33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Improved Blood Glucose Estimation through Multi-Sensor Fusion

Feiyu Xiong, *Graduate Student Member, IEEE*, Brian R. Hipszer, *Member, IEEE*, Jeffrey Joseph, and Moshe Kam, *Fellow, IEEE*

***Abstract*—Continuous glucose monitoring systems are an integral component of diabetes management. Efforts to improve the accuracy and robustness of these systems are at the forefront of diabetes research. Towards this goal, a multi-sensor approach was evaluated in hospitalized patients. In this paper, we report on a multi-sensor fusion algorithm to combine glucose sensor measurements in a retrospective fashion. The results demonstrate the algorithm's ability to improve the accuracy and robustness of the blood glucose estimation with current glucose sensor technology.**

I. Background

Continuous glucose monitoring (CGM) systems provide an estimate of an individual's blood glucose level on a near-continuous basis (every few minutes) [1]. Current CGM systems are considered adjunctive devices and, as such, a user must obtain a capillary fingerstick glucose measurement prior to instituting any change in therapy. Even with current limitations, CGM use in children and adults with diabetes has been associated with reductions in both HbA1c levels and the frequency of severe hypoglycemia [2-4].

In the future, the hope is that CGM data will be used to make automatic adjustments in therapy achieve better glucose control with the ultimate goal being a fully automated insulin delivery system. CGM performance is a significant factor to achieving this goal [5]. However, the accuracy of CGM systems available today may not be well-suited for this application [6]. Rather than focusing on hardware improvements, we evaluate the use of redundancy with current technology by developing a means to intelligently combine the data from multiple glucose sensors to improve CGM accuracy and robustness.

II. Algorithm Introduction

A. Data type applied

The algorithm described in this paper will apply to any CGM data set with several sensors of the same type. These sensors monitor the same patient at the same time and provide periodic near-continuous measurements (Table I). This algorithm has been applied to CGM data collected in ten patients admitted to Thomas Jefferson University Hospital (Philadelphia, PA). Each patient data set contains, on average, 53 hours of data from six interstitial fluid glucose sensors. The output (in nanoamperes) from each sensor is reported every minute.

B. Step 1 processing

The purpose of first step in CGM data processing is to remove isolated glitches in individual sensor data. A relatively small window of size w_1 is selected.

TABLE I
PATIENT DATA SET

Time	1	...	t	t+1	...	$t+w_1$	...	T
Sensor 1	$s_{1,1}$	...	$s_{1,t}$	$s_{1,t+1}$	…	$s_{1,t+w_1}$	…	$s_{1,T}$
Sensor 2	$s_{2,1}$	…	$s_{2,t}$	$s_{2,t+1}$	…	$s_{2,t+w_1}$	…	$s_{2,T}$
⋮								
Sensor N	$s_{N,1}$	…	$s_{N,t}$	$s_{N,t+1}$	…	$s_{N,t+w_1}$	…	$s_{N,T}$

In a patient data set with N sensors, each individual sensor data point is represented as $s_{n,t}$ where n=1,2,…,N denotes the n^{th} sensor and t=1,2,…,T denotes the chronological order of the point. The window moves point by point, calculating the difference $d_{n,t}$ between $s_{n,t}$ and $s_{n,t+w_1}$ for each sensor, generating an N-element array at time t.

A k-mean algorithm with k=2 is applied to this array to separate the sensors into two groups. The group with the numerical majority is assumed to contain the sensors that are tracking the true glucose concentration. The data points between t and $t+w_1$ associated with the majority are labeled valid. If the two groups have equal number of sensors, all data points between t and $t+w_1$ are labeled valid. Every data point data will be evaluated w_1 times except for the first and last w_1-1 data points. If the majority of the evaluations are valid, the data point will be considered a true data point; otherwise, it will be marked as a false data point.

Sensors with high percentage of false data points are identified for removal during final processing. The frequency of false data points is compared to a predefined threshold. The i^{th} element of an N-element array p_1 will be set to 1 if the i^{th} sensor is to be removed and 0 otherwise.

False data points from the remaining sensors are replaced with interpolated values. For example, if data points between s_{n,t_1} and s_{n,t_2} are false, a line is constructed

between these two points and sampled at times $t_1 < t < t_2$. The false data points are replaced by the corresponding sampled values.

In our application, the number of sensors is 6 (N=6), the number of data points per sensor varies depending on the patient data set ($1666 \leq T \leq 3325$), the window size is 10 minutes (w_1=10), and the threshold frequency for false data points is 0.3. w_1 is chosen to be a relatively small compared to T and the threshold frequency is empirically determined to provide the best separation between good and poor performance in several simulation tests.

