Fusion of Inertial and Depth Sensor Data for Robust Hand Gesture Recognition

Kui Liu, *Student Member, IEEE*, Chen Chen, *Student Member, IEEE*, Roozbeh Jafari, *Senior Member, IEEE*, and Nasser Kehtarnavaz, *Fellow, IEEE*

***Abstract*— This paper presents the first attempt at fusing data from inertial and vision depth sensors within the framework of a hidden Markov model for the application of hand gesture recognition. The data fusion approach introduced in this paper is general purpose in the sense that it can be used for recognition of various body movements. It is shown that the fusion of data from the vision depth and inertial sensors act in a complementary manner leading to a more robust recognition outcome compared with the situations when each sensor is used individually on its own. The obtained recognition rates for the single hand gestures in the Microsoft MSR data set indicate that our fusion approach provides improved recognition in real-time and under realistic conditions.**

***Index Terms*— Sensor fusion, fusion of inertial and depth sensor data, hand gesture recognition.**

I. Introduction

THE literature includes a large collection of works where either vision sensors or inertial body sensors have been used for measurement or recognition of human body movements. Each of the above two sensors has been used individually for body movement measurements and recognition. However, each sensor has its own limitations when operating under realistic conditions. The major contribution of this paper is the fusion of data from two different modality sensors that are captured at the same time. The two utilized sensors of vision depth sensor and inertial body sensor are used in a complementary manner where erroneous data that may get generated by each individual sensor are compensated by the other sensor. In other words, the introduced fusion approach involves the fusion of data from a cost-effective inertial body sensor and a cost-effective vision depth sensor in order to achieve more robust hand gesture recognition in real-time compared to the situations when these sensors are used individually. The focus of this paper is on hand gesture recognition. However, it should be noted that the introduced approach in this paper is general purpose in the sense that it is applicable to other body movement applications.

As far as vision sensors are concerned, comprehensive reviews on hand pose estimation or hand gesture recognition have previously appeared in [1]–[3]. Two major matching techniques have been deployed for hand gesture recognition. These techniques include Dynamic Time Warping (DTW) [4] and Elastic Matching (EM) [5]. Statistical modeling techniques such as particle filtering [6], [7], and hidden Markov model (HMM) [8] have also been utilized for hand gesture recognition. The application of depth sensors, in particular Kinect [9], has been steadily growing for body movement measurements and recognition. Several studies utilizing the depth sensor Kinect have been reported in the literature for hand gesture recognition. For example, in [10], depth images captured by Kinect were used to achieve American Sign Language (ASL) recognition. In [11], both depth and color information captured by Kinect were used to achieve hand detection and gesture recognition. In [12], a HMM was trained to identify the dynamic gesture trajectory of seven gestures using the Kinect sensor.

As far as inertial body sensors are concerned, many body measurement and recognition systems involving such sensors have appeared in the literature. For example, a human body motion capture system using wireless inertial sensors was presented in [13]. In [14], wireless body sensors were used to recognize the activity and position of upper trunk and lower extremities based on a DTW-based hierarchical classifier. In [15], a customizable wearable body sensor system was introduced for medical monitoring and physical rehabilitation. In [16], a support vector machine (SVM) classifier was used as part of a body sensor network to estimate the severity of Parkinsonian symptoms. In [17], Kalman filtering in a body sensor network was used to obtain orientations and positions of body limbs.

The simultaneous utilization of both inertial body sensor and depth sensor that have appeared in the literature have been studied for the registration of images [18], [19], the estimation of the position and orientation of a camera [20], and the use of gravity to recover the focal distance of a camera [21]. In [22], an angle estimation approach involving both an inertial sensor and a Kinect sensor was discussed where Kalman filtering was applied to correct or calibrate the data drifting of the inertial sensor. Our fusion approach presented in this paper differs from all the previous works in the sense that both inertial and depth sensor data are used at the same time and together as the input to a probabilistic classifier in order to increase the robustness of recognition. Another attribute of our approach is that the computational complexity is kept low so that its real-time implementation is made possible. Furthermore, it is worth noting that these

Manuscript received January 10, 2014; accepted February 10, 2014. Date of publication February 12, 2014; date of current version April 16, 2014. The associate editor coordinating the review of this paper and approving it for publication was Dr. M. R. Yuce.

