Surgical Phase Recognition using Movement Data from Video Imagery and Location Sensor Data

A. Nara[1], C. Allen[1], K. Izumi[2]

[1]Department of Geography, San Diego State University, 5500 Campanile Dr., San Diego, CA 92182-4493
+1-619-594-1305
Email: anara@mail.sdsu.edu; allen14@rohan.sdsu.edu

[2]Graduate School of Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656
Email: izumi@sys.t.u-tokyo.ac.jp

Abstract

In this paper we investigated the performance of context recognition from moving object data collected by imaging and tracking. We employed an optical flow algorithm and trajectory clustering techniques to extract movement characteristics of surgical staff from video imagery and time-stamped location data collected by an ultrasonic location aware system respectively. Then we applied a Support Vector Machine to time-stamped location data, optical flow estimates, trajectory clusters, and combinations of these three data to examine the intraoperative context recognition rate. Our results showed that the integration of both video imagery and location sensor data improves context awareness of neurosurgical operations.

Keywords: Context Awareness; Moving Object; Optical flow; Trajectory Clustering.

1. Introduction

Advances in location sensing and computing technologies enable automatic tracking of moving objects at a high level of detail in space and time. Context awareness from such moving object data is one of the key research challenges in data mining and ubiquitous computing. Activity recognition and situation awareness associated with locations, time, and moving objects facilitate the interaction between users and computing system, which ultimately supports decision making in applications such as transportation (Andrienko et al. 2011), video surveillance (Rougier et al. 2011), and offender monitoring systems (Yuan and Nara 2015). In the Operating Room (OR) environment, context awareness promotes better patient treatment and higher hospital efficiency; for instance, automatic surgical phase recognition supports dynamic scheduling and resource allocations (Sutherland et al. 2006), and workflow analysis (Padoy et al. 2012) aids workflow optimization and standardization.

To achieve context recognition of intraoperative activities, various monitoring approaches have been proposed: patient's vital signs (Xiao et al. 2005), instrument signals (Padoy et al. 2007), surgeon's elbow and wrist movements using two video cameras (Ohnuma et al. 2006), eye-gaze tracking data (James et al. 2007), and standardized free-hand movement by a Kinect sensor (Yoshimitsu et al. 2014). While previous research has demonstrated that various sensors can recognize activities, workflows, and phases during an operation, most of them ignore the comparative study of sensor technologies and their context recognition performances.

This paper investigated the performance of context recognition from moving object data collected by imaging and tracking. We employed an optical flow algorithm and trajectory clustering techniques to extract movement characteristics of surgical staff from video imagery and time-stamped location data collected by an ultrasonic location aware system respectively. Then we apply a Support Vector Machine to time-stamped location data, optical flow estimates, trajectory clusters, and combinations of these three data to examine the intraoperative context recognition rate.

2. Data Collection

2.1 Video Imagery

In this study, single channel intraoperative video imagery was recorded. The camera is mounted on the wall near the entrance of the operating room to shoot the surgical field and staff (Fig. 1).

