Abu Zaher Md Faridee, Sreenivasan Ramasamy Ramamurthy, Nirmalya Roy
Department of Information Systems, University of Maryland, Baltimore County

Editors: Aruna Balasubramanian and Lin Zhong

HappyFeet: Challenges in Building an Automated Dance Recognition and Assessment Tool

In this paper, we discuss our experience in building an automated dance assessment tool with IMU and IoT devices and highlight the major challenges of such an endeavor. In a typical dance classroom scenario, where the students frequently outnumber their instructors, such a system can add an immense value to both parties by providing systematic breakdown of the dance moves, comparing the dance moves between the students and the instructors, and pinpointing the places for improvement in an autonomous way. Along that direction, our prototypical work, *HappyFeet* [1], showcases our initial attempts of developing such an intelligent *Dance Activity Recognition (DAR)* system. Our CNN based Body Sensor Network proves more effective (by ≈7% margin at 94.20%) at accurately recognizing the micro-steps of the dance activities than traditional feature engineering approaches. These metrics are derived by purposely evaluating the setup on a dance form known for its gentle, smooth and subtle limb movements. In this paper, we articulate how our proposed DAR framework will be generalizable for diverse dance styles involving very pronounced movements, human body kinematics and energy profiles.

With the proliferation of wearable sensors and IoT devices over the last few years, exciting new applications are evolving every day that exploits their ubiquity for breaking new ground in health care, sports, fitness and entertainment. Similarly, an automated dance activity recognition and assessment framework poses unique opportunities and challenges. Modeled as a learning companion for aspiring dancers and a teaching aid for instructors, the detailed identification, deconstruction and comparison of the dance moves can become a great explorative tool to augment and, in some cases, supplant the traditionally crowded dance classrooms and unidirectional teaching methodology. It can provide the instructors with an in-depth breakdown of the dance moves that individual students are having difficulties performing (where and by how much) and help to give personalized feedback that is tailored towards each student's physical attributes, dexterity and training level.

We share our experience in building an early prototype, HappyFeet [1], a dance activity recognition framework, in this paper. We discuss the relevant details of sensor architecture, ground truth annotation and activity recognition model development. We specifically elaborate on the challenges we faced during the optimization of the number, types and placement of the IMU sensors, handling the heterogeneity (introduced by sensors of different make and native sampling frequency) and dealing with missing data points (from faulty sensor and application stack), synchronization of multiple sensor data streams, maintenance of granularity and precision during ground truth data annotation. We also posit the challenges in defining a scoring and comparison metric for assessing individual dance performances.

The rest of the paper is organized as follows. We start by defining the ideal characteristics of a Dance Activity Recognition (DAR) System. Next, we elaborate the hurdles and rationale for our IMU setup. We also discuss the different tools and techniques used during the annotation process. Finally, we share our insights on the scoring process.

IDEAL CHARACTERISTICS OF A DAR SYSTEM

We have investigated existing tools, frameworks and hardware solutions for dance activity recognition in our preliminary study [1], however we have yet to find a system specifically designed to handle the dance classroom environment in an automated way. Most of the existing works are primarily focused on providing entertainment value, but they are not optimized to handle a large number of participants, and they often rely on depth sensing or heavyweight pressure sensing, which restricts the participants' movement zones. We define a Dance Activity Recognition (DAR) system as: a combination of hardware and software solutions specifically designed to identify and assess the grammar, pace, sequence, directions, limb rotation and facial expressions of different dance steps (at changing granularities) with minimal human intervention. We postulate that an ideal Dance Activity Recognition (DAR) system should embody the following functional characteristics:

(i) It should minimize the number of sensors required to achieve sufficient accuracy in identifying the details of the micro-dance steps.
(ii) It should utilize commodity sensors and cameras and help minimize costly/proprietary sensor usage.
(iii) It should conduce a real-time data collection, annotation, training/inference and user-feedback pipeline with an end-to-end solution.
(iv) The DAR system should warrant minimal/no technical knowledge from the end users' perspective to achieve full runtime functionality.
(v) The sensor placement, annotation and feedback process should be minimally intrusive and distracting, so as not to obstruct the natural dance learning environment of the dancers.
(vi) The learned model should be interpretable from the instructors' perspective so that they can understand the nature and extent of the mistakes the students make to provide better feedback.