C. Step 2 processing

The purpose of the second step is to identify and remove sensors that have a weak correlation with the group. A window with a size larger than w_1 is used here. The window of size w_2 minutes moves hour by hour until it reaches the end of the patient data set. At time t, the window contains the data from t to $t+w_2$ for all sensors. Then an N×N correlation matrix is generated that contains the linear correlation coefficients between each pair of sensors (Table II).

TABLE II
CORRELATION MATRIX

Sensor	1	2	3	4	5	6
1	1	0.3851	0.3461	0.8614	0.7823	0.6721
2	0.3851	1	0.7538	0.6784	0.6351	0.6968
3	0.3461	0.7538	1	0.6487	0.5078	0.7693
4	0.8614	0.6784	0.6487	1	0.8933	0.8342
5	0.7823	0.6351	0.5078	0.8933	1	0.7698
6	0.6721	0.6968	0.7693	0.8342	0.7698	1

The sum of each row of the correlation matrix is calculated and the k-mean algorithm with k=2 is applied to this N-element array to separate it to two groups. The group with larger centroid contains the best-correlated sensors at time t.

In the similar fashion as in step 1, the data points between t and $t+w_2$ associated with the best-correlated sensors are labeled valid. Every data point data will be evaluated $w_2/60$ times except for the data points in the first and last hour. If over half of the evaluations indicate this data point is valid, it will be considered as true data point; otherwise, it will be marked as a false data point.

After window has moved over the entire data set, the number of hours containing false data points is tabulated for each sensor and stored in an N-element array p_2 for use final processing.

In our application, the window size for step 2 is 360 minutes (w_2=360).

D. Final processing

The final step of the algorithm integrates the results from steps 1 and 2 to identify the best-performing sensors in the patient data set. Excluding the sensors identified for removal in step 1, the sensors with the least false hours in step 2 are selected. These sensors are averaged to produce a final composite output.

The selection logic of this final step is illustrated in the following example. The patient data set has 6 sensors. The results from steps 1 and 2 are given in Table III. Sensors 1 and 2 are excluded based on Step 1 processing. Out of the remaining sensors, 3 and 5 have the lowest values in p_2. These sensors are selected and averaged.

TABLE III
AN EXAMPLE FOR FINAL PROCESSING

sensor	1	2	3	4	5	6
p_1	1	1	0	0	0	0
p_2	5	29	6	10	6	23

III. DATA AND EVALUATION CRITERION

The data used to evaluate algorithm performance is from the project entitled "Artificial Pancreas for Control of BG and Insulin Levels in Hospitalized Patients with Diabetes and Stress Hyperglycemia" sponsored by the Department of the Army under the Technologies in Metabolic Monitoring Initiative. The project's purpose was to evaluate two glucose sensing technologies in the preoperative setting in diabetic patients and patients with stress hyperglycemia. One of the technologies was the Telemetered Glucose Monitoring System (TGMS). The TGMS is a research-only device and predecessor of the Guardian REAL-Time System (Medtronic Diabetes, Northridge, CA).

Ten surgical patients were studied (TABLE IV). Immediately prior to surgery, six TGMS sensors were placed on each patient. The sensors were placed into the fatty tissue of the upper arm, upper thigh, and/or abdomen (Fig. 1) and connected to transmitters that wirelessly transferred the sensor data to a bedside computer.

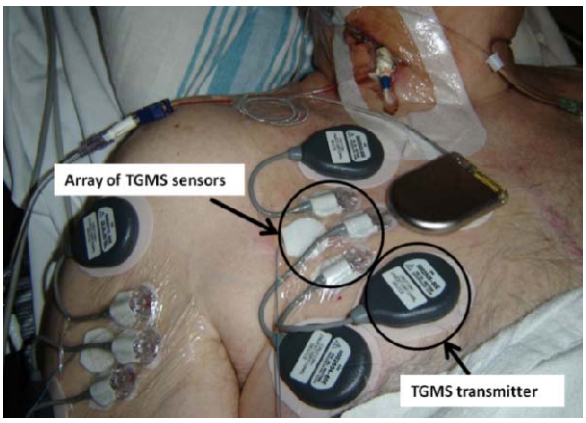

Fig. 1. Telemetered Glucose Monitoring System (TGMS)

Patients were studied for up to 60 hours in the

perioperative period. Each sensor provided measurement of the interstitial fluid glucose concentration every minute. In addition, reference blood samples were collected to calibrate and evaluate sensor data. Reference blood samples were obtained from an artery catheter every 20 minutes and central venous line every 60 minutes. All samples are assayed for the concentration of glucose using an OMNI 9 Blood Gas Analyzer (Roche Diagnostics). The reference blood source used for each patient differed depending on the ability to consistently obtain blood from either the artery or central vein (TABLE IV). Reference data may not be available for the entire sensor wear period if the sampling catheter failed prematurely. On average, venous reference data covered a longer period than arterial reference data although arterial blood was sampled more frequently.

