# Accurate 3D Human Pose Recognition via Fusion of Depth and Motion Sensors

S. B. Nam, S. U. Park, J. H. Park, M. D. Zia Uddin, and T. S. Kim

***Abstract*—In this work, we present an accurate 3D human pose recognition (HPR) work via multi-sensor fusion. Lately, 3D HPR is widely performed using a depth imaging sensor, but this approach has limitations: 1) orientations of body parts cannot be accurately recognized and 2) it suffers from occlusion.**

**To achieve an accurate and stable recognition of human poses in real-time, in this study, we propose to use inertial measurement units (IMUs) which are used to estimate the orientation of body limbs and solve the occlusion problem. Via fusion of depth and IMU sensors, our results demonstrate significantly improved 3D human pose reconstruction: our results show the accurate recognition of twist and location of the arms even under occlusion. Our presented approach could be critical if 3D HPR is to be used for medical applications such as musculoskeletal analysis via in 3D as demonstrated in this study.**

***Index Terms*—Human pose recognition, depth sensors, inertial measurement units, sensor fusion, musculoskeletal analysis.**

## I. Introduction

Human Pose Recognition (HPR) is a technique of accurately estimating a human body posture in 3D. Recently, HPR has been an active research topic not only in computer vision, user interface, virtual reality, and body shape analysis but also in biomedical fields such as rehabilitation.

There are classical approaches to estimate human pose. One approach, which is widely used for motion capture in movies, is to utilize optical markers. However, this technique requires attaching multiple optical markers on human body parts and a special setup with multiple cameras. Another approach is to use multiple Inertial Measurement Units (IMUs) requiring at least 10 IMUs attached to the human body parts. Although these sensors are very sensitive to motion, this approach is not practical because of their cost and inconvenience.

Recently, a marker-free method for HPR is proposed. In particular, a RGB-Depth (RGB-D) sensor which produces both RGB images and a distance image is actively employed. For instance, in the work of Luong *et al*. [1], HPR based on body parts recognition using a single depth image was presented. Another example of HPR based on inertial sensors could be found in the work of Roetenberg *et al*. and Brigante *et al*. [2], [3] in which human motion tracking using miniature inertial sensors was presented.

However, either depth or inertial sensors only based approaches have their own strengths and limitations. Although the depth-based methodology works without optical markers, it cannot detect twist information of body parts. Also it fails under occlusions. In contrast, the inertial sensor based approach is stable in tracking during fast motion, but it needs wearing of multiple sensor units as a set of 17 sensors was used in [3]. Accordingly, it would be beneficial if these approaches or sensors are fused for HAR as attempted in [4] where both sensors were used for a single arm tracking. However, their approach had a limited success since only skin color information from RGB images was utilized.

In this work, we present a novel 3D HPR work via fusion of a single depth sensor and a minimum number of IMUs in order to overcome the limitations of each approach. In the presented methodology, we reconstruct human body poses using a single depth camera for a whole body as done in [5], [6]. In addition, we acquire information of the body limb (i.e., position and twist information of arms or legs) from two IMUs which are separately attached onto the right and left wrists or legs. Thereby, our proposed method takes both advantages of each sensor by acquiring accurate joint position and orientation information of the limbs, and finally reconstructs an accurate 3D whole body pose.

Our results show the proposed methodology is a practical solution for accurate pose estimation in 3D. Also with it, one can obtain the joint twist information of the arms and solve the occlusion problem. We have applied our proposed methodology for a medical application in which a musculoskeletal analysis of the recognized human pose is performed as a function of a future rehabilitation system.

## II. Methods

In this section, we present a hybrid HPR methodology using a depth sensor and a set of two IMUs. A flow chart of our proposed methodology is shown in Fig. 1 consisting of the following steps. First, we estimate a 3D human pose from a single depth sensor. Second, we also estimate rotation (i.e. twist) and location information of the lower arms from IMUs. Finally, we reconstruct a full body pose with fused information on a combined ellipsoidal and hexahedral rigid body model. To verify our methodology, we have performed real-time pose recognition in 3D. In addition, as a biomedical application where precise information of musculoskeletal analysis is needed, we have reconstructed musculoskeletal

Manuscript received March 31, 2015; revised September 16, 2015. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (2014R1A2A2A09052449).

