Unobtrusive Measurement of Subtle Nonverbal Behaviors with the Microsoft Kinect. IEEE VR Workshop on Ambient Information Technologies, pp. 10–13, 2012.

Unobtrusive Measurement of Subtle Nonverbal Behaviors with the Microsoft Kinect

Nathan Burba, Mark Bolas, David M. Krum, Evan A. Suma
Institute for Creative Technologies
University of Southern California
Los Angeles, CA
{*burba, bolas, krum, suma*}*@ict.usc.edu*

***Abstract*—We describe two approaches for unobtrusively sensing subtle nonverbal behaviors using a consumer-level depth sensing camera. The first signal, respiratory rate, is estimated by measuring the visual expansion and contraction of the user's chest cavity during inhalation and exhalation. Additionally, we detect a specific type of fidgeting behavior, known as "leg jiggling," by measuring high-frequency vertical oscillations of the user's knees. Both of these techniques rely on the combination of skeletal tracking information with raw depth readings from the sensor to identify the cyclical patterns in jittery, low-resolution data. Such subtle nonverbal signals may be useful for informing models of users' psy-chological states during communication with virtual human agents, thereby improving interactions that address important societal challenges in domains including education, training, and medicine.**

***Keywords*-nonverbal behavior; breathing; fidgeting; depth sensors**

Unobtrusive Measurement of Subtle Nonverbal Behaviors with the Microsoft Kinect

Nathan Burba, Mark Bolas, David M. Krum, Evan A. Suma
Institute for Creative Technologies
University of Southern California
Los Angeles, CA
{*burba, bolas, krum, suma*}*@ict.usc.edu*

***Abstract*—We describe two approaches for unobtrusively sensing subtle nonverbal behaviors using a consumer-level depth sensing camera. The first signal, respiratory rate, is estimated by measuring the visual expansion and contraction of the user's chest cavity during inhalation and exhalation. Additionally, we detect a specific type of fidgeting behavior, known as "leg jiggling," by measuring high-frequency vertical oscillations of the user's knees. Both of these techniques rely on the combination of skeletal tracking information with raw depth readings from the sensor to identify the cyclical patterns in jittery, low-resolution data. Such subtle nonverbal signals may be useful for informing models of users' psychological states during communication with virtual human agents, thereby improving interactions that address important societal challenges in domains including education, training, and medicine.**

***Keywords*-nonverbal behavior; breathing; fidgeting; depth sensors**

I. Introduction

Nonverbal behavior, such as body posture, gaze, facial expressions, and gesture, plays a critical role in human communication. These behaviors can convey a multitude of information about the attention, emotions, attitudes, and physiological states of conversation participants [1] [2] [3]. As such, user state sensing is a major technological challenge for providing realistic and meaningful experiences with virtual human agents, commonly used in a variety of domains including education [4], training [5], and medicine [6]. However, in addition to measuring overt behaviors, it is also beneficial to detect much more subtle signals that can provide additional insight into the psychological and physiological state of the user.

Recent advances in depth-sensing technology has led to the widespread proliferation of low-cost devices that can capture scene data and human motion in 3D. Sparked by their relatively low cost and the fact that they can track motion without any markers or worn devices, there is a great deal of burgeoning interest in using these sensors for unobtrusively measuring user behavior. Chief among them is the Microsoft Kinect sensor, which was released in late 2010 and currently holds the world record as the fastest selling consumer electronics device in history.

In this paper, we describe two approaches for unobtrusively measuring subtle nonverbal signals using the Microsoft Kinect sensor. First, we demonstrate how respiratory rate information can be extracted by observing the cyclical expansion and contraction of the user's chest. Next, we apply a similar approach to detect a specific type of fidgeting behavior known as "leg jiggling." We believe that these subtle signals, combined with other well-known methods of interpreting nonverbal human behavior, will be useful for improving future virtual human interactions and other context-aware computing applications.

II. Previous Work

Recognition and analysis of human motion is one of the most active topics in the field of computer vision, and the quantity of literature on this topic is vast. A fairly recent survey of techniques can be found in [7]. Many existing approaches focus on analyzing motion in standard videos, and consequently may not always be fast enough for time-sensitive real-time interactions. However, the release of the Microsoft Kinect has prompted a number of action recognition papers that have leveraged its capability to generate both RGB images and corresponding depth maps of the scene (e.g. [8] [9]). Additionally, the skeletal tracking capabilities provided by SDKs from Microsoft Research and OpenNI/PrimeSense have been leveraged for a multitude of applications such as repurposing traditional video games [10], rehabilitation [11], interactive storytelling [12], and gesture recognition for social robots [13]. However, most of the existing literature focuses on body posture and hand gestures, and to the best of our knowledge, this paper represents the first attempt to recognize the subtle motions of breathing and fidgeting with consumer depth sensing technology.