TABLE IV
REFERENCE BLOOD SOURCE AND PATIENT DEMOGRAPHICS

Patient ID	Reference Blood Source	Sex	Age (yr)	BMI (kg/m2)	Sensor Wear (hrs)
A2	Arterial	F	53	46.1	58.5
B2	Arterial	M	73	31.6	43.6
A3	Arterial	F	47	19.1	59.9
B3	Arterial	M	55	22.7	35.7
C2	Arterial	M	68	22.3	32.6
C3	Venous	M	58	22.9	59.4
D3	Venous	M	59	29.5	59.7
E3	Venous	F	51	20.7	60.9
F3	Venous	M	56	27.8	59.8
D2	Venous	M	63	36.4	60

The algorithm was implemented on a continuous block of data that began 6 hours after the last sensor was inserted, starting the postoperative period. In order to quantify the performance of the algorithm, the Mean Absolute Relative Deviation (MARD) was calculated. The formulation for MARD is

$$MARD = \frac{1}{n}\sum \frac{|x-y|}{x}$$

where x and y are paired values of calibrated sensor data and the reference data, respectively, and n is the number of paired values. Sensor data are recalibrated every 6 hours using reference glucose data. Reference data used for calibration were excluded from MARD calculations.

IV. RESULTS AND DISCUSSION

A. Results

Algorithm results for all patients except A2, A3 and E3 are shown in the Table V where the best-performing sensor(s), as chosen by the algorithm, are highlighted. For example, Sensor 6 is selected as the best-performing sensor in D2. If all the raw sensor data and reference data are plotted together (Fig. 2), it is apparent that trajectory of Sensor 6 matches reference data best, demonstrating the algorithm selected the correct sensor as best-performing sensor without having any knowledge of the actual blood glucose levels.

Alternatively, to quantitatively assess the algorithm performance, MARD from the algorithm's composite of the "best-performing" sensor(s) is compared to the MARD from (1) the average of all sensors and (2) the best individual sensor. In all cases, the composite MARD was lower than the all-sensors average (Table VI). Averaged over all patients, the MARD for the composite is 0.0917 versus 0.1175 and 0.0990 for the all-sensors average and the best sensor. The algorithm provides an average MARD decrease of 2.6% compared to the straight-forward averaging of all sensors. Moreover, the composite MARD is lower than the best individual sensor in all cases except C3.

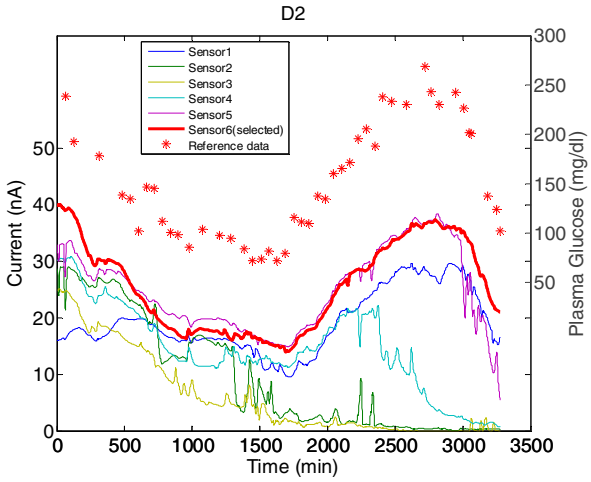

Fig. 2. Algorithm results and comparison for raw sensor data and reference glucose data for Patient D2

B. Exceptions

The data for patients A2, A3, and E3 are considered exceptions because unique circumstances impact the algorithm's performance. For patient A2, the row-sum of the correlation matrix is [5.79 5.33 5.66 5.79 5.80 5.80]. These numbers indicate that all the 6 sensors are highly correlated. Indeed, p_2 is [0 4 1 1 7 1]; its mean is the lowest among all the patients. Every sensor placed on patient A2 is performing well (Fig. 3(a)) but the current algorithm cannot make that distinction. For patient E3, all sensors exhibit a high level of noise. Every sensor placed on patient E3 is performing poorly (Fig. 3(b)). The row-sum of the correlation matrix is [3.28 3.08 2.08 3.52 3.03 2.86], indicating poor correlation among all the sensors. The current algorithm can neither identify nor improve the situation with patient E3. For patient A3 (Fig. 3(c)), algorithm fails. In step 2 processing, the algorithm

determines Sensors 2 and 3 should be excluded. The vast majority of the step 2 false points for these sensors occur in the latter half of the data set. However, these sensors initially outperform all other sensors and their exclusion greatly impacts the MARD calculation.