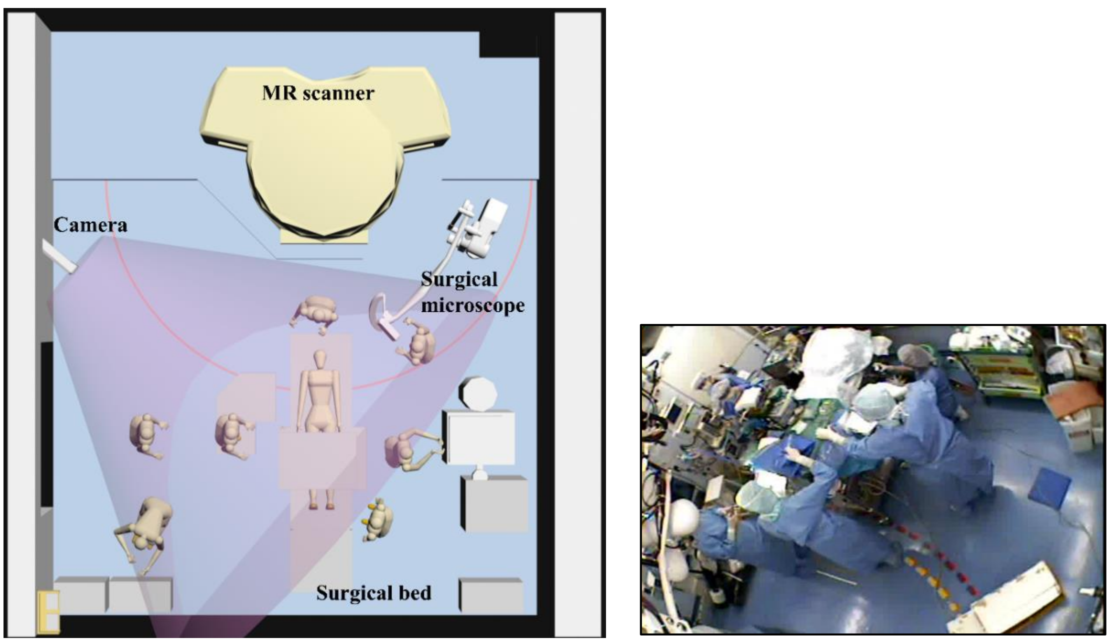

Figure 1. Operating room layout (left) and camera view (right).

2.2 Ultrasonic Location Aware System

The ultrasonic location aware system (Fig. 2) consists of ultrasonic tags (transmitters), receivers, and four control units. The receivers receive ultrasonic pulses emitted from multiple tags. Four control units identify each tag's identification and detect associated 3D positions in the OR. To estimate a location, the system records the time-of-flight, which is the travel time of the signal from transmission to reception. Based on more than three time-of-flight results, the system computes 3D position using the trilateration method based on robust estimation algorithm known as random sample consensus (RANSAC) (Fischler et al. 1981). Table 1 represents the system specification in a typical environment.

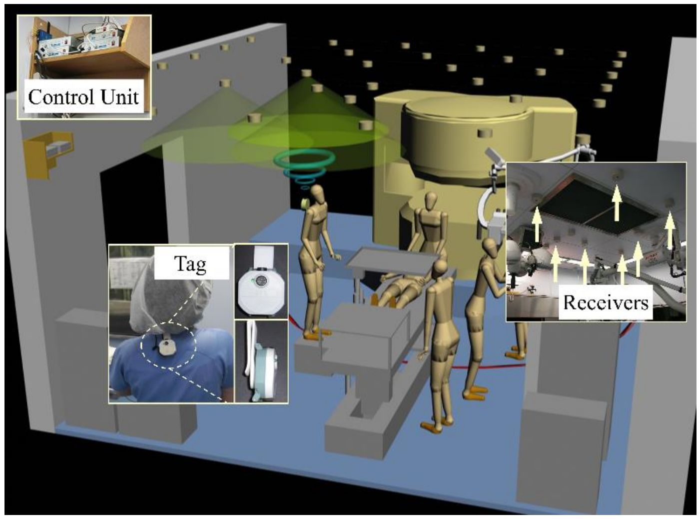

Figure 2. Ultrasonic Location Aware System.

Frequency of ultrasound	40KHz
Position estimation error	less than 80mm
Sampling frequency	up to 50 Hz
Measurement range (Distance)	Vertical distance from a reader ≈7m
Measurement range (Angle)	Vertical angle from a reader ≈100°
Max number of tags	2048

Table 1. Specification of the ultrasonic location aware system

3. Methods

3.1 Tag Movements

We installed the ultrasonic location aware system in an OR at the Tokyo Women's Medical University (TWMU), Tokyo, Japan. The room is 5.8m (Width) x 4.8m (Depth) x 2.9m (Height) in size (Fig. 2). We deployed 33 ultrasonic receivers on the ceiling and set four control units on the wall nearby the room entrance. The wearable ultrasonic tag is 44mm (Width) x 75mm (Height) x 24mm (Depth) in size, and 40g (Tag: 30g, Battery: 10g) in weight. For the purpose of minimum disturbance during a surgical operation, a single tag was hooked on surgical clothes around the nape of surgical staff's neck. We have also verified that the system does not conflict with other surgical devices including MR (Magnetic Resonance) scanners.

At TWMU, a typical neurosurgical operating team includes surgeons, anesthetists, engineers, scrub nurses, and assistant nurses. We have collected tag movement data from surgeons, anesthetists, scrub nurses, and assistant nurses and excluded engineers who are not continuously present in the OR during neurosurgical operations. In this study, we used movement data from 10 neurosurgery cases.

3.2 Optical Flow

Optical flow measures attempt to track the movement of individual features from one frame to the next and produce a set of motion vectors that describe the direction and magnitude of these movements. The most straight-forward of these algorithms use the concept of block matching, which divides a frame into sub-blocks and searches for corresponding sub-blocks in the second frame. To search for candidate sub-blocks, block matching techniques often rely upon metrics such as the root-mean-square error (RMSE) or the sum of absolute differences (SAD). Block matching has been effectively applied to many computer vision problems, such as segmenting moving objects (Bradsky and Davis 2002) and measuring cyclical motion in artery walls (Golemati et al. 2003).