In this paper, we describe how we tried to fulfill some of these requirements during our early prototype development of HappyFeet [1] and discuss how the rest of them can also be accomplished.

OPTIMIZING OUR IMU SETUP

Dance is an art form that involves sequenced and rhythmic movements of human limbs. Because of the complexity of the movement kinematics, gestures and postures of human body and mind, and cognitive and physical variability across students, dance can be categorized as one of the most challenging human activity recognition problems. Recognizing the dance activity is fundamentally different from recognizing and learning the traditional *Activities of Daily Living (ADLs)*. Dancing requires grace and finesse, and involves repetitive movements of the fingers, hands, forearms, elbows, arms, legs, toes, waist, heads, etc., in a rhythmic fashion. It also reflects the delicacy and rhythm of different postures along with the cognitive ability and physical fitness of an individual. One step alone may consist of multiple micro steps, which span across the various movements of legs, hands, fingers, shoulders, elbows, etc. Capturing these movements with a minimal number of motion (accelerometer) and vision (camera) sensors, recognizing and delimiting these micro steps, and defining a repetitive pattern to recognize an entire dance episode are non-trivial activity recognition problems. This makes a Dance Activity Recognition (DAR) system [1] unique in its own context than in the traditional Human Activity Recognition problems [2]–[8]. While a single IMU sensor is often enough for capturing and classifying a lot of ADLs to capture acceptable details of the dance steps, we realized that more sensors were required. We also argue that the type of dance being classified plays a role in selecting the number and placement of sensors. One might be able to capture a feet-heavy dance performance (e.g., waltz) with two sensors in the ankles or just one at the hip, but for capturing a jazz dance session, those might not be adequate. In our experiment, we chose to study a classical Indian dance style: *Lasya*, a subcategory of Manipuri [9] dance form, which is noted for its gentle, smooth and subtle limb movements. Our justification for this deliberate choice was that the developed detection and assessment methodology should be applicable for dance styles involving much pronounced movements and energy profiles, which are easier to detect.

Figure 1 shows an example of a choreography that we used in one of our sessions. We designed a specific dance script for lasya, which a beginner would learn during the first few dance sessions. Four accelerometer sensors are placed on the limbs: two on the wrists and the others on both ankles. We have 10 different dance steps during a session that is described in Figure 3. As we can see, depending on the dance steps being performed, certain sensors show more pronounced activity than the other sensors. For example, for Step 1 (wave both hands from left to right), the wrist sensors show more activity than the ankle sensors, but for step 2 (step right leg forward), the left ankle sensor shows more activity than the others. We ran our experiment incrementally with different combinations of the number of sensors and placement, and concluded that for our case, at least four sensors are needed to achieve high accuracy (above 90%).