S. B. Nam, S. U. Park, J. H. Park, and T. S. Kim are with the Department of Biomedical Engineering, Kyung Hee University, Seocheon-Dong, Giheung-Gu, Yongin-Si, Gyeonggi-Do, 446-701, Republic of Korea (e-mail: sbnam@khu.ac.kr, supark@khu.ac.kr, psmt2655@khu.ac.kr, tskim@khu.ac.kr).

M. D. Zia Uddin is with the Department of Computer Education, Sungkyunkwan University, Jongno-Gu, Seoul, Republic of Korea (e-mail: ziauddin@skku.edu).

doi: 10.18178/ijfcc.2015.4.5.412

poses from the recognized poses with the proposed method.

### *A. Sensors*

For the depth based HPR, we obtain the human depth silhouettes from a single depth camera (PrimeSense Inc.). The imaging parameters of the depth camera are an image size of 640 × 480, field of view 57.5, 45 and 69 degree (horizontal, vertical, and diagonal respectively), frame speed of 30fps, and operating range of 0.8~3.5m.

In order to estimate the motion information of the arms, we have used two wireless IMUs (EBIMU24GV2, E2box). The parameters of the IMUs are the baud rate of 9600bps, data bit of 8bits, digital low pass filter of 98Hz, data processing speed of 225Hz, and output format of quaternion angle.

### *B. Training Random Forests*

We utilize Random Forests (RFs) which are a combination of trees to classify all depth pixels of a silhouette into thirty-one body parts [7], [8]. Our implemented RFs include 5 trees, maximum tree depth of 20, 2,000 feature vectors, 2,000 random pixels, threshold range from -500 to 500, and 31 classes. The RFs are trained with our training database created using multiple synthetic human body models from a commercial software (3Ds Max 2012, [9]) with motion database from Carnegie-Mellon Graphics Lab [10] as shown in Fig. 2.

### *C. Body Parts Recognition and 3D Joint Proposals*

From a given depth silhouette image $I$, a feature $f$ is calculated by taking difference of two depth values $d_I(\cdot)$ at two vectors of $u$ and $v$ from a pixel $x$ as in (1). $d_I(x)$ is a normalization factor which decides a window size for $u$ and $v$

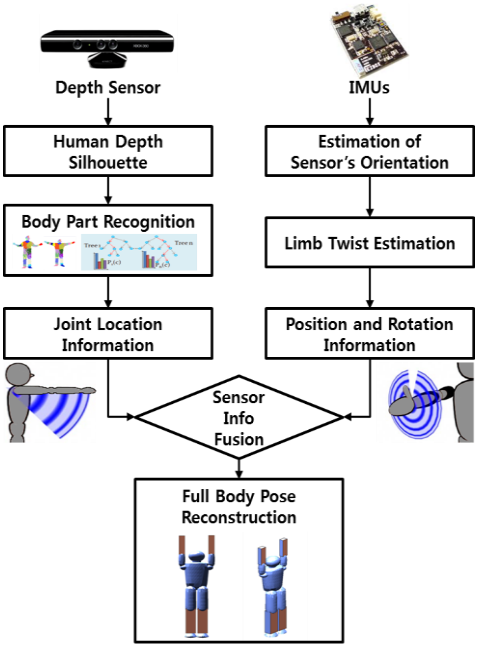



Fig. 1. A flow chart of our proposed hybrid HPR methodology.

$$f_\theta(I,x) = d_I\left(x+\frac{u}{d_I(x)}\right) - d_I\left(x+\frac{v}{d_I(x)}\right) \qquad (1)$$

For all pixels in the silhouette, the trained RFs decide what body parts each pixel belongs to. Finally, a recognized body part map gets obtained. Then, we estimate 3D joint positions from the body parts map utilizing Mean Shift algorithm [11].

### *D. Depth Based HPR*

We reconstruct a 3D human pose from the estimated joint proposals. From each joint to the next joint, a directional vector gets estimated to represent the direction of each body part. Then the set of directional vectors are converted into the spherical coordinates to represent the pose. For each joint, a rotation matrix $R_U$ is obtained by multiplying two rotation matrices computed by two angles of each joint.