III. Respiratory Rate

Respiration is a necessary human activity that is known to vary closely with emotional state, making it a potentially useful indicator of psychological changes such as stress or relaxation [14]. Typically, accurately measuring respiration requires a physical apparatus placed either on the body or near the mouth (e.g. [15]). To estimate the user's breathing rate unobtrusively, we use the Kinect depth sensor to measure the average depth over time of the point cloud that corresponds to the user's chest. This method relies

978-1-4673-1246-2/12/$31.00 ©2012 IEEE

upon the cyclical expansion and contraction of the chest cavity during breathing. However, measuring this behavior with the Microsoft Kinect sensor introduces several notable challenges. First, we were concerned that the sensor would not provide enough fine-grained accuracy to reliably detect the subtle motions of breathing. Additionally, we observed that in terms of appearance, this complex muscular process manifests itself differently between people depending on body type and posture.

The first step in creating a generic solution is to isolate the area corresponding to the chest cavity in the depth map. To do so, we take advantage of the fact that the Microsoft Research Kinect SDK also provides a skeletal tracker. The chest cavity is defined as a rectangular region outlined using the position of the shoulders and the length of the torso in the skeletal model. We only measure the top half of the torso, as this generally represents the area that has the greatest depth displacement from breathing motions.

Once the chest cavity is defined, we calculate the average depth (distance away from the sensor along the Z axis) of all pixels within the chest cavity. While the depth of each individual pixel may fluctuate randomly due to sensor noise, we expect that as the chest expands and contracts, this number will fluctuate in a reliable and cyclical manner. While the measurement of this data is somewhat trivial when the user is in a fixed position relative to the sensor, it becomes more challenging when the user moves around. Because the accuracy of the sensor and skeletal tracking algorithms are imperfect, macro-scale body motions sometimes mask the subtle movement we are attempting to estimate. One troublesome source of noise is the fact that the skeletal positions that determine the bounding box of the chest cavity are constantly changing. Thus, since we are assuming that the user is in a sitting position, we define the chest cavity in a fixed position, and only recalculate the bounds if the user moves relative to the Kinect by an amount greater than a certain positional threshold. In informal testing, we determined a minimum threshold of 10 pixels to be a reasonable value that increases the reliability of the chest cavity depth calculation.

To estimate the respiratory rate, we need to look for cyclical patterns in the raw depth measurement. When the user inhales, we would expect the chest to expand, resulting in a depth measurement that decreases. Conversely, the depth measurement will increase when the user exhales, causing the chest cavity to contract. These two states alternate, with each breath consisting of a period of inhalation followed by a period of exhalation. However, since the data from the Kinect is noisy, the average chest depth may rapidly fluctuate up and down even though it is following a general trend consistent with a breathing motion. Thus, to find the inflection points in the breathing signal, we look for periods of consistent rising and falling over time. The depth measurements are divided temporally into four consecutive periods of three frames each, resulting in a total window of 12 frames. For each period, the average depth is calculated across the three frames to remove fluctuations due to noise. If the differences between all four periods are positive, then we consider this an exhalation. If the differences show consecutive negative moment, then it is labeled as an inhalation.

Figure 1. Measuring breathing with the Microsoft Kinect sensor. The graphed line shows the cyclical fluctuation of the chest cavity expansion and contraction from inhalation and exhalation. The breathing state (IN or OUT) along with the respiratory rate is also displayed on screen.

Once the user's breathing motions has been identified, estimating the respiratory rate at a given instant in time is a simple process of measuring the elapsed time between consecutive breaths. In our implementation, we maintain a list of the previous recorded data for the previous 10 seconds. A weighted average is then applied, with a higher weight assigned to the more recent data. This resulting calculation provides a balance between providing a stable measurement over time while also being responsive to sudden changes in behavior. Figure 1 shows an example of our respiratory rate application.

One important caveat of this technique is the user moves physically closer or further away from the sensor by more than a small amount, it will momentarily disrupt the breathing measurement until the a stationary pose is resumed. In general, due to the accuracy of the sensor, large or quick displacements of the chest cavity will mask the fine-grained fluctuations from breathing. In the future, this approach may be improved by using skeletal information to filter and stabilize the depth map. However, for cases where a user is standing still or sitting in a chair, we were pleasantly surprised that such a subtle motion could be unobtrusively measured using a relatively coarse sensor such as the Kinect. Of course, readings vary with different body shapes, breathing styles, and, like many other Kinect applications, clothing style. However, our informal tests have indicated that this technique can return surprisingly accurate results which, under the right conditions, are comparable to other more intrusive measurement methods.