The authors are with the Department of Electrical Engineering, University of Texas at Dallas, Dallas, TX 75080 USA (e-mail: kxl105220@utdallas.edu; cxc123730@utdallas.edu; rjafari@utdallas.edu; nxk019000@utdallas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSEN.2014.2306094

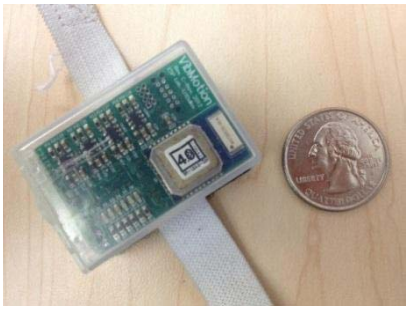

Fig. 1. Wireless inertial sensor.

two differing modality sensors that are deployed are both cost-effective which makes their joint utilization practical in various applications. Our approach uses the HMM classification as this classifier has been proven effective in various recognition applications due to its probabilistic framework.

In section II, a brief overview of the Kinect and inertial sensor used is mentioned. In section III, the details of our fusion approach are presented. The results obtained are then reported in section IV. This section also includes a comparison with the situations when the sensors were used individually. Finally, the conclusion is stated in section V.

II. Overview of Kinect and Inertial Sensor

Kinect is a low-cost RGB-Depth sensor introduced by Microsoft for human-computer interface applications. Two software packages are publically available for this sensor (OpenNi/NITE and Kinect SDK) that allow gesture and movement recognition. The introduction of Kinect has led to successful recognition in many applications including video games, virtual reality and gesture recognition.

Fig. 1 shows a 9-axis wireless body sensor having a size of $1'' \times 1.5''$ that was designed and built in the ESSP Laboratory at the University of Texas at Dallas [23]. It consists of (i) an InvenSense 9-axis MEMS (micro-electro-mechanical system) sensor MPU9150 which captures 3-axis acceleration, 3-axis angular velocity and 3-axis magnetic strength data, (ii) a Texas Instruments 16-bit low power microcontroller MSP430 which provides data control, (iii) a dual mode Bluetooth low energy unit which streams data wirelessly to a laptop/PC, and (iv) a serial interface between MSP430 and MPU9150 enabling control commands from the microcontroller to the MEMS sensor and data transmission from the MEMS sensor to the microcontroller. For the magnetometer to provide an accurate reference, a controlled magnetic field without any distortion is required. Thus, for the application reported here, the data consisting of the accelerometer and the gyroscope were used since a controlled magnetic field is not normally available in practice.

III. Developed Fusion Approach

A. Resampling and Filtering

The sampling rates of the Kinect and inertial sensor used are 30 Hz and 200 Hz, respectively. Thus, in order to fuse the data from these two sensors, the inertial sensor data is

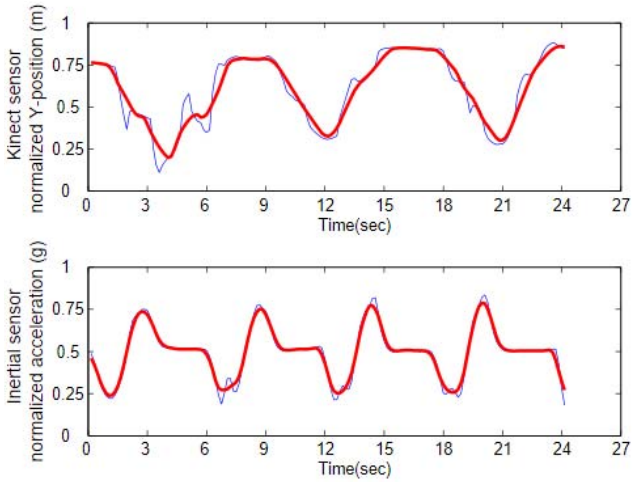

Fig. 2. Raw signal versus filtered signal: (top) Kinect normalized Y-coordinate signal and (bottom) inertial sensor normalized Z-gyro signal.