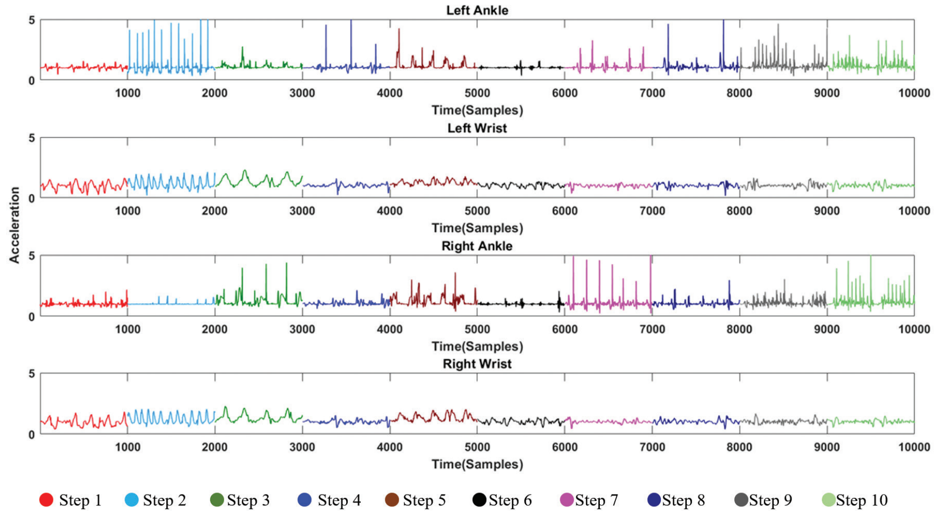


FIGURE 1. Accelerometer signals of different activities captured from four actigraph sensors.

COMPLEXITIES WITH MULTIPLE IMU SENSORS

Using multiple IMU to collect detailed data of the dance steps has its drawbacks, the first one being the added cost of the sensors. Utilization of inexpensive sensors or allowing the students to use their personal wearable devices can help lower the cost. We experimented with several off-the-shelf sensors: ActiGraph wGT3X-BT, Empatica Embrace and E4, Microsoft Band and Fitbit Flex (compared in Figure 2). This introduced a new set of challenges for us: Fitbit Flex API only allowed for data extraction of 1/60 Hz sampling frequency, so it is totally unsuitable for high speed activity detection, such as dance. Each Empatica E4 costs 1500 USD, and the Embrace only has a sampling frequency of 32Hz. The Microsoft Band failed the Zero-G test (when kept at rest it is supposed to show a constant 1g acceleration but our readings were contaminated with noise). We also found that the companion app for Microsoft Band had constant connectivity issues. The sampling rate did not stay constant throughout our data collection session and random data points would often go missing. The nature of data storage of the sensors is also very critical in assessing their usability in DAR system design. Dedicated activity trackers, such as ActiGraph wGT3X-BT, Empatica Embrace and E4, primarily store the collected data offline inside the sensor's memory. The data needs to be transferred to a PC through a physical connection and using the vendors' proprietary application stack, which impedes real-time data processing. These sensors do have Bluetooth data transfer options to smartphones via their SDK, but the licensing price can be steep. So, for most practical cases, these sensors are limited by their offline storage space. Conversely, general purpose wearable sensors, such as Microsoft Band and Fitbit, often stream data to a Bluetooth paired phone or directly in a cloud without any added cost, hence creating more

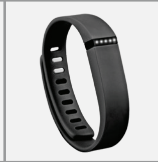
	Actigraph wGT3X-BT	Empatica Embrace	MS Band	Fitbit Flex
Devices				
Price	$225	$249	$169	$149
Max. Sampling Frequency	100Hz	32 Hz	128 Hz	1/60 Hz
Sensor Biases (Zero g test)	Passed	Passed	Failed	N/A

FIGURE 2. Comparison of sensor properties.

1. Waving both hands from left to right
2. Stepping left leg forward
3. Clockwise rapid rotational movement
4. Taking two forward steps with extended arms
5. Anti-clockwise rotational movement
6. Move both wrists to left side
7. Clockwise step-by-step slow rotation
8. Anti-clockwise step-by-step slow rotation
9. Clockwise rapid tiptoe rotation
10. Anti-clockwise rapid tiptoe rotation

FIGURE 3. Details of the dance steps.

flexibility for real-time stream analysis. At the end, we selected ActiGraph wGT3X-BT as the primary sensor for our initial experiment, as it had a high sampling rate (100Hz), passed Zero-G test, had relatively high dynamic range (±G) and had a moderate price among the sensors, but this setup also forced us to use offline processing.

Redundant data were collected by two Microsoft Bands and one Empatica E4 by placing them on wrists and the waist, so that we could analyze heterogeneity issues. We realized that when combining the inertial data collected from multiple sensors with different makes and models, three main types of Heterogeneity [10] problems can arise:

(i) *Sensor Bias*, which is caused by the difference in precision, resolution and range values of the devices.