Since the depth based HPR can only estimate the directional information of body parts, it has two degree of freedom (DOF) without twist information of all body parts. Therefore, the reconstructed pose is represented with several ellipsoids for torso, arms, legs, and the head.

### *E. Hybrid Depth and IMU HPR*

In our proposed hybrid HPR approach, we reconstruct a body pose first using the depth based HPR, and then we utilize two IMUs which are attached to the lower arms to get the twist and location information of the arms. In order to avoid the gimbal-lock effect (i.e. the loss of one DOF in 3D), quaternion angles are used. The quaternion angle is represented by 2 elements {$W$, $\boldsymbol{V}$} where $W$ is a scalar for rotation angle, and $\boldsymbol{V}$ is a vector of rotation axis in 3D, $\boldsymbol{V}$= {$V_x$, $V_y$, $V_z$}. We compute a rotation matrix for the lower arms $R_L$ from the normalized quaternion angle using (2).

$$R_L = \begin{bmatrix} V_x^2+V_y^2-V_z^2-W^2 & 2(V_yV_z-V_xW) & 2(V_yW+V_xV_z) \\ 2(V_yV_z+V_xW) & V_x^2-V_y^2+V_z^2-W^2 & 2(V_zW-V_xV_y) \\ 2(V_yW-V_xV_z) & 2(V_zW+V_xV_y) & Vx^2-Vy^2-Vz^2+W^2 \end{bmatrix} \qquad (2)$$

To reflect the twist information of the arms, we have modified the ellipsoidal model to the rigid human body model as shown in Fig. 3 the lower arms are represented with the hexagonal boxes. The orientation of hexahedral body part is derived from IMUs data. The twist of the arm is represented by rotating the hexagonal body part.

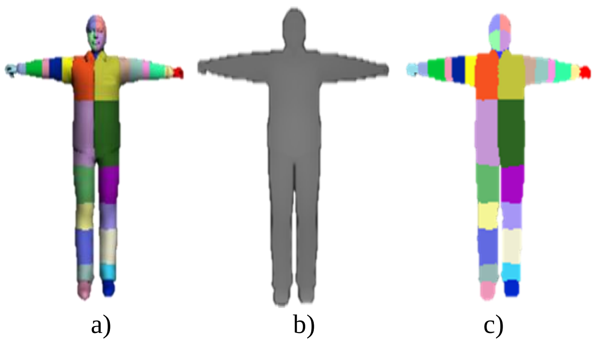



Fig. 2. Examples from a synthetic database. a) a synthetic human body model, b) its corresponding depth silhouette, and c) color labeled body parts map.

The three dashed arrows in different color in Fig. 3 indicate the palm (red), fingertip (black), and outer side (green) of the left hand. In the same way, the red surface of each lower arm is the palm, yellow is the backside of the hand, and blue and green parts indicate the inner and outer side from the initial pose.

In order to keep the kinematic chain of all body parts, (4) is used to get the transform matrix for the upper arm $T_U$. As shown in (5), the transformation matrix for the lower arm $T_L$ is obtained by multiplying the transformation matrix for the upper arm $T_U$ to the lower arm $R_L$. Finally, we have a fully reconstructed human body pose with the arm twist information.

$$T_U = t_{torso} R_{torso} R_U \tag{4}$$

$$T_L = T_U R_L \tag{5}$$

where $t_{torso,}$, $R_{torso}$ are the translation and rotation matrix of torso respectively, and $R_U$ the rotation matrix of the upper limb from the depth based HPR approach.

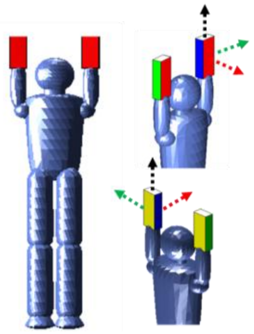
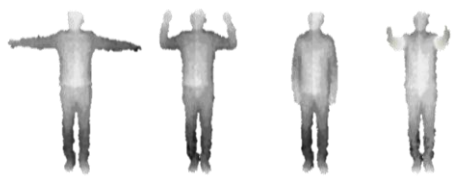


Fig. 3. Our whole human body model (left, front view), its front side view (right top), and its back side view (right bottom).