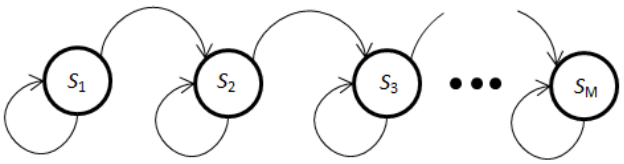

Fig. 3. Left-right HMM topology.

downsampled to match the sampling frequency of the Kinect. The downsampling is performed as follows. The Kinect signal samples are collected through the Kinect SDK software at the rate of 30Hz. Whenever this software indicates the presence of a Kinect signal sample, a signal sample from the inertial sensor gets collected at that time. This approach allows the synchronization of the signals from the Kinect and the inertial sensor. In other words, inertial samples at the rate of 200HZ closest to Kinect samples at the rate of 30Hz are considered to form the inertial sensor signals.

Because of the presence of various noise sources in an actual operating environment, jitters often appear in the Kinect skeleton signals as well as in the inertial signals. A moving average window is thus used in order to reduce jitters in the signals. Based on experimentations, it was found that a moving window of size between 9 and 19 generated high recognition rates by adequately reducing jitters in the signals. Fig. 2 shows an example of the raw and filtered signals from the Kinect and inertial sensor.

B. HMM Classifier

HMM has been used extensively to model random processes. The HMM model characterizes a state transfer probability matrix A and an observation symbols probability matrix B. Given an initial state matrix π, an HMM is described by the triplet $\lambda = \{\pi, A, B\}$. Since hand gesture recognition involves temporal signal sequences, a left-right HMM topology is adopted here, see Fig. 3.

An overview of the HMM equations are provided in this section. More details on HMM can be found in many

references, see [24]. Suppose a random sequence $O = \{O_1, O_2, \ldots, O_T\}$ is observed; let $V = \{v_1, v_2, \ldots, v_T\}$ denote all possible outcomes and let $S = \{S_1, S_2, \ldots, S_M\}$ denote all HMM states with q_t representing the state at time t, where T indicates the number of time samples. The three components of the HMM model π, A, B are computed by the following equations:

$$\pi = \{p_i = P(Q_1 = S_i)\}, 1 \le i \le M; \quad (1)$$

$$A = \{a_{ij} = P(q_t = S_j | q_{t-1} = S_i)\}, \quad 1 \le i, \quad j \le M; \quad (2)$$

$$B = \{b_j(k) = P(O_t = v_k | q_t = S_j)\}, \quad 1 \le j \le M, \quad 1 \le k \le T; \quad (3)$$

where

$$\sum_{i=1}^{M} \pi_i = 1, \quad \sum_{j=1}^{M} a_{ij} = 1 \text{ and } \sum_{k=1}^{T} b_j(k) = 1 \quad (4)$$

For HMM training, the above components need to be initialized. Among all the initialization matrices, the initialization of the transition matrix A is of importance here. By zeroing out all the non-adjacent probabilities in this matrix, the state transitions are made limited to the sequence of adjacent states, thus representing a hand gesture. That is to say, for our hand gesture recognition application, all possible state transitions are constrained to only occur from left-to-right and between two adjacent states. The initial transition matrix A is thus considered to be

$$A = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

Let $O = \{O_1, O_2, \ldots, O_T\}$ be the observation sequence of a hand gesture, $Q = \{q_1, q_2, \ldots, q_T\}$ be the corresponding state sequence with the probability of the observation sequence O obtained by this equation

$$P(O|Q, \lambda) = \prod_{t=1}^{T} P(O_t | q_t, \lambda) \quad (6)$$

According to the Baum-Welch algorithm [24], the equation below represents the probability of the observation O at time t, where π denotes the initial state probabilities,

$$P(O|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} a_{q_3 q_4} \cdots a_{q_{T-1} q_T} \quad (7)$$

In every time step 1 through T, this probability is updated as follows:

$$P(O|\lambda) = \sum_{Q} P(O|Q,\lambda) P(Q,\lambda) = \sum_{q_1, q_2, \ldots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \ldots a_{q_{T-1} q_T} b_{q_T}(O_T) \quad (8)$$

