Proceedings of the 2025 SICE Festival with Annual Conference
September 9-12, 2025, Chiang Mai, Thailand

Lung Volume Estimation using Upper Chest Motion Tracking with a Intel RealSense Depth Camera

P. Vejjanugraga[1†], P. Siritanawan[2], J. Izawa[3], and K. Kotani[4]

[1]College of Arts, Media, and Technology, Chiang Mai University, Chiang Mai, Thailand
[2]Shinshu University, Nagano, Japan
[3]Komatsu University, Ishikawa, Japan
[4]Kanazawa University, Ishikawa, Japan
(E-mail: pikul.v@cmu.ac.th)

Abstract: Contactless respiratory monitoring has emerged as a promising alternative to traditional spirometry, offering advantages in comfort, hygiene, and continuous assessment in both clinical and homecare settings. This study proposes a method for estimating lung volume by tracking chest motion using a 3D Intel RealSense depth camera. By capturing real-time depth data from the thoracic region, the system identifies and tracks key surface landmarks throughout the respiratory cycle. The 3D velocity vector of these landmarks reflects the dynamic expansion and contraction of the chest during breathing. Statistical features such as maximum, minimum, average, and variability of motion are extracted and used to train regression models. Linear regression model based on a three-point chest tracking (F_3-LR-20) demonstrated the highest predictive performance. The model's outputs were compared to established reference values from the Baldwin and Japanese Respiratory Society (JSR) equations, with results showing a closer alignment to the JSR standard.
Keywords: Lung volume estimation, depth camera, respiratory monitoring, chest motion tracking, Baldwin, JSR

1. INTRODUCTION

Pulmonary function tests (PFTs) have been the standard measurements for assessing respiratory health and conditions. The most traditional measurement is the spirometry, which measures lung volumes and airflow by requiring patients to perform forced breathing maneuvers into spirometers [1]. The traditional spirometry, while accurate, requires direct contact with the patient and active cooperation, which may be unsuitable in certain clinical conditions or prolonged monitoring scenarios, especially under the concerns of contagious respiratory diseases during the pandemics [2].

With the advancement of computer vision and perceptual sensing technologies, the ability to estimate respiratory parameters through visual data has become feasible. Generally, the previous works tend to focus on analyzing temporal changes in small regions of interest (ROI) located on the chest or abdomen in 2D image sequences captured from the frontal view. Motion estimation techniques are one of the most widely adopted techniques in the field, such as dense optical flow [3], or feature tracking [4]. Similarly, the pixel intensity variation was adopted under similar image capturing conditions [5]. In addition, respiratory rate can be measured through remote plethysmography (rPPG) calculated from the changes over facial regions [6].

However, these image-based methods face several limitations. They are sensitive to variations in lighting conditions and the reflective properties of clothing. Moreover, the observable chest motion in the frontal image plane is often minimal compared to unrelated body movements, making it difficult to accurately extract breathing signals. Due to these limitations, such methods are insufficient for estimating key pulmonary function parameters needed for respiratory disease diagnosis.

To address these challenges, depth cameras such as the Intel RealSense offer the capability to monitor chest surface motion in three dimensions, enabling indirect lung volume estimation without physical contact. This research investigates the feasibility of using such technology to monitor respiratory activity by first extracting 3D keypoints from the chest surface. The motion of each keypoint is then tracked throughout the breathing cycle. Finally, the dynamic changes in these 3D points are analyzed and converted into an estimate of lung volume (VC).

2. METHODOLOGY

This research was conducted by following these 4 main steps: (1) Data Acquisition, (2) Region of Interest (ROI), (3) Chest Motion Tracking, and (4) 3D Lung Volume (VC) Analysis as shown on Fig.1.

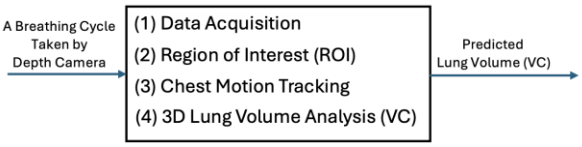

Fig. 1 Overview of the system

2.1 Data Acquisition

An Intel RealSense D435 camera was used to record depth and RGB video of subjects in a seated position. The camera was positioned 1.2 meters in front of the subject to focus on the upper thoracic region. Depth data was recorded in `.bag` format for offline processing using the RealSense SDK or Viewer, with a resolution of 640×480 pixels at 30 FPS.