## III. Result and Discussion

### A. 3D HPR via Depth Only

Fig. 4 shows the pose recognition results using the depth only. Fig. 4 a) shows some sample depth silhouettes of four different poses. Fig. 4 b) shows the labeled body parts map and estimated joint proposals. Fig. 4 c) shows the reconstructed poses in 3D. Since the twist information of the arms is not available, they are represented only with the ellipsoids.



a)

b)

c)

Fig. 4. Result of depth only based HPR. a) input depth silhouette, b) labeled body parts and joint proposals, and c) reconstructed body poses using the ellipsoidal model.

### B. Hybrid HRP via Depth and IMUs

We have evaluated our proposed hybrid pose recognition system by comparing the results to the pose reconstructed from the depth only. Fig. 5 shows the result of the comparison work. Fig. 5 a), the first row, shows the RGB images of various poses, b) their corresponding depth silhouettes, c) the reconstructed poses from depth only, and d) the reconstructed poses from the proposed hybrid. Our hybrid HPR system produces the poses with the precise poses of the arm. The colors of the hexagonal arms show their correct directions. In contrast to the results from the hybrid, the correct directions of the lower arms are not shown in Fig. 5 (c).

Table I shows the Euler angles of the left lower arm (LLA) and right lower arm (RLA) for the first six frames in Fig. 5. The depth based approach could estimate two angles only excluding the rotation angle of the arm which is indicated as N/A. Their correct angular values from the hybrid system are given in the table.

The last three columns in Fig. 5 show the result for the cases of occlusion. When the lower arm is behind the body, the depth based approach could not recognize the correct pose. Also when the arm is in front of the body b close to the body, the depth based approach could not distinguish the arm from the main body. However, the hybrid approach could recognize the correct poses even in these occlusion cases.

TABLE I: Euler Angles in Degree of the Lower Arms from the Depth Only and the Hybrid HPR Systems

| Frame # | | 1 | | | 2 | | | 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Depth Only | LLA* | -118 | 23.9 | N/A | -149 | 94.4 | N/A | 82.7 | 10.8 | N/A |
| | RLA* | 147.5 | 13.2 | N/A | 173.8 | 109.4 | N/A | 99.3 | 9.5 | N/A |
| Hybrid | LLA | -118 | -80.6 | -84.5 | -99.9 | 18.5 | -125 | 9.9 | 60.2 | -170 |
| | RLA | 156.1 | -72.4 | 19.4 | 104.6 | 21.4 | 120.7 | -6.1 | 53.4 | 156.8 |
| **Frame #** | | **4** | | | **5** | | | **6** | | |
| Depth Only | LLA | -47.9 | 10.9 | N/A | -162 | 99.7 | N/A | -67.1 | 16.3 | N/A |
| | RLA | 37.5 | 17.4 | N/A | 167.2 | 116.6 | N/A | 128.9 | 8.6 | N/A |
| Hybrid | LLA | 76.3 | 3.5 | -112 | -85.0 | 4.7 | 3.7 | 5.5 | 4.0 | 163.4 |
| | RLA | -84.5 | 5.9 | 77.8 | 75.0 | 16.3 | -14.6 | -10.0 | -4.8 | 171.0 |

*LLA: Left Lower Arm; RLA: Right Lower Arm

### C. 3D Musculoskeletal Analysis

As an application of our hybrid system, we have established a link to the musculoskeletal model of OpenSim [12] which is a publically available biomechanics analysis package. Fig. 6 shows the musculoskeletal pose reconstructed from its corresponding pose from the hybrid system. Fig. 6 b) shows the result from depth only, c) from the hybrid. In this example, both palms are facing front, but the result of the depth only shows incorrect poses of the arms. However, our proposed hybrid approach reconstructs the arm poses correctly as shown in Fig. 6 c).

This kind of analysis can be used for home-rehabilitation services in the future as smart physical therapy. For stroke patients, muscle parameters such as muscle force and length of muscles information during physical exercise could be monitored at home.