IV. Fidgeting

Another subtle, yet possibly suggestive behavior is fidgeting. These involuntary restless movements are often associated with nervousness, and many studies have observed increased incidence of fidgeting behaviors during situations involving conflict or stress (e.g. [16]). As a general term, fidgeting can also include shifting in one's chair, playing with jewelry or hair, repeated head motions, and an assortment of other movements. We focus specifically on "leg jiggling," a rapid and repetitive up and down movement of the leg that has been previously suggested by researchers to be a possible indicator of tension in both American and Asian cultures [17]. Given the potential implications of this behavior on the user's mental state, it is a potentially useful signal to measure unobtrusively.

Similar to the respiratory rate technique described in the previous section, our method of detecting fidgeting uses depth map data combined with skeletal information from the Microsoft Research Kinect SDK. Knee oscillation is different from chest oscillation in one fundamental way. When the user it sitting down, the cyclical variation is vertical in relation to the sensor rather than fluctuations in depth. At first, one might think that we can simply measure the vertical motion of the knee joint of the skeletal model. However, even though the skeleton is tracked at 30 frames per second, the skeletal data is too approximate, jittery, and slightly delayed compared to the depth cloud data. Instead, we use the skeletal information to determine the approximate position of the knee joints within the 2D depth map.

Once we have the approximate knee position (which is typically around the center), the next step is to determine a more exact point along surface of the user's leg immediately above the knee. To do this, we cast a ray upwards, measuring the depth of each pixel along the line. When the depth abruptly increases, we assume that the top of the knee has been reached. This method works regardless of whether or not the knee is facing the sensor, because an abrupt change in depth will either refer to the slope of the user's thigh, the torso, or the scene background, depending on how the user is sitting.

Using the top of the knee as determined by the depth map, we can then measure the cyclical vibration of the leg. A single vibration is considered to be a rapid motion upward followed by a rapid motion downward. To distinguish these motions from other types of body movement, we apply a minimum speed threshold to consider the motion a fidgeting behavior. In informal testing, we determined that a threshold of approximately 9 pixels/sec on the 320x240 depth map produced reliable results at typical distances away from the sensor. The frequency of vibration is calculated and displayed using the same method described for respiratory rate (Section III). Figure 2 shows an example of our test application that displays the vibration rate of each leg independently.

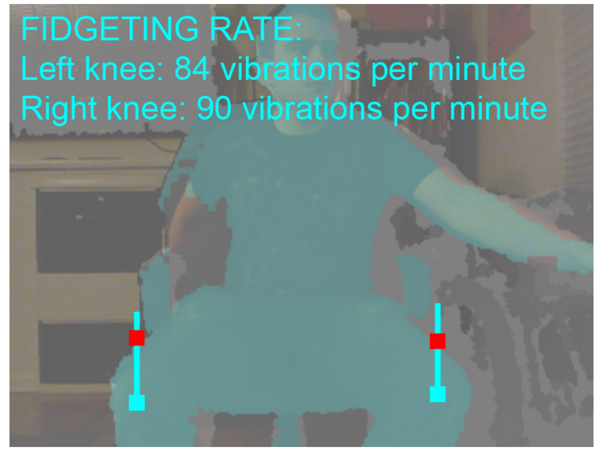

Figure 2. Measuring "leg jiggling" fidgeting behavior with the Microsoft Kinect sensor. The two vertical lines show the approximate knee position returned from the skeleton tracker (cyan dot) as well as the point along the surface of the leg that is used to measure leg oscillation (red dot).

While this technique when developed assuming that the user would be in a neutral seated position with legs uncrossed, our informal tests have shown that it is robust to many different poses. If the user's legs are crossed, the application can still detect the oscillations of the knee height that are necessary to identify fidgeting behavior.

V. Discussion and Conclusion

Measuring the user's respiratory rate can be considered a form of directed depth testing. In the future, the technique presented in this paper may be extended by isolating the chest cavity using more complex methods. For example, to adapt to users with different breathing styles, a calibration could be performed to isolate the specific areas of the chest that most strongly oscillate during breathing. Additionally, methods of filtering and stabilizing the depth map as the user moves around would be an important step towards detecting such subtle fluctuations in situations where it can not be assumed that the user is in a stationary seated or standing position. Additionally, directional audio sensing and echo cancellation using the Kinect's four microphones could be used to isolate breathing-related sounds. This could be used to improve robustness of breathing detection when visual information alone is unreliable.