In this study, we generated frame images using a one-second sampling, which is consistent with the sampling frequency of the tag sensor data. Subsequently, we calculated optical flow vectors between two consecutive frames. Due to runtime performance considerations, we used the Lucas-Kanade method (Lucas and Kanade 1981), which is optimized for real-time analysis. The magnitudes for each of these vectors was summed in order to derive a global measure of activity at a given time. Because our camera is stationary, we did not need to account for any movement of the camera when calculating optical flow. Note that this global approach diverges from most studies that use optical flow to track individual objects; here we are merely concerned with aggregate motion within a scene.

3.3 Trajectory Clustering

The ultrasonic location aware system collects a set of trajectories from multiple surgical staff {Trajectory Set: $T = T_1, T_2, T_3, \ldots, T_i$, where i denotes the number of surgical staff} during a surgical operation. Each trajectory is composed of a sequence of 4-dimensional points {{$T_i = p_1, p_2, p_3, \ldots, p_j$, where j denotes the number of points in the trajectory i}, {$p_j = x, y, z, t$}}. To extract intraoperative movement characteristics of surgical staff, we employed trajectory data mining techniques, which include two procedures, trajectory partitioning and trajectory clustering (Nara et al. 2011). The trajectory partitioning process partitioned an entire trajectory of each surgical staff into trajectory partitions (sub-trajectories). By grouping trajectory partitions for each surgical role, the unsupervised clustering process describes surgical events and procedures that have similar trajectory patterns.

For each trajectory partition, we obtain multi-dimensional vectors to characterize the partition trajectory. The vector values include total distance (x-y axes), distance between start and end nodes (x-y axes), total distance (z axis), and time duration. Then, k-means cluster analysis is run based on standardized values of these vectors. To estimate the quality of clusters for determining the number of cluster k in k-means automatically, we applied the IGRC (Information Gain Ratio for Cluster) index (Yoshida et al. 2006).

4. Results and Discussion

Figure 3 shows the relationship between Optical Flow and tag moving distances during one neurosurgical operation. The thick red line represents the amount of Optical Flow, while the thick black line represents the total tag moving distance. It visually shows a strong correlation between two values suggesting that both movements obtained from video imagery and the ultrasonic location aware system describe similar movement behaviors. The blue vertical bars explains the amount of moving distance by each role (i.e., surgeons, anesthetists, scrub nurses, assistant nurses).

Figure 4 draws the result of trajectory clustering, which makes groups of similar moving and stopping behaviors. For each staff's role, we quantified moving behaviors and created time-sequence vectors by counting the total amount of duration represented for each trajectory cluster for a specified time window. In this study, we selected 1 minute, 5 minutes, and 10 minutes as time windows. In order to compare the context recognition rate with the trajectory cluster duration, we used a time-window averaging method with the size of 1 minute, 5 minutes, and 10 minutes to create time-sequence vectors for Optical Flow and tag moving distances.

Finally, we applied a Support Vector Machine to tag moving distances, Optical Flow, and trajectory clustering, and combinations of these three data to evaluate the performance of the intraoperative context recognition using 10-fold cross-validation. Our results show that the integration of both video imagery and location sensor data improves context awareness of neurosurgical operations particularly when utilizing trajectory clustering outcomes (numbers in red). As shown on Figure 3, tag moving distances and Optical Flow describe similar moving behaviors; therefore, combining these two data does not improve the recognition rate. On the other hand, trajectory clustering is able to extract unique movement characteristics from the same data and thus provides a more comprehensive description of surgical phases.