Let the updated HMM model be $\bar{\lambda} = \{\overline{\pi}, \overline{A}, \overline{B}\}$ and let the probability of the joint event that the sequence $O_1, O_2, \ldots, O_t$ is observed be $\alpha_t(i)$, thus

$$\alpha_t(i) = P(O_1, O_2, \ldots, O_T, \quad q_T = S_i | \lambda) \quad (9)$$

Algorithm HMM training

Input: Observation sequence $O = \{O_i\}_{i=1}^{T}$, state sequence $Q = \{q_i\}_{i=1}^{T}$, initial triplet $\lambda_0 = \{\pi_0, A_0, B_0\}$

while not the last observation **or** $\log\{P(O|\lambda)\} - \log\{P(O|\bar{\lambda})\} < \varepsilon$ **do**

Calculate $P(O|\lambda)$ according to Eq.(7)

Calculate $\bar{\lambda} = \{\overline{\pi}, \overline{A}, \overline{B}\}$ according to Eqs.(12), (13) and (14)

end while

$\lambda = \bar{\lambda}$

Output: triplet λ

Fig. 4. HMM training algorithm.

In a backward way, let

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \ldots, O_T, q_T = S_i | \lambda) \quad (10)$$

The probability being in state S_i at time t and state S_j at time $t+1$ is thus given by

$$\xi_t(i, j) = P\left(q_t = S_i, q_{t+1} = S_j | O, \lambda\right) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \quad (11)$$

By letting $\gamma_t(i)$ be the probability of being in state S_i at time t, one gets $\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i, j)$, and the updated model $\bar{\lambda} = \{\overline{\pi}, \overline{A}, \overline{B}\}$ is expressed as follows:

$$\overline{\pi}_i = \gamma_t(i) \quad (12)$$

$$\overline{a_{ij}} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (13)$$

$$\overline{b_j}(k) = \frac{\sum_{t=1, O_t = v_k}^{T-1} \gamma_t(j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (14)$$

By considering a very small threshold value, e.g. $\varepsilon = 10^{-6}$, when $log\{P(O|\lambda)\} - log\{P(O|\overline{\lambda})\} < \varepsilon$, the training is terminated. An algorithmic description of the training process is shown in Fig. 4.

A test or validation sequence is then fed into five trained HMM models each corresponding to a hand gesture in order to calculate the probabilities. Then, a high (e.g., 95%) confidence interval is applied to the five probabilities to classify the sequence. Let μ and σ represent the mean and variance of the probabilities. For the 95% confidence interval, whenever none of the five probabilities is larger than $\mu + 1.96\frac{\sigma}{\sqrt{n}}$, where n denotes the number of gestures, the sequence is rejected and the gesture is considered to be a not-done-right gesture. If the sequence is not rejected, the gesture with the maximum probability is considered to be the recognized gesture. The testing process is illustrated in Fig. 5.

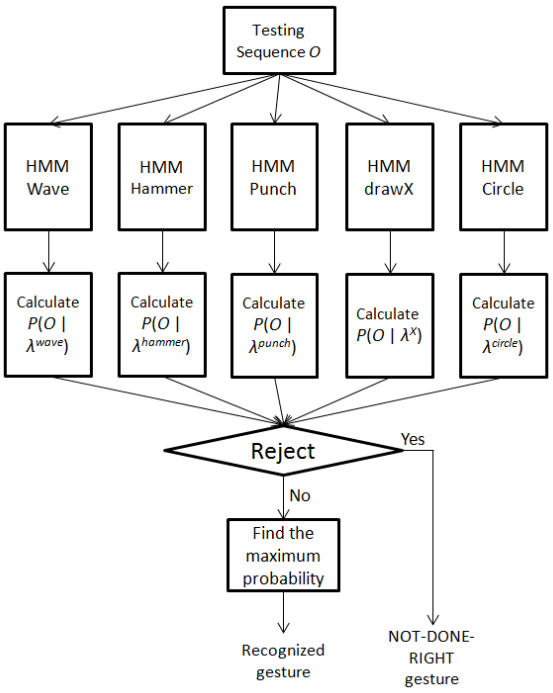

Fig. 5. Flowchart of HMM testing or recognition.