"Leg jiggling" is a specific form of fidgeting, but there are many other manifestations of this behavior, such as rapid repetitive motions of the fingers. Due to resolution limitations, the ability of the Kinect to isolate and track individual fingers at full-body distances is highly questionable. However, it may be possible to detect the presence of rapid finger motions, even if the sensor does not have the required fidelity for finger tracking. Additionally, other fidgeting behaviors such as stroking one's beard or playing with hair would also likely be detectable using the sensor. Thus, developing other types of signals that indicate fidgeting or agitation is an open area for future work.

In both cases, robust evaluation of performance of these algorithms with a variety of users, distances, orientations, and movement frequencies is necessary before they can be applied in a practical setting. Our long term goal is to leverage these signals, along with other methods for both verbal and nonverbal behavior understanding, to inform a model of the user's psychological state. This is an important step towards creating lifelike virtual human agents that can communicate more effectively and address important societal challenges in domains including education, training, and medicine.

References

[1] M. Argyle and M. Cook, *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press, 1976.

[2] P. Ekman and E. L. Rosenberg, *What the face reveals: Basic and applied studies of spontaneous expressions using the Facial Action Coding System (FACS)*. New York: Oxford University Press, 1997.

[3] A. Kendon, "Language and Gesture: Unity or Duality," in *Language and Gesture: Window into Thought and Action*. Cambridge: Cambridge University Press, 2000, pp. 47–63.

[4] E. Rader, M. Echelbarger, and J. Cassell, "Brick by Brick: Iterating Interventions to Bridge the Achievement Gap with Virtual Peers," in *ACM Conference on Human Factors in Computing Systems (CHI)*, 2011, pp. 2971–2974.

[5] P. Kenny, A. Hartholt, J. Gratch, W. Swartout, D. Traum, S. Marsella, and D. Piepol, "Building Interactive Virtual Humans for Training Environments," in *Interservice/Industry Training, Simulation, and Education Conference*, no. 7105, 2007.

[6] A. A. Rizzo, B. Lange, J. G. Buckwalter, E. Forbell, J. Kim, K. Sagae, J. Williams, B. O. Rothbaum, J. Difede, G. Reger, T. Parsons, and P. Kenny, "An Intelligent Virtual Human System for Providing Healthcare Information and Support," in *Medicine Meets Virtual Reality*, 2011.

[7] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine Recognition of Human Activities: A Survey," *IEEE Transactions on Circuits and Systems For Video Technology*, vol. 18, no. 11, pp. 1473–1488, 2008.

[8] Z. Ren, J. Meng, J. Yuan, and Z. Zhang, "Robust hand gesture recognition with kinect sensor," in *ACM International Conference on Multimedia*, 2011, pp. 759–760.

[9] S. Sempena, N. U. Maulidevi, and P. R. Aryan, "Human Action Recognition Using Dynamic Time Warping," in *International Conference on Electrical Engineering and Informatics*, 2011, pp. 1–5.

[10] E. A. Suma, B. Lange, A. Rizzo, D. M. Krum, and M. Bolas, "FAAST : The Flexible Action and Articulated Skeleton Toolkit," in *IEEE Virtual Reality*, 2011, pp. 245–246.

[11] B. Lange, E. A. Suma, B. Newman, T. Phan, C.-y. Chang, A. Rizzo, and M. Bolas, "Leveraging Unencumbered Full Body Control of Animated Virtual Characters for Game-Based Rehabilitation," in *HCI International*, 2011, pp. 243–252.

[12] F. Kistler, D. Sollfrank, N. Bee, and E. André, "Full Body Gestures Enhancing a Game Book for Interactive Story Telling," in *Interactive Storytelling*. Springer Berlin / Heidelberg, 2011, pp. 207–218.

[13] V. Gonzalez-Pacheco and M. A. Salichs, "Integration of a Low-Cost RGB-D Sensor in a Social Robot for Gesture Recognition," in *International Conference on Human-robot Interaction*, 2011, pp. 229–230.

[14] I. Stevenson and H. S. Ripley, "Variations in respiration and in respiratory symptoms during changes in emotion," *Psychosomatic Medicine*, vol. 14, pp. 476–490, 1952.

[15] M. Vegfors, L.-G. Lindberg, H. Pettersson, and P. Öberg, "Presentation and evaluation of a new optical sensor for respiratory rate monitoring," *International Journal of Clinical Monitoring and Computing*, vol. 11, no. 3, pp. 151–156, 1994.

[16] M. H. Krout, "An Experimental Attempt to Produce Unconscious Manual Symbolic Movements," *The Journal of General Psychology*, vol. 51, no. 1, pp. 93–120, 1954.

[17] L. Sechrest and L. Flores, "The occurrence of a nervous mannerism in two cultures," *Asian Studies*, vol. 9, pp. 55–63, 1971.