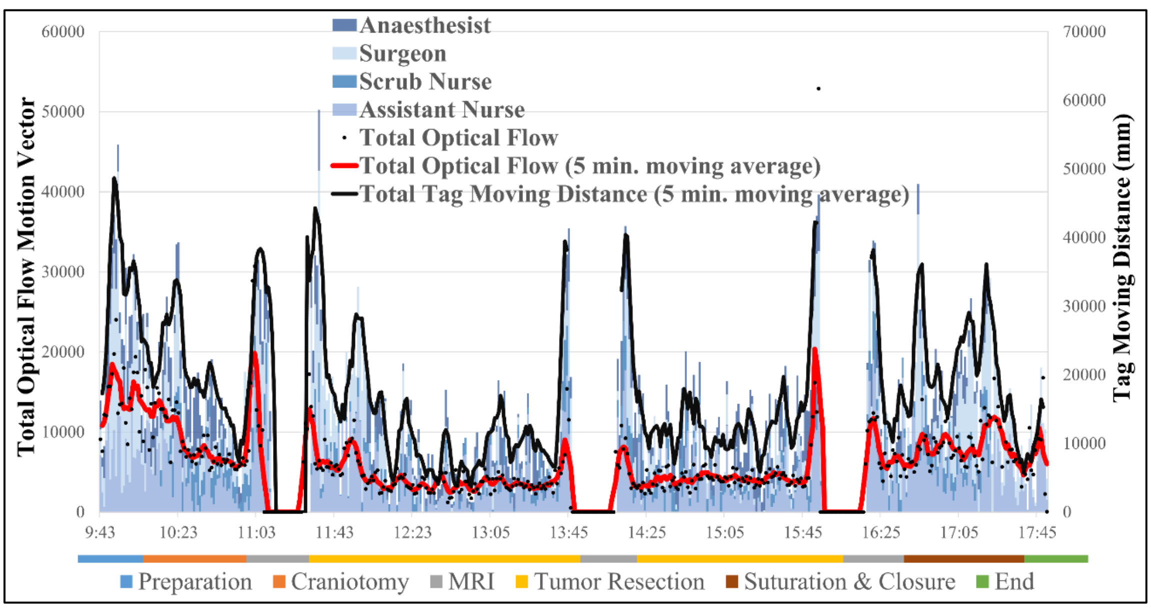

Figure 3. The relationship between Optical Flow and tag movements

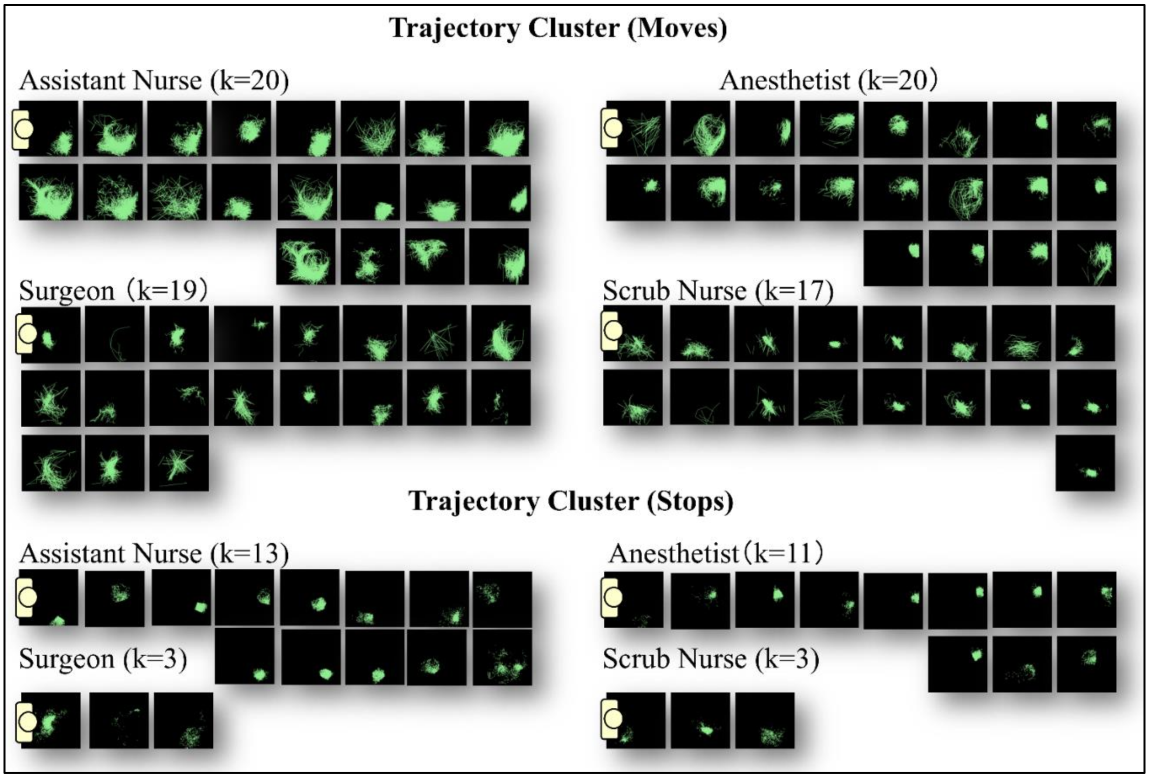

Figure 4. Trajectory clustering results

	without time-elapse			with time-elapse		
Moving Window Average	1min	5min	10min	1min	5min	10min
Optical Flow	40.08	48.37	49.09	67.14	68.57	70.82
Tag Movement	34.78	36.94	36.42	61.83	60.82	63.38
Trajectory Cluster	45.71	57.86	58.15	69.46	72.45	74.25
Optical Flow + Tag Movement	49.56	50.41	50.50	72.44	73.57	72.44
Optical Flow + Trajectory Cluster	50.30	62.55	65.19	71.24	76.43	77.46
Tag Movement + Trajectory Cluster	49.02	60.41	60.97	71.28	74.59	73.84
All	55.43	65.92	67.20	75.05	78.27	77.67

Table 2. Comparison of surgical phase recognition rate (%)

5. Acknowledgements

This study was partially funded by NEDO (New Energy and Industrial Technology Development Organization) through the Intelligent Surgical Instruments Project. We thank for Drs. Hiroshi Iseki, Takashi Suzuki, Takashi Maruyama, and Masahiko Tanaka, and other staffs of the Tokyo Women's Medical University for participating in data collection.

6. References

Andrienko G, Andrienko N, Heurich M, 2011, An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, 25(9):1347-1370.

Bradski GR and James WD, 2002, Motion segmentation and pose recognition with motion history gradients. *Machine Vision and Applications*,13(3): 174-84.

Fischler MA and Bolles RC, 1981, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6): 381-395.

Golemati S, Sassano A, Lever MJ, Bharath AA, Dhanjil S and Nicolaides AN, 2003, Carotid artery wall motion estimated from B-mode ultrasound using region tracking and block matching. *Ultrasound in Medicine & Biology*, 29: 387-399.

James A, Vieira D, Lo B, Darzi A and Yang, GZ, 2007, Eye-gaze driven surgical workflow segmentation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2007*, 110-117.

Lucas BD and Kanade T, 1981, An iterative image registration technique with an application to stereo vision. In *IJCAI*, 81: 674-679.