(ii) *Sampling Rate Heterogeneity*, which occurs when at least two devices start collecting data at two different sampling rates (e.g. 100Hz vs 79Hz).
(iii) *Sampling Rate Instability*, which is the irregularity between successive timestamps of consequent data points.

In addition to these, missing data points due to faulty sensors or software can introduce heterogeneity challenges and these heterogeneities cannot be fixed easily by simple up-sampling/interpolation techniques.

USING VIDEO TO ANNOTATE ACCELEROMETER STREAMS

Annotation of dance activities is challenging, as the duration of each activity is very small, and the activity classes alternate or change frequently, unlike activities of daily living. For example, with a 60Hz sampling frequency, a micro dance step of 800 millisecond only gives 48 data points, a simple overlapped labeling by 400 milliseconds therefore translates to a 50% labeling error. Such incorrect labeling with adjacent classes drastically affect the classification accuracy, and the usage of multiple IMU compounds the errors. The hardware clock of the IMU needs to be synchronized with others down to millisecond level granularity or we run the risk of severe labeling error again. In a controlled lab experiment, this extra synchronization overhead might not be a major issue, but in a live dance classroom environment, it can become quite difficult. The next challenge lies in the utilization of proper tools to visualize the IMU data streams, so that the teachers and dancers can annotate their own dance sessions. Unlike other Activities of Daily Living, the dance micro-steps can be very specific to a choreography, the preference of the dance instructor and the training and dexterity level of the students. Due to the wide variety of possible choreography within a single dance genre, each dance lesson might require relabeling of the micro-steps even though they are thematically very similar. This low generalizability also means the activities cannot be annotated without the help of the domain experts (e.g., dance instructors or students) and frequent manual labeling needs more attention.

Even with visualizing the data points from the IMU sensors through graphs/charts, the data streams are difficult to interpret, let alone pinpoint the labeling boundaries for the individual steps. To go around this problem, we recorded each dance session using a video camera at 60 frames per seconds and let the dancers identify and annotate the dance steps from the video rather than the IMU data. After that, synchronization of the video with IMU data streams remains the only tricky part. To streamline the synchronization of the signals from each IMU sensor, the video and the time stamps associated with them, we used ELAN software [11]. At the start of each dance routine, the participants were asked to jump as high as possible three times. These three jumps showed a peak in the resultant acceleration signal. The annotator used the peaks to synchronize the sensor data stream and the video feed, all at the same time. We deduced the starting and ending frame for each of the micro level dance moves and labeled them accordingly. When doing the alignment, we noted the video and accelerometer synchronization offset. With this information, we could derive the standardized time stamps across devices and crop the data that is of interest. Because of the initial clock synchronization, all sensor samples are also properly aligned with each other in the end. Figure 4 shows the data alignment and annotation process with ELAN software [2].

CHALLENGES IN DEVELOPING A SCORING METRIC

The main modus operandi of a DAR platform is to automatically learn the dance steps from the instructor's dance routines, and then use that as a template to understand, classify and score the students' dance steps that follows. The whole process can be divided into three major steps:

(i) Accurately recognize the individual dance steps performed by the instructor.
(ii) Create a generalizable model from the instructor's model, so that it can be applied to recognize the same dance steps performed by the students.
(iii) Scoring the dances according to a well-defined criterion that can be used by the instructor to pinpoint what the students are doing wrong and provide corrective feedback.