† P. Vejjanugraha is the presenter of this paper.

2.1.1 The instruction for recording the data set is as follows:

- Clothing: Subjects are required to wear non-reflective, tight-fitting clothing (e.g., a T-shirt) to minimize artifacts and ensure accurate depth measurements. The green point markers are put over the Posture. At the beginning of the experiment, subjects are instructed to sit comfortably with their back straight and feet flat on the floor, maintaining a stable posture throughout the recording.
- Breathing Protocol: To measure Vital Capacity (VC), subjects are asked to perform a maximum inhalation followed immediately by a full exhalation. They are instructed to breathe in as deeply as possible, and then exhale completely until no air remains in their lungs.
- Instructions During Recording: Subjects are advised to avoid speaking or moving their upper body during the recording session to prevent noise in the motion tracking data.
- Sitting Position: the distance between the chair and camera are fixed at 120 cm as shown in Fig.2.
- PDPA: The consent form with an explanation about file storage and private information were informed to the subjects.

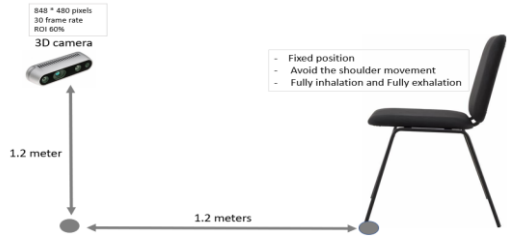

Fig. 2 Sitting position

2.1.2. 2D-3D Transformation

To make use of image features such as keypoints or ROIs, some processing steps were performed on the 2D image plane and then mapped to 3D coordinates using the aligned depth data. For this transformation, each 2D image coordinate (x, y) with its corresponding depth value Z can be converted into a 3D point (X, Y, Z) using the following equation:

$$X = \frac{(x - c_x)Z}{f_x}, Y = \frac{(y - c_y)Z}{f_y} \quad (1)$$

where f_x and f_y are the focal lengths of the camera in pixels along the x and y axes, and (c_x, c_y) is the principal point (optical center) of the image. This projection allows for accurate spatial interpretation of chest surface features detected in 2D, enabling reliable 3D tracking and volume analysis

2.2 Region of Interest (ROI)

To capture the area most indicative of respiratory motion, a Region of Interest (ROI) was defined over the central thoracic region of the subject's chest, typically centered around the sternum, where chest expansion and contraction are most prominent during breathing. Depth values within this ROI were extracted from each frame to measure displacement corresponding to the respiratory cycle.

In this study, a fixed rectangular ROI was applied consistently across frames. This approach was feasible because subject movement was minimized under controlled environmental conditions during data acquisition. While manual ROI selection was sufficient in this setup, future implementations may benefit from automated landmark detection to adaptively position the ROI based on anatomical features.

2.3. Chest Motion Tracking

To measure chest motion during respiration, point markers were physically placed on the subject's shirt over the thoracic region. These markers served as chest landmarks that could be detected on RGB images taken from the Intel Realsense sensor. The 2D positions of the markers were extracted from each frame, then projected into 3D coordinates using the aligned depth map and camera intrinsics (as described in Section 2.1.2). Each landmark's 3D position was tracked over time to capture its motion throughout the respiratory cycle. The average displacement was calculated for each frame to represent chest displacement. This formed a time-series signal that reflects local movement of chest surface during breathing.