Nara A, Izumi K, Iseki H, Suzuki T, Nambu K and Sakurai Y, 2011, Surgical workflow monitoring based on trajectory data mining. In *New Frontiers in Artificial Intelligence*, 283-291.

Ohnuma K, Masamune K, Yoshimitsu K, Sadahiro T, Vain J, Fukui Y and Miyawaki F, 2006, Timed-automata-based model for laparoscopic surgery and intraoperative motion recognition of a surgeon as the interface connecting the surgical and the real operating room. *International journal for computer assisted radiology and surgery*, 1: 442-445.

Padoy N, Blum T, Ahmadi SA, Feussner H, Berger MO and Navab N, 2012, Statistical modeling and recognition of surgical workflow. *Medical Image Analysis*, 16(3): 632-641.

Padoy N, Blum T, Essa I, Feussner H, Berger MO, Navab N, 2007, A boosted segmentation method for surgical workflow analysis. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2007*, 102-109.

Rougier C, Meunier J, St-Arnaud A, Rousseau J, 2011, Robust video surveillance for fall detection based on human shape deformation. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(5): 611-622.

Sutherland J, and van den Heuvel WJ, 2006, Towards an intelligent hospital environment: adaptive workflow in the OR of the future. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences 2006*, 100b.

Xiao Y, Hu P, Hu H, Ho D, Dexter F, Mackenzie CF, Seagull FJ, and Dutton RP, 2005, An algorithm for processing vital sign monitoring data to remotely identify operating room occupancy in real-time. *Anesthesia and Analgesia*, 101(3): 823-829.

Yoshida T, Shoda R and Motoda H, 2006, Graph clustering based on structural similarity of fragments. *Lecture Notes in Computer Science*, 3847: 97-114.

Yoshimitsu K, Muragaki Y, Maruyama T, Yamato M and Iseki H, 2014, Development and initial clinical testing of “OPECT”: An innovative device for fully intangible control of the intraoperative image-displaying monitor by the surgeon. *Neurosurgery*, 10: 46-50.

Yuan M and Nara A, 2015, Space-time analytics of tracks for the understanding of patterns of life. In Kwan MP, Richardson D, Wang D and Zhou C (eds), *Space-Time Integration in Geography and GIScience*. Springer Netherlands, 373-398.