Transferring a generalizable "knowledge" from the instructor's model to that of the students can be tricky as, due to their different physical attributes, the dances from trained dancers will vary widely with respect to correct dance moves. The algorithm needs to generalize to an extent to capture the "essence" of the dance without penalizing the natural variations due to the students' natural physical traits. This property should also be maintained when deriving the scores of the dance routines. A balanced scoring mechanism needs to take the following attributes into consideration when calculating the student scores: (i) timing, (ii) rhythms, (iii) shape of arm position while dancing, (iv) maintaining healthy position while dancing, (v) posture, body, knee & foot alignments, (vi) dynamic alignments of head, rib, and pelvis, (vii) centers of hip and balance and (viii) efficiency of movements. Dynamic alignment of the body is important for dance training, as it's easy for dancers to place themselves in alignment while they are stationary, but it becomes challenging when they are moving through space. The next critical aspect is the "timing." When the students start out a dance form, they generally try to follow a slow pace and pick up speed as they become more familiar and proficient with the steps. Both the detection and scoring algorithm need to be able to accommodate both the slower and faster version of the dance performance.

We have tried a simple scoring mechanism by training the model with the instructor's dance steps and then testing the model out on three students' performance. We calculate the confusion matrix on the 10 dance steps to get some idea on which dance steps the students are having trouble with. In the initial dance session, the students scored 23%, 12% and 26% respectively in terms of accuracy. While this crude approach is good for initial analysis, we believe that the model is too simplistic to capture the "essence" of the dance and therefore penalizes students whose movement patterns are quite different from that of the instructor. Moreover, all the students in our experiment were male, whereas the instructor was female, and we suspect that this might affect the scores of the students to some extent.

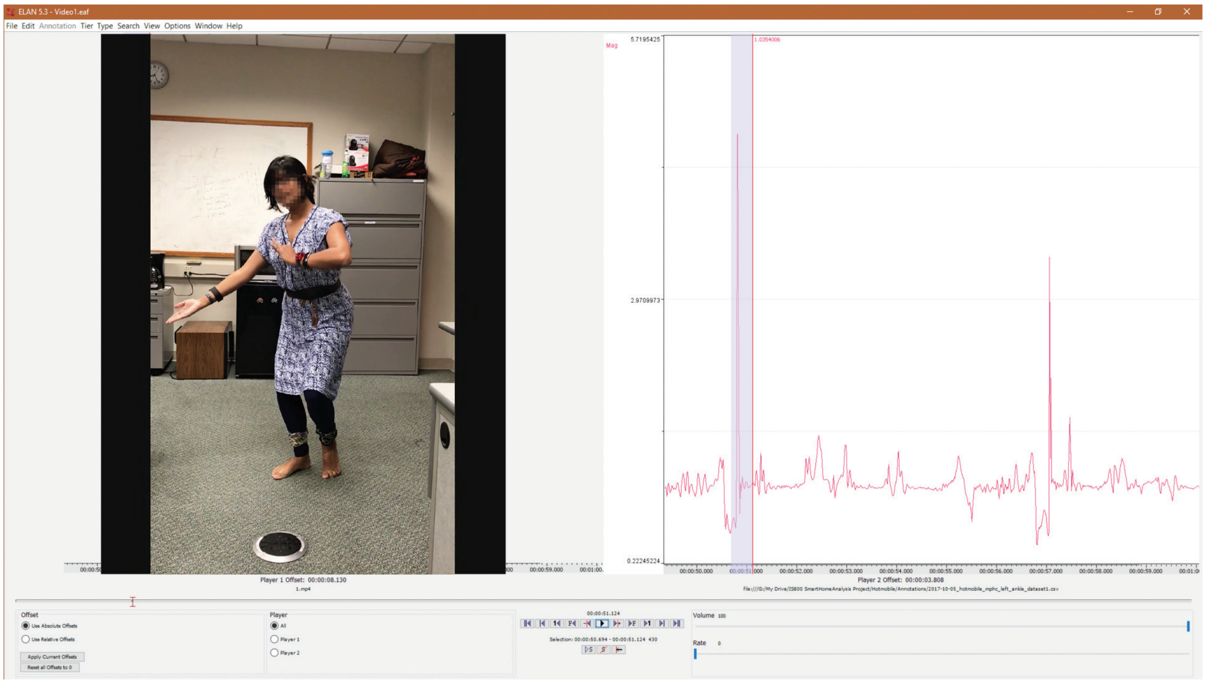


FIGURE 4. Synchronization of the video and IMU data.