Fig. 6. Experimental setup.

TABLE I
HAND GESTURE RECOGNITION RATES (%) WHEN USING DATA FROM KINECT ONLY

	wave	hammer	punch	drawX	circle	reject
wave	81	3	3	8	5	0
hammer	2	90	6	0	1	1
punch	6	4	83	1	2	4
drawX	16	1	1	69	12	1
cirlce	6	2	0	1	91	0
Incomplete/other gestures	1	2	6	1	0	90

IV. RECOGNITION RESULTS AND DISCUSSION

Extensive experimentations were carried out to show the increase in robustness when the data from the two differing modality sensors were used together as the input to the HMM classifier compared to when using a single sensor individually. The code is written in C running in real-time on a PC platform with a quad core 1.7GHz processor and 4G memory. The input signals were captured with a Microsoft Kinect sensor and the wireless sensor described in section II. The wireless sensor was placed and tied to a subject's wrist with the subject staying within the operating distance of the Kinect, that is within a distance of 1.2m-3.5m from the Kinect sensor, see Fig. 6.

We considered the five single hand gestures present in the Microsoft MSR dataset [25]. These gestures are illustrated in Fig. 7. Ten subjects were asked to perform these five gestures 30 times in front of different backgrounds. Different backgrounds included different scenes appearing in different lighting conditions including outdoor day light, indoor flores-cent and indoor incandescent lights. Each subject performed the gestures at different speeds which were timed to last between 1 to 3 seconds. The number of the HMM states were considered to be between 8 to 12 as this range of states allowed covering all the major transitions in the training sequences.

The 3-axis accelerometer and the 3-axis gyroscope signals from the wireless inertial sensor and the 3-axis $\{X, Y, Z\}$ coordinates signals from the Kinect sensor were captured in real-time and simultaneously to form the observation sequence $O = \{O_1, O_2, \ldots, O_T\}$ of the HMM classifier. The signals from the accelerometers denoted speed changes associated with linear motions along the three directions X, Y, and Z and the signals from the gyroscopes denoted rotational motion speeds about the three axes X, Y, and Z while the signals from the Kinect denoted the 3-D coordinates X, Y, and Z of the centroid of the hand depth blob. Each hand gesture training sample can be viewed as a 9-dimensional feature vector where all the above components are fused together (3 dimensions for gyroscope, 3 dimensions for accelerometers and 3 dimensions for Kinect position coordinates). 30 variations of a hand gesture made by each subject were considered. The data from 9 subjects were used for training and the data form the remaining subject was used for testing. The recognition process was repeated 10 times, each time choosing a different set of 9 training subjects. The testing outcomes were then averaged. In other words, the training data consisted of a 3-dimensional matrix of size T*270*9. For the incom-plete/other gesture category, 100 gestures were performed by different subjects with 50 of them done in an incomplete way and with the other 50 done differently than the five gestures. The similarity between a test or validation feature vector with all the training classes or gestures was then obtained via the HMM probabilistic classification.

The recognition rates obtained are shown in the form of three confusion matrices in Tables I through III for the five gestures of "wave," "hammer," "punch," "drawX," "circle". Table I corresponds to the situation when using just the Kinect sensor, Table II when using just the inertial sensor, and Table III when using both of the sensors together.

TABLE II
HAND GESTURE RECOGNITION RATES (%) WHEN USING DATA FROM INERTIAL SENSOR ONLY

	wave	hammer	punch	drawX	circle	reject
wave	85	2	5	3	5	0
hammer	0	91	5	3	1	0
punch	0	6	94	0	0	0
drawX	4	3	5	76	12	0
cirlce	4	3	2	0	91	0
Incomplete/other gestures	0	1	1	4	2	92

TABLE III
HAND GESTURE RECOGNITION RATES (%) WHEN FUSING DATA FROM KINECT AND INERTIAL SENSOR

	wave	hammer	punch	drawX	circle	reject
wave	92	1	1	5	1	0
hammer	5	91	2	2	0	0
punch	3	5	91	0	0	1
drawX	0	0	6	88	6	0
cirlce	1	0	0	0	99	0
Incomplete/other gestures	0	1	1	1	0	97

TABLE IV
AVERAGE RECOGNITION RATES (%) FOR HMM AND DTW

	Kinect	Inertial	Fusion
DTW	69	71	80
HMM	84	88	93

Fig. 7. Five hand gestures examined: "wave," "hammer," "punch," "drawX," and "circle."

