingers. The Leap Motion, on the other hand, can identify individual finger movements but does not track above the wrist.

The skeleton inference relies on statistical techniques (termed 'machine learning') that fit the depth image to a large database of pre-computed examples. For each example, both the depth image sequence and skeleton position are known. This makes it possible to distinguish the skeleton from the background and ensure that the skeleton's joints move anatomically. The example data set is key to the success of the inference. It must have examples similar to those to be tracked and cover a broad range of anatomical positions. The Kinect's algorithm is based on a data set with over 100,000 examples, but centres on young, healthy adults rather than older people or those with abnormal movement patterns.

Once a skeleton is successfully tracked, more work is required by a research team to use this data to calculate clinical measures. In some systems, a direct measure is calculated such as volume reach or gait speed. In others, "learned" correlations between a known measure, such as the Expanded Disability Severity Score (EDSS), and the movement performed by the patient, are used to predict the correct measure with a new patient. See Table 1 for a range of studies that calculate clinical measures in different ways. As a whole, many of the published studies have validated vision-based tracking against "gold-standard" measures as accurate enough for clinical contexts [7].

Conventional View

Field of view

Kinect Sensor

Infrared Camera

An infrared pattern is projected onto the scene. The pattern is detected by an infrared camera.

Depth Image

A depth image is calculated by observing the distortion of the infrared pattern at each point in the image. Depth is shown here through a colourmap (blue=close, red=far)

Skeleton Inference

The human figure is identifed from the depth image using a large database of known skeletons and their associated depth images. This represents the human skeleton as 20 joints, each with a 3D cartesian position

Distortion Analysis

Skeleton Inference

Infrared View

Depth Image

Skeletal View

Measure / Activity	Clinical Application	Captured Measures	Features					Reference
			Home Use	*Clinical Use*	*Technical Val*	*Clinical Study*	*Skeleton Use*	
Gait	General functional analysis	Stride duration Arm swing	✓		✓		✓	Gabel et al. Full body gait analysis with Kinect. In Proceedings of Engineering in Medicine and Biology Society (EMBC), pp. 1964-1967; 2012.
Posture	Postural Control	Single leg standing balance Forward reach Lateral reach		✓	✓	✓	✓	Clark et al. Validity of the Microsoft Kinect for assessment of postural control. Gait & Posture 2012; 36: 372-377..
Spinal Loading	Lower back and spine disorders	Linear acceleration Angular acceleration		✓	✓		✓	Ning et al. Assessing spinal loading using the kinect depth sensor: A feasibility study. Sensors Journal 2013; 13: 1139-1140.
Gait	Falls Risk	Stride time Stride length	✓		✓			Stone, et al. Mapping Kinect-based in-home gait speed to TUG time: a methodology to facilitate clinical interpretation. In Proceedings of Pervasive Computing Technologies for Healthcare (PervasiveHealth) pp. 57-64; 2013.
Gait	Falls Risk	Stride time Stride length	✓		✓			Stone et al. Unobtrusive, continuous, in-home gait measurement using the microsoft kinect. IEEE Transaction on Biomedical Engineering 2013; 60:2925-2932..
Upper limb functional movement	Upper extremity strength	Functional reaching volume Movement velocity Rate of fatigue		✓	✓	✓		Lowes, L.P. et al. Proof of concept of the ability of the Kinect to quantify upper extremity function in dystrophinopathy
Standing position	Foot posture	Lateral malleolar curvature Calcaneal inversion/eversion Talo-navicular joint bulging Medial longitudinal arch Forfoot abduction/adduction		✓	✓	✓		Mentiplay, et al. Reliability and validity of the Microsoft Kinect for evaluating static foot posture. J Foot and Ankle Research 2013; 6:14.
Upper extremity function	Neurological function	Upper and lower extremity functions Truncal stability		✓	✓	✓		Kontschieder et al. Quantifying progression of Multiple Sclerosis via classification of depth videos. Proceedings of The Medical Image Computing and Computer Assisted Intervention Society 2014.
Limb motion	Clinical functional analysis	Shoulder abduction Elbow flexion Hip abduction Knee flexion	✓	✓	✓	✓	✓	Bonnechere et al. Validity and reliability of the Kinect within functional assessment activities: Comparison with standard stereophotogrammetry. Gait and Posture 2014; 39: 593-598.