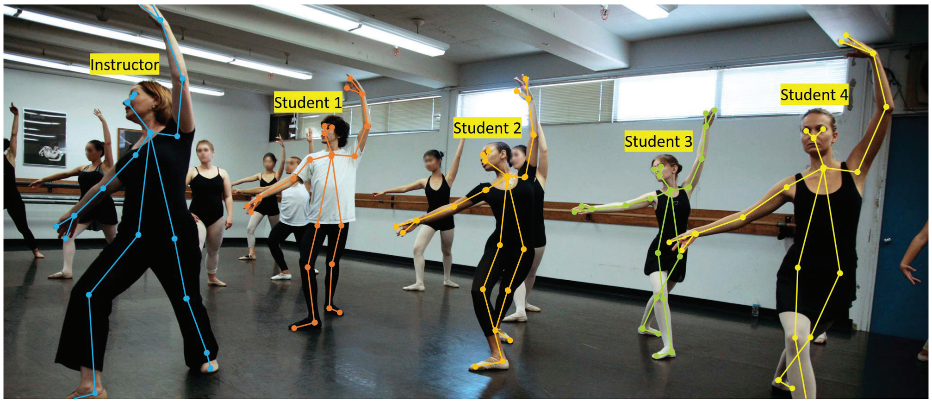


FIGURE 5. Classroom scenario of a dance session. *Courtesy of http://smc.edu/AcademicPrograms/Dance*

RECOGNIZING THE DANCE ACTIVITY IS FUNDAMENTALLY DIFFERENT FROM RECOGNIZING AND LEARNING THE TRADITIONAL *ACTIVITIES OF DAILY LIVING (ADLs)*

FUTURE CHALLENGES

To build a real-time, low-cost and low-maintenance DAR platform that can accommodate heterogeneities discussed previously, we propose a *Homogenization Module*. By collecting a large amount of unlabeled data dance performances with redundant sensors on the same limb where a costly high precision offline sensor is paired with a cheaper low precision online sensor, deep learning models can be trained to learn the intrinsic probability distribution of the dance activities. These deep models can later be used to reconstruct missing data points or remove bias/errors from the low precision sensor data streams. *Sampling Rate Instability* is analogous to image in-painting problems in computer vision domain and *Sampling Rate Heterogeneity* parallels audio and image super-resolution. Therefore, we propose to leverage well-studied models for those problems, such as Convolution-Deconvolution networks [12], and generative models, such as Restricted Boltzmann Machine (RBM)[13] and Generative Adversarial Networks [14] in our future studies.

In HappyFeet [1], we only focused on using IMU sensors; the accelerometer and gyroscopes (that are built into most modern smart watches) provide a very convenient and cost-effective way to capture dance activities. Their wearable nature gives the user unparalleled flexibility on the place and time of their performance. 3-axis acceleration and angular velocities provide a very detailed albeit local 3D view of the movement performed without much computational complexity. However, in a large classroom environment even the cost of cheaper IMU can gradually add up to a significant amount. These sensors are also very sensitive to the precise placement on the body, which might not be possible all the time during strenuous dance sessions. Hence, we believe that a secondary modality of data capture mechanism is required to build a cost-effective and reliable DAR system. Depth sensing has been traditionally used for capturing 3D movement data and human action recognition for many years. Marker based motion capture and depth inference from stereo cameras can be expensive and not practical for a classroom scenario. Infrared depth sensing (e.g., Microsoft Kinect) provides a very reliable way to capture 3D human performance data and is capable of tracking up to six human skeletal motions in a 70.6 by 60 degrees FOV in real time, but these values are still inadequate for a large classroom environment and the setup is still expensive. Recent advances in deep learning–based pose estimation models, such as *OpenPose* [15] and *DensePose* [16] has shown that it is possible to create real time posture key-point detection and tracking with commodity hardware, at a considerably lower infrastructural cost. An appropriately placed single camera can capture the dance performance of a roomful of people and the inferred key points (skeletal joints shown in Figure 5, and the resulting skeletal model can be used to accurately classify a broad range of human activities [17]. However, vision-based approaches, especially with single camera, have their own limitations, such as sensitivity to lighting conditions, occlusion, background clutter and limited field of view. The issues can be mitigated by using multiple cameras, but that increases the setup complexity and cost. Most commodity video cameras rarely record at more than 60 frames per second, while most inertial sensors can easily go beyond 100Hz. Vision and depth-based models can estimate the whole-body posture with great efficiency, however, it cannot estimate the micro level movements of the limbs with high accuracy, something the inertial sensor-based systems excel at [18]. Therefore, we conclude that a multi modal view of the dance moves that efficiently fuses complementary input from visual and inertial sensors would be the optimum solution when assessing dance moves in a group setting. There are multiple ways to build the fusion model such as:

(i) Using the IMU data streams to create a calibrated vision model;
(ii) Tracking the 3d skeletal joints and using those to augment IMU streams;
(iii) Concatenating the visual and inertial features in a joint model.

As ELAN is a specialized software targeted at researchers, it is not intuitive enough to be used by the dance instructors and students by themselves. Also, this annotation process is offline; there is significant delay between the dance capture and feedback generation for the dancers, which can be an additional overhead and source of distraction for them. The tedious manual labeling can take a long time depending on the length of the dance session. To streamline the process of annotation by the dancers and make the ordeal real-time, we propose a smartphone-based *Annotation Module*. Most online IMU sensors work by Bluetooth pairing with smartphones. As we integrate more of these sensors into HappyFeet (as suggested in the previous section), we can directly visualize these activities on the dancers' smartphones. A wide-angle camera (which is part of the *Multi-modal Fusion Module*) can capture the dance moves of the whole class; the video from it can be embedded in the synchronization view with the iOS/Android app. This is effectively creating a smartphone specific, real-time version of the ELAN interface optimized for the dancers' workflow that we show in Figure 4. We argue that such an application can significantly reduce the inconveniences of the annotation process. The functionality of the app can be further augmented by *Change-point Detection* [19], *Semi-supervised* and *Active Learning* [5], [20] techniques. For example, after a dance session has been designed, captured and annotated by the instructor, the dance template can be used to generate annotation suggestions for the students in a semi-supervised way. Active Learning methods can be leveraged to query the users to provide annotations for only where the algorithm has a low confidence rating. As the system gathers more data from the dance sessions over time, the suggestions will have a higher confidence rating and users will query less for annotation clarifications.

CONCLUSION

In this paper, we discussed our in-depth experiment of developing a DAR platform for a dance classroom scenario. IMU sensors give a lot of flexibility on the dance capture process, but for large scale deployment, we believe that integration with a vision-based approach should be the way to go forward. Device heterogeneity is an issue that we are actively working on to rectify; automatic segmentation and annotation of the dance sessions remains a challenge as well. Without a large amount of annotated data, it is very difficult for the deep learning–based models to extract knowledge of dances to build a generalizable model to calculate the scores effectively. We also wish to explore and validate the qualitative aspect of such a framework with feedback from the dancers; the framework should always be designed so that it never gets in the way of the natural dance learning environment of the students. ■

Abu Zaher Md Faridee is currently a PhD student in the Department of Information Systems at the University of Maryland, Baltimore County. His primary research areas are human activity recognition, active and transfer learning and IoT. Before joining UMBC, he contributed in multiple open source projects for The National Evolutionary Synthesis Center and Apertium Machine Translation Platform through the Google Summer of Code program. He completed his undergraduate degree from Bangladesh University of Engineering and Technology.

Sreenivasan Ramasamy Ramamurthy is a PhD student in the Department of Information Systems at University of Maryland Baltimore County. He received his Master's in Biomedical Engineering from Vellore Institute of Technology, India. His research interests include signal processing, machine learning, and multi-inhabitant activity recognition.