As can be seen from Table I, a relatively low recognition rate (83%) was obtained for the gesture "punch". This was caused due to the low depth resolution of the Kinect sensor for this gesture. While for the gesture "drawX," a lower recognition rate (69%) was obtained. The errors were traced back to the hand blobs crossing the face area creating an overlap of the face and hands in the depth map, thus leading to severe jitters in the wrist and hand joints. In Table II, it is seen that

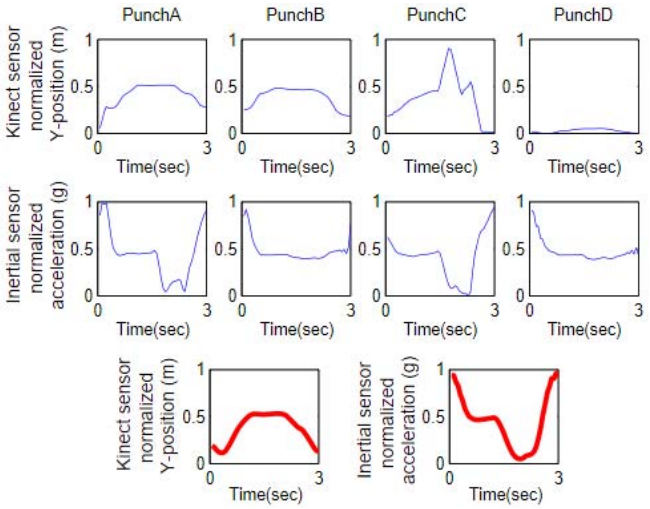

Fig. 8. Examples illustrating the complementary nature of the signals from the two sensors: (top) Kinect Y-coordinate position signal expressed in meters, (middle) inertial sensor Z-gyro signal expressed in g (9.8m/s^2), and (bottom) reference Kinect and inertial signals.

although the inertial sensor provided higher recognition rate for this gesture, the recognition rate was still not high (76%). However, as seen from Table III, by combing the data from the Kinect and the inertial sensor, in particular for the two gestures of "wave" and "drawX," the overall recognition rate was improved by 15% over the Kinect sensor alone and by 10% over the inertial sensor alone. This clearly demonstrated the complementary nature of data from these two differing modality sensors.

Essentially, jitters in the signals from the two sensors are caused by different sources. For example, the Kinect sensor exhibits difficulty when the tracking is lost and the inertial sensor exhibits difficulty when it is worn differently on the wrist and due to signal drifts. Examples are shown in Fig. 8 to illustrate the complementary aspect of the data from the two sensors. The expected punch signal is exhibited on the bottom pane of the figure (thick curve), which was created by taking the average of the training signals. In the example punch A, the test signals (thin curves) from both the Kinect and the inertial sensor exhibited a good match to the expected signals. While in the example punch B, the inertial sensor signal produced no significant orientation information, leading to a failure when using this sensor alone. However, with the HMM getting its input from the Kinect sensor, a correctly recognized probability could still be achieved. In the example punch C, the Kinect signal suffered from skeleton or joint jitters while the signal from the inertial sensor still allowed achieving a correctly recognized probability. In the example punch D, both the Kinect and the inertial sensor signals failed to provide a correctly recognized probability.

In another experimentation, the HMM classifier was replaced by a Dynamic Time Warping (DTW) classifier noting that DTW has been widely used for human body movement recognition. Table IV provides the comparison between HMM and DTW. As can be seen from this table, the HMM classifier provided higher recognition rates compared to the DTW classifier primarily due to the DTW not being able to

adequately cope with scale invariance. However, still in this classification, the fusion of the data from the two sensors provided a higher recognition rate compared to the individual sensor cases.