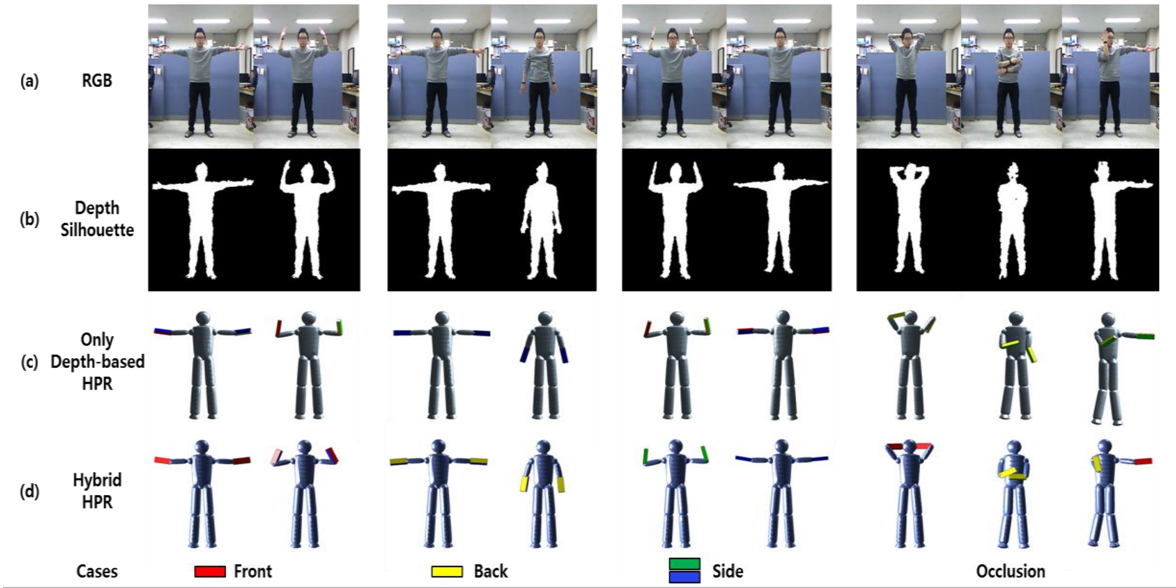



Fig. 5. Comparison of our proposed hybrid HPR and depth based HPR Systems. a) Input RGB and depth images, b) depth silhouette, c) depth based and d) hybrid HPR results using rigid body model. Each column indicates one frame case. The last row indicates the direction of the arm: front in red, back in yellow, side in green or blue.

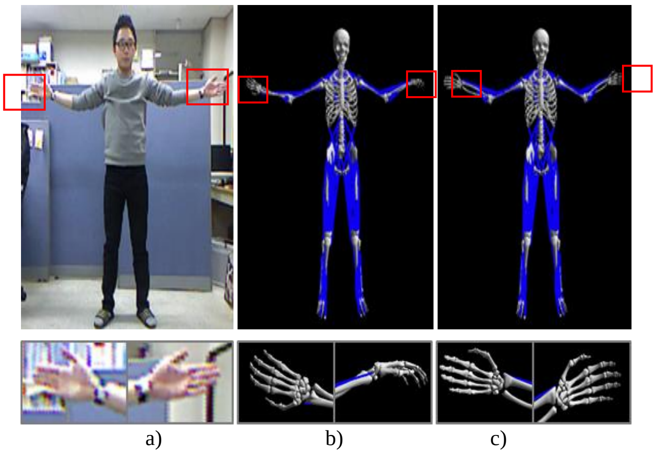



Fig. 6. Example of the musculoskeletal analysis using OpenSim. a) RGB image, b) result of the depth based and c) hybrid HPR.

## IV. Conclusion

In this paper, we present accurate 3D HPR via multi-sensor fusion approach. Our proposed methodology performs in high accuracy at its low cost and practicality. Also, the proposed approach achieves fast tracking even for complex movement by taking advantages of both depth and IMU. Our proposed approach estimates not only direction but also rotation information of the limbs for accurate HPR. We expect that our research can be utilized in various biomedical applications.

## References

[1] D. D. Luong, M. J. Lim, N. D. Thang, S. Y. Lee, and T. S. Kim, "Real-time 3D human pose recovery from a single depth image using principal direction analysis," *Appl Intell.*, vol. 41, pp. 473-486, 2014.