Nirmalya Roy is currently an associate professor in the Information Systems department at University of Maryland Baltimore County, where he directs the Mobile, Pervasive, and Sensor Computing (MPSC) Lab. His research spans vast fields of pervasive computing and middleware, wireless sensor networks and smart environments with a primary focus on sensor-driven human activity recognition, smart healthcare, and building energy analytics. He is a recipient of the 2018 NSF CAREER award.

REFERENCES

[1] A. Z. M. Faridee, S. Ramasamy Ramamurthy, H. M. S. Hossain, and N. Roy, "HappyFeet: Recognizing and assessing dance on the floor," in *Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications - HotMobile '18*, 2018, pp. 49–54.

[2] S. Ramasamy Ramamurthy and N. Roy, "Recent trends in machine learning for human activity recognition – a survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, p. e1254, Jul. 2018.

[3] M. A. U. Alam, N. Roy, A. Misra, and J. Taylor, "CACE: Exploiting behavioral interactions for improved activity recognition in multi-inhabitant smart homes," in *2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*, 2016, pp. 539–548.

[4] N. Roy, A. Misra, and D. Cook, "Ambient and smartphone sensor assisted ADL recognition in multi-inhabitant smart environments," J. *Ambient Intell. Humaniz. Comput.*, vol. 7, no. 1, pp. 1–19, Feb. 2016.

[5] H. M. S. Hossain, M. A. A. H. Khan, and N. Roy, "Active learning enabled activity recognition," *Pervasive Mob. Comput.*, vol. 38, pp. 312–330, Jul. 2017.

[6] N. Roy, A. Misra, and D. Cook, "Infrastructure-assisted smartphone-based ADL recognition in multi-inhabitant smart environments," in *2013 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2013, pp. 38–46.

[7] M. A. U. Alam, N. Roy, S. Holmes, A. Gangopadhyay, and E. Galik, "Automated functional and behavioral health assessment of older adults with dementia," in *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 2016, pp. 140–149.

[8] M.A.H. Khan, N. Roy, and A. Misra, "Scaling human activity recognition via deep learning-based domain adaptation," *2018 IEEE Int. Conf. Pervasive Comput. Commun. Athens, Greece, March 19-23*, Mar. 2018.

[9] R. Massey, *India's dances : their history, technique, and repertoire*. Abhinav Publications, 2004.

[10] A. Stisen et al., "Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition," *Proc. 13th ACM Conf. Embed. Networked Sens. Syst. – SenSys '15*, 2015.

[11] H. Lausberg and H. Sloetjes, "Coding gestural behavior with the NEUROGES-ELAN system," *Behav. Res. Methods*, vol. 41, no. 3, pp. 841–849, Aug. 2009.

[12] V. Kuleshov, S. Z. Enam, and S. Ermon, "Audio super resolution using neural networks," Aug. 2017.

[13] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," Springer, Berlin, Heidelberg, 2012, pp. 599–619.

[14] I. Goodfellow et al., "Generative adversarial nets." pp. 2672–2680, 2014.

[15] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields *."

[16] R. A. Güler, N. Neverova, and I. Kokkinos, "DensePose: Dense human pose estimation in the wild."

[17] W. Zhu et al., "Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks," *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. AAAI Press, pp. 3697–3703, 2016.

[18] C. Chen, R. Jafari, and N. Kehtarnavaz, "A survey of depth and inertial sensor fusion for human action recognition," *Multimed. Tools Appl.*, vol. 76, no. 3, pp. 4405–4425, Feb. 2017.

[19] D. Picard, "Testing and estimating change-points in time series," *Adv. Appl. Probab.*, vol. 17, no. 04, pp. 841–867, Dec. 1985.

[20] H. M. S. Hossain, M. A. Al Haiz Khan, and N. Roy, "DeActive: Scaling activity recognition with active deep learning," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 2, pp. 1–23, Jul. 2018.