V. Conclusion

In this paper, a data fusion approach to hand gesture recognition based on the probabilistic HMM classification was introduced. It was shown that fusing or merging the data from two differing modality sensors, consisting of an inertial sensor and a vision depth sensor, led to an overall recognition rate of 93% for five motional hand gestures under realistic conditions such as different gesture speeds and backgrounds. This recognition rate was higher than when using each sensor individually on its own. For future works, considering that the introduced fusion framework involving the two sensors of inertial body sensor and Kinect depth sensor is general purpose, this framework will be applied to other human body movement applications.

References

[1] A. Erol, G. Bebis, M. Nicolescu, R. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," *J. Comput. Vis. Image Understand.*, vol. 108, nos. 1–2, pp. 52–73, Oct. 2007.

[2] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Trans. Syst., Man, Cybern., C, Appl. Rev.*, vol. 37, no. 3, pp. 311–324, May 2007.

[3] G. Murthy and R. Jadon, "A review of vision based hand gesture recognition," *Int. J. Inf. Technol. Knowl. Manag.*, vol. 2, no. 2, pp. 405–410, Jul./Dec. 2009.

[4] K. Liu and N. Kehtarnavaz, "Real-time robust vision-based hand gesture recognition using stereo images," *J. Real-Time Image Process.*, Feb. 2013, doi: 10.1007/s11554-013-0333-6, print to appear in 2014.

[5] S. Uchida and H. Sakoe, "A survey of elastic matching techniques for handwritten character recognition," *IEICE Trans. Inf. Syst.*, vol. E88-D, no. 8, pp. 1781–1790, Aug. 2005.

[6] P. Djuric, M. Vemula, and M. Bugallo, "Target tracking by particle filtering in binary sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2229–2238, Jun. 2008.

[7] P. Pan and D. Schonfeld, "Video tracking based on sequential particle filtering on graphs," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1641–1651, Jun. 2011.

[8] H. Lee and J. Kim, "An HMM-based threshold model approach for gesture recognition," *IEEE Trans. Pattern Recognit. Mach. Intell.*, vol. 21, no. 10, pp. 961–973, Oct. 1999.

[9] C. Chen, K. Liu, and N. Kehtarnavaz, "Real-time human action recognition based on depth motion maps," *J. Real-Time Image Process.*, Aug. 2013, doi: 10.1007/s11554-013-0370-1, print to appear in 2014.

[10] C. Keskin, F. Kirac, Y. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Barcelona, Spain, Nov. 2011, pp. 1228–1234.

[11] Z. Ren, J. Meng, J. Yuan, and Z. Zhang, "Robust hand gesture recognition with kinect sensor," in *Proc. ACM Int. Conf. Multimedia*, Scottsdale, AZ, USA, Dec. 2011, pp. 759–760.

[12] Y. Wang C. Yang X. Wu, and S. Xu, "Kinect based dynamic hand gesture recognition algorithm research," in *Proc. IEEE Int. Conf. Intell. Human Mach. Syst. Cybern.*, Nanchang, China, Aug. 2012, pp. 274–279.

[13] Z. Zhang, Z. Wu, J. Chen, and J. Wu, "Ubiquitous human body motion capture using micro-sensors," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun.*, Galveston, TX, USA, Mar. 2009, pp. 1–5.

[14] L. Wang, T. Gu, H. Chen, X. Tao, and J. Lu, "Real-time activity recognition in wireless body sensor networks: From simple gestures to complex activities," in *Proc. IEEE Int. Conf. Embedded Real-Time Comput. Syst. Appl.*, Macau, China, Aug. 2010, pp. 43–52.

[15] M. Zhang and A. Sawchuk, "A customizable framework of body area sensor network for rehabilitation," in *Proc. IEEE Int. Symp. Appl. Sci. Biomed. Commun. Technol.*, Bratislava, Slovakia, Nov. 2009, pp. 1–6.

[16] S. Patel, K. Lorincz, R. Hughes, and N. Hugguns, "Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 6, pp. 864–873, Nov. 2009.

[17] R. Zhu and Z. Zhou, "A real-time articulated human motion tracking using tri-axis inertial/magnetic sensors package," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no. 2, pp. 295–302, Jun. 2004.