STORIES FROM THE FIELD

Walking tests with a bit of twist

Walking tests are an important way of assessing movement ability, but at the same time are challenging to capture with vision-based body tracking. Depth cameras such as the Kinect, when mounted at its usual height, track best between three and six meters from the sensor and, at its widest point, approximately three meters across. This does not provide enough room for the camera to see many steps during a walking test and certainly not the 10 meters often requested.

CM substituted turning on the spot for walking as a movement that is similar but more appropriate to capture using a single Kinect. Other groups have been successful in mounting the Kinect much higher (e.g. above a door) or lower (e.g. at knee height) than normal to obtain good results [8,9]. However, this is a trade-off between gaining a result for walking, and not capturing other movements. HM took another approach and used multiple Kinects spaced out over a 10-meter walkway to capture a few choice moments during a traditional walking test. This is closest to the clinical measure, but poses substantial algorithmic challenges to overcome overlapping infrared fields from the multiple sensors.

Measuring me or my home

When considering vision-based body tracking measures in the home, our clinical collaborators tend to request us to reproduce measures currently used in the clinic (e.g. toe-tapping tests). However, with vision-based body tracking, it is also possible to measure more contextually situated movement changes such as the location of freezing gait episodes in the home. In this case, we can use the environment as a way to trigger joint reflection on how a patient might manage their condition better in the home. For example, HM used the Kinect to take a photo with the on-board camera every time the body-tracking software detected a freezing gait incident. This photo can support a better clinical impression of the context leading to a frozen gait (e.g. number of people around, time of day). It is an alternative to relying solely on the vision-based body tracking software to 'measure' the patient's gait.

Congratulations. You've done 62% correct movements.

Feedback to patients is important for vision-based body tracking applications in the home. Yet, it is not obvious what the feedback should be. The easiest metric is percentage of correct movements during rehabilitation exercises. Such a number however, does not necessarily make sense. It may even have negative consequences of inducing a sense of failure leading to reticence towards further engagement in exercises. The Emo-pain research team at UCL, for example, discovered that clinicians believed that correcting the movement would increase anxiety. They successfully piloted giving patients feedback through non-verbal sounds corresponding with the envelope of the movement defined individually for each patient [10].

A helping hand is not always so helpful
In many cases, it is necessary for a health professional to support a patient in some way. Holding the hands of MS patients for example when doing a hopping test; or standing at their side with arms out stretched in front of the patient while doing a balance test. The close or touching arrangement of bodies often confuses the body tracking system in distinguishing the skeletons of the patient and helper. CM developed a strict protocol stipulating how the helper physically interacts with the patient in order to manually differentiate people in the data. A similar problem happens in serious game based assessment. The helper has to become skilled in hiding behind the player and avoiding large movements while providing physical support so that game activity is controlled only by the patient and not the "helping hand" by accident.

Monkey see, monkey does not do
The Finger-to-Nose test is a common neuro-assessment movement. It can be done in a myriad of ways. Some doctors ask people to stretch their arm out to the side and touch their nose, while others ask patients to stretch the arm to the front and then touch their nose. While these two variations may show a doctor exactly the same thing, intention tremor as the finger approaches the nose, they look very different to a computer vision algorithm. To help focus on variations in movement between healthy volunteers and patients, CM developed an avatar-based instructional system to show the patients the neuro-assessment movements in a standardised way.

Going solo in the home
While the ideal of 'remote' rehabilitation in the home will no doubt increase patients' independence and reduce healthcare costs, there are safety issues with even the simplest move. Home-rehabilitation is always associated with a certain level of risk, but the constraints associated with vision-based body tracking could increase these. For example, using props to stabilise the patient, such as tables or walking frames can confuse the system and are therefore not used. To create a safe system, TP created rules that give detailed instructions even for simple movements, such as sitting back down in a chair. An alternative approach used by colleagues at University College London was to use cheap plastic exercise planks that could be stacked to the appropriate height of the individual for a sit to stand test, but were made of a material that did not disturb the system.

All patients like games, right?
Rehabilitation exercises can become tedious, and non-compliance in the home is well recognised. Serious games have been one proposal to motivate rehabilitation exercise which can be realised using vision based body tracking. However, in some cases, off-the-shelf games that focus on fitness can not only be de-motivating to patients, but can also be unsafe. TP noted the anxious responses of chronic pain patients to a fast moving 'shoot-out' game attempting to encourage them to be more active. Options for users to change settings to easier or slower versions are

often required. These should be determine through consultation with all users, including patients.