[2] C. M. N. Brigante, N. Abbate, and A. Basile, "Towards miniaturization of a MEMS-based wearable motion capture system," *IEEE Trans. on Industrial Electronics*, vol. 58, pp. 3234-3241, April 2011.

[3] D. Roetenberg, H. Luinge, and P. R. Slycke, "Xsens MVN: Full 6DOF human motion tracking using miniature inertial sensors," XSENS Technologies White Paper, 2009.

[4] Y. Tao, H. Hu, and H. Zhou, "Integration of vision and inertial sensors for 3D arm motion tracking in home-based rehabilitation," *The International Journal of Robotics Research*, vol. 26, pp. 607-624, 2007.

[5] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, pp. 116-124, January 2013.

[6] N. D. Thang, T. S. Kim, Y. K. Lee, and S. Y. Lee, "Estimation of 3-d human body posture via co-registration of 3-d human model and sequential stereo information," *Appl Intell*, vol. 35, no. 2, pp. 163-177, 2011.

[7] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, New York, 2008.

[8] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5-32, October 2011.

[9] Autodesk 3Ds MAX. (2012). [Online]. Available: http://www.autodesk.co.kr/

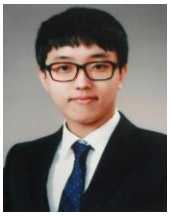
[10] CMU Mocap Database. [Online]. Available: http://mocap.cs.cmu.edu

[11] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. PAMI*, 2002.

[12] S. L. Delp, F. C. Anderson, A. S. Arnold, P. Loan, A. Habib, C. T. John, E. Guendelman, and D. G. Thelen, "OpenSim: Open-source software to create and analyze dynamic simulations of movement," *IEEE Trans. on Biomedical Engineering*, vol. 54, pp. 1940-1950, November 2007.

**S. B. Nam** received his B.S. degree in biomedical engineering from Kyung Hee University, South Korea. He is currently working toward his M.S. degree in the Department of Biomedical Engineering at Kyung Hee University, Republic of Korea. His research interests include image processing, pattern recognition, artificial intelligence, and computer vision and deep learning.

**S. U. Park** received his BE degree in biomedical engineering & computer engineering from Kyung Hee University, Republic of Korea. He is currently working toward his M.S. degree in the Department of Biomedical Engineering at Kyung Hee University, Republic of Korea. His research interests include machine learning such as deep learning, pattern recognition, artificial intelligence, image processing, and computer vision.



**J. H. Park** received his B.S. degree in biomedical engineering from Kyung Hee University, South Korea. He is currently working toward his M.S. degree in the Department of Biomedical Engineering at Kyung Hee University, Republic of Korea. His research interests include image processing, pattern recognition, artificial intelligence, and computer vision.

**M. D. Zia Uddin** received his Ph.D. degree in biomedical engineering from Kyung Hee University, Republic of Korea in 2011. He is currently working as an assistant professor in the Department of Computer Education, Sungkyunkwan University, Republic of Korea. Dr. Zia's research interest includes pattern recognition, image processing, computer vision, and machine learning.

**Tae-Seong Kim** received the B.S. degree in biomedical engineering from the University of Southern California (USC) in 1991, the M.S. degrees in biomedical and electrical engineering from USC in 1993 and 1998 respectively, and the Ph.D. in biomedical engineering from USC in 1999. After his postdoctoral work in cognitive sciences at the University of California, Irvine in 2000, he joined the Alfred E. Mann Institute for Biomedical Engineering and Dept. of Biomedical Engineering at USC as a research scientist and research assistant professor. In 2004, he moved to Kyung Hee University in Korea where he is currently an associate professor in the Biomedical Engineering Department. His research interests have spanned various areas of biomedical imaging including magnetic resonance imaging (MRI), functional MRI, E/MEG imaging, DT-MRI, transmission ultrasonic CT, and magnetic resonance electrical impedance imaging. Lately, he has started research work in proactive computing at the u-Lifecare Research Center where he serves as the vice director. Dr. Kim has published more than 80 peer reviewed papers and 150 proceedings, 6 international book chapters, and holds 10 international and domestic patents. He is a member of IEEE from 2005-2012, KOSOMBE, and Tau Beta Pi, and listed in Who's Who in the World '09-'12 and Who's Who in Science and Engineering '11-'12.