[18] E. Foxlin, Y. Altshuler, L. Naimark, and M. Harrinton, "Flight tracker: A novel optical/inertial tracker for cockpit enhanced vision," in *Proc. IEEE ACM Int. Symp. Mixed Augmented Reality*, Nov. 2004, pp. 212–221.

[19] J. Hol, "Sensor fusion and calibration of inertial sensors, vision, ultra-wideband and GPS," Ph.D. dissertation, Division Autom. Control, Linkoping Univ., Linköping, Sweden, 2011.

[20] J. Hol, B. Schon, F. Gustafsson, and P. Slycke, "Sensor fusion for augmented reality," in *Proc. IEEE Int. Conf. Inf. Fusion*, Florence, Italy, Jul. 2006, pp. 1–6.

[21] J. Lobo and J. Dias, "Vision and inertial sensor cooperation using gravity as a vertical," *IEEE Trans. Pattern Recognit. Mach. Intell.*, vol. 25, no. 12, pp. 1597–1608, Dec. 2003.

[22] O. Banos, A. Calatroni, M. Damas, and H. Pomares, "Kinect=IMU? learning MIMO signal mappings to automatically translate activity recognition systems across sensor modalities," in *Proc. IEEE 16th Int. Symp. Wearable Comput.*, Newcastle, U.K., Jun. 2012, pp. 92–99.

[23] M. Bidmeshki and R. Jafari, "Low power programmable architecture for periodic activity monitoring," in *Proc. ACM/IEEE 4th Int. Conf. Cyber-Phys. Syst.*, Philadelphia, PA, USA, Apr. 2013, pp. 81–88.

[24] L. Rabiner, "A tutorial on hidden Markov model and selected application in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[25] J. Yuan, Z. Liu, and Y. Wu, "Discriminative subvolume search for efficient action detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 2442–2449.

Kui Liu received the B.E. degree in electrical engineering from Nanchang University, Nanchang, China, in 2005, and the M.S. degree in electrical engineering from Mississippi State University, Starkville, MS, USA, in 2011. He is currently a Graduate Research Assistant with the Department of Electrical Engineering, University of Texas at Dallas, and a member of the Signal and Image Processing Laboratory. His current research interests include real-time image processing, 3-D computer vision, and machine learning.

Chen Chen received the B.E. degree in automation from Beijing Forestry University, Beijing, China, in 2009, and the M.S. degree in electrical engineering from Mississippi State University, Starkville, MS, USA, in 2012. He is currently pursuing the Ph.D. degree at the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX, USA. His current research interests include compressed sensing, signal and image processing, pattern recognition, and computer vision.

Roozbeh Jafari is an Associate Professor with the University of Texas at Dallas. He received the Ph.D. degree in computer science from UCLA and completed a Post-Doctoral Fellowship with UC-Berkeley. His current research interests include in the areas of wearable computer design and signal processing. He has published over 100 papers in these areas. His research has been funded by the NSF, NIH, DoD (TATRC), AFRL, AFOSR, DARPA, SRC, and industry (Texas Instruments, Tektronix, Samsung, and Telecom Italia). He has served as technical program committee chairs for several flagship conferences in the area of wireless health and wearable computers, including the ACM Wireless Health 2012, the International Conference on Body Sensor Networks 2011, and the International Conference on Body Area Networks 2011. He is an Associate Editor for the IEEE Sensors Journal.

Nasser Kehtarnavaz received the Ph.D. degree in electrical and computer engineering from Rice University in 1987. He is a Professor of Electrical Engineering and Director of the Signal and Image Processing Laboratory at the University of Texas at Dallas. His current research interests include signal and image processing, real-time signal and image processing, biomedical image analysis, and pattern recognition. He has authored or co-authored eight books and more than 290 journal papers, conference papers, patents, industry manuals, and editorials in these areas. He has had industrial experience in various capacities at Texas Instruments, AT&T Bell Laboratories, the U.S. Army TACOM Research Laboratory, and the Houston Health Science Center. He is currently the Editor-in-Chief of the *Journal of Real-Time Image Processing*, the Vice Chair of the IEEE Dallas Section, and the Chair of the SPIE Conference on Real-Time Image and Video Processing. Dr. Kehtarnavaz is a Fellow of SPIE, and a licensed Professional Engineer.