THE FUTURE

While the Kinect was originally intended as a living room entertainer under the guise of xBox, the technology has quickly been repurposed to form the core of innovative healthcare applications in clinical assessment and rehabilitation. As shown by our stories from the field, there are clinical as well as practical challenges to make such systems a reality. Our stories highlight three major themes that health professionals need to consider when collaboratively designing applications of vision-based body tracking with professionals.

- Substantial effort needs to go into determining what types of measures and feedback vision-based body tracking should provide. This needs to account for the practicalities of the technology (e.g. range) as well as the opportunities of new environments (e.g. the home).

- Practical considerations need to be accounted for when planning capture in a particular environment so that data is analysable, whether it be finding a chair substitute, ways to support a patient, or ensuring everyone does the exercise in the same manner.

- The home is a place of opportunity with vision-based body tracking, but it would be naïve to think that we can do what we do now in the clinic (e.g. balance tests) or in the home (e.g. play games), without appropriate modifications to what constitutes a practically capturable, clinically relevant measure.

This article articulates how vision-based body tracking works and when it does not, to continue to inspire our clinical colleagues to imagine how new applications might be developed to support their practice in close collaboration with technologists.

Acknowledgements

Declaration of Interests

CM is an employee of Microsoft Research. HM has been an employee of Microsoft Research in the previous five years.

Authors' Contributions

CM structured the article, managed the discussions between contributors, and did the final edits. CM is the guarantor. PC, HM, and TP were equal contributors to the discussion and writing of this article. All approved the final version.

Authors' Information

The authors of this paper are researchers developing a diverse set of vision-based tracking applications for clinical assessment and rehabilitation. CM is working on the clinical assessment of Multiple Sclerosis; PC on movement analysis for the assessment of childhood development; HM on movement analysis for Parkinson's disease; and TP on rehabilitation systems to support pain management. With a common desire to use vision-based body tracking to deliver clinically meaningful, cost-effective systems, we were drawn together to write this article after an expert workshop on the topic held at Microsoft Research Cambridge in November 2013.

References

1. Simon SR. Quantification of human motion: gait analysis-benefits and limitations to its application to clinical problems. Journal of biomechanics. 2004;37(12):1869–80.

2. Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet neurology. 2007;6(12):1094–105.

3. Kharrazi H, Lu AS, Gharghabi F, Coleman W. A Scoping Review of Health Game Research: Past, Present, and Future. Games for health journal. 2012;1(2).

4. Lu AS, Kharrazi H, Gharghabi F, Thompson D. A Systematic Review of Health Videogames on Childhood Obesity Prevention and Intervention. Games for health journal. 2013;2(3):131–141.

5. Weightman A, Preston N, Levesley M, Bhakta B, Holt R, Mon-Williams M. The nature of arm movement in children with cerebral palsy when using computer-generated exercise games. Disability and rehabilitation. Assistive technology. 2013 Apr 18.

6. Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, Kipman A, Blake A. Real-time human pose recognition in parts from single depth images. CVPR 2011. 2011 Jun:1297–1304.

7. Bonnechère B, Jansen B, Salvia P, Bouzahouene H, Omelina L, Moiseev F, Sholukha V, Cornelis J, Rooze M, Van Sint Jan S. Validity and reliability of the Kinect within functional assessment activities: Comparison with standard stereophotogrammetry. Gait & Posture. 2014;39(1):593–598.

8. Gabel M, Gilad-Bachrach R, Renshaw E, Schuster A. Full body gait analysis with Kinect. In: IEEE Engineering in Medicine and Biology Society Conference. Vol. 2012. ; 2012. p. 1964–7.

9. Stone EE, Skubic M. Unobtrusive, continuous, in-home gait measurement using the Microsoft Kinect. IEEE transactions on bio-medical engineering. 2013;60(10):2925–32.

10. Singh A, Klapper A, Jia J, Fidalgo A, Jimenez AT, Kanakam N, Bianchi-Berthouze N, Williams A. Motivating People with Chronic Pain to do Physical Activity: Opportunities for Technology Design. In: Proceedings of CHI'14. ; 2014.