C. *Discussion*

The algorithm appears to be robust when handling different situations. It is able to identify a single optimally performing sensor as is the case with patients D2 and D3. Whereas, with patients B2 and B3, it identifies a group of four sensors whose composite outperforms both the best single sensor and the average of all six sensors. In all cases, the algorithm did not have any knowledge of the reference glucose concentrations during the selection process. However, this analysis is retrospective and only included a small number of patients. Moreover, because of the limited sample size, it is impossible to determine whether sensor performance is related to subject weight, gender or sensor location.

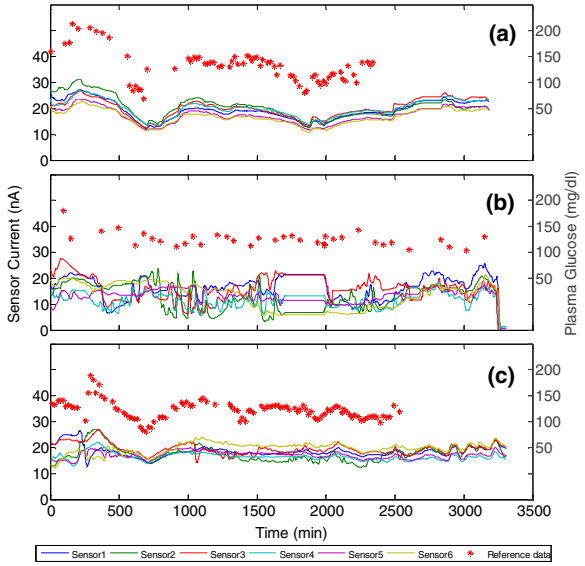

Fig. 3. Raw sensor data and reference glucose data for Patients A2 (panel a), A3 (panel b) and E3 (panel c).

V. Conclusion and Future Work

Multiple sensors can improve the accuracy and robustness of blood glucose estimation. The algorithm presented here retrospectively processes the data. In addition to moving toward an online implementation, the current algorithm needs to be optimized to (1) identify situations when all sensors are performing well, or poorly, and (2) remove only portions of sensor data, Future work will focus on real-time processing of multiple sensors and the collection of more patient data sets to validate the algorithm performance.

TABLE V
PROCESSED RESULTS

ID	Sensor	1	2	3	4	5	6
B2	p_1	1	0	0	0	0	0
	p_2	33	5	10	5	5	5
B3	p_1	0	0	1	0	1	0
	p_2	0	0	17	0	10	0
C2	p_1	1	0	0	0	0	0
	p_2	0	8	11	0	0	10
C3	p_1	0	1	0	0	1	0
	p_2	0	15	13	0	21	0
D2	p_1	0	1	0	0	0	0
	p_2	14	27	35	19	10	5
D3	p_1	0	1	1	1	0	0
	p_2	38	15	19	5	7	18
F3	p_1	1	1	0	0	0	0
	p_2	5	29	6	10	6	23

TABLE VI
MARD FOR PATIENT

ID	All-Sensors Average	Best Sensor	Algorithm Composite
B2	0.1024	0.1010	0.0884
B3	0.0711	0.0680	0.0646
C2	0.1312	0.0970	0.0847
C3	0.1080	0.0947	0.0960
D2	0.1449	0.1276	0.1274
D3	0.1371	0.1369	0.1165
F3	0.1278	0.0680	0.0646

References

[1] Burge MR, Mitchell S, Sawyer A, Schade DS. Continuous glucose monitoring: The future of diabetes management. Diabetes Spectrum. 2008 April 01;21(2):112-9.

[2] Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group, Tamborlane WV, Beck RW, Bode BW, Buckingham B, Chase HP, et al. Continuous glucose monitoring and intensive treatment of type 1 diabetes. N Engl J Med. 2008 Oct 2;359(14):1464-76.

[3] Battelino T, Phillip M, Bratina N, Nimri R, Oskarsson P, Bolinder J. Effect of continuous glucose monitoring on hypoglycemia in type 1 diabetes. Diabetes Care. 2011 Apr;34(4):795-800.

[4] Bailey TS, Zisser HC, Garg SK. Reduction in hemoglobin A1C with real-time continuous glucose monitoring: Results from a 12-week observational study. Diabetes Technol Ther. 2007 Jun;9(3):203-10.

[5] Hovorka R. Continuous glucose monitoring and closed-loop systems. Diabet Med. 2006 Jan;23(1):1-12.

[6] Castle JR, Ward WK. Amperometric Glucose Sensors: Sources of Error and Potential Benefit of Redundancy. J Diabetes Sci Technol. 2010 Jan 1;4(1):221-5.