To measure chest motion during respiration, green point markers were physically placed on the subject's shirt over the thoracic region. These markers served as chest landmarks that could be detected on RGB images taken from the Intel Realsense sensor. The 2D positions of the markers were extracted from each frame, then projected into 3D coordinates using the aligned depth map and camera intrinsics (as described in Section 2.1.2). In this work, the color of landmarks is green. The green point is detected by converting RGB frames to BGR images and convert to HSV for green color detection and then the upper and lower range of green color was set to match with the landmark's intensity. The green landmark is detected and return its coordinates (P_{xi}, P_{yi}) in order to track the chest movement overtime where i represents the index of landmarks. Then, each landmark's 3D position was tracked over time to capture its motion throughout the respiratory cycle. The average displacement was calculated for each frame to represent chest displacement. This formed a time-series signal that reflects local movement of chest surface during breathing.

The chest landmark in Fig.3 (left) shows the location of the reference points. The landmark was aligned based on the anatomy, the upper landmarks both left and right side are set five cm. below the middle of the collarbone, the lower landmarks are set at the lower 10th rib. The

middle landmarks are the midway between the upper and the lower landmarks. In Fig.3 (right), the midway between the upper and middle landmarks and between the middle and lower landmarks are added in order to measure more movement on the chest while breathing.

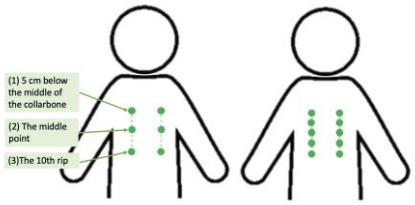

Fig. 3 Chest Landmark: (left) six landmark points, (right) ten landmark points.

Fig.4 shows the example of the subjects and the green alignment points on the body for both front and side. The color heatmap image shows the depth on the body surface. After the adjustment of body alignment, the recording process is started following the instruction from end-inspirstory to end-expiratory. It was found that the direction of the motion on each landmark point significant some pattern.

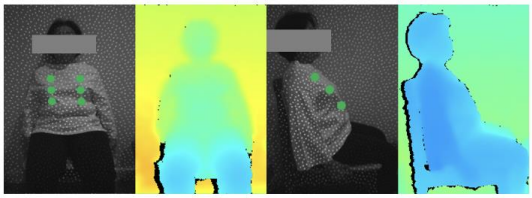

Fig. 4 Infrared and depth images with landmarks for both front and side.

2.4 Volume Estimation

Lung volume estimation is a critical component in respiratory health assessment and pulmonary function analysis. Various empirical equations have been proposed to estimate lung volumes based on anthropometric measurements such as height, age, and gender. One widely used method is the Baldwin equation [7], which provides a regression-based approach to estimate total lung capacity (TLC) and other pulmonary volumes. These equations as shown in Eq. 2, derived from large-scale population data, have been validated across multiple demographics and serve as reliable references in clinical practice. The Baldwin equations are typically expressed as:

Baldwin:

$$VC_m^{bw}(ml) = (27.63 - 0.112 * age) * HT(cm),$$
$$VC_f^{bw}\ (ml) = (21.78 - 0.101 * age) * HT(cm) \quad (2)$$

where VC_m^{bw} is the estimated lung volume of male, $VC_f^{bw}\ (ml)$ is the estimated lung volume of female, age is the age of the subject, HT is the height of the subject in cm.

It was found that Baldwin approach is widely used to estimate static lung volumes from spirometry data. The research builds upon traditional regression-based methods like Baldwin's to enhance lung volume estimation accuracy [8]. Anothr research developed novel formulas based on FLV measurements, which align with the principles of the Baldwin equations by considering individual anatomical differences [9]. They compared functional lung volume (FLV) measurement with the segment-counting method for predicting postoperative pulmonary function.

In addition, the Japanese Respiratory Society (JRS) has proposed a set of equations tailored specifically for the Japanese population as shown in Eq.3, accounting for ethnic and physiological differences [10]. The JRS formulas similarly relate lung volume to body size and age, and are given by:

JSR:

$$VC_m^{JSR} = 0.045 \times HT(cm) - 0.023 * age - 2.258,$$
$$VC_f^{JSR} = 0.032 \times HT(cm) - 0.018 * age - 1.178 \quad (3)$$

Where VC_m^{JSR} is the estimated lung volume of male, VC_f^{JSR} is the estimated lung volume of female, age is the age of the subject, and HT is the height of the subject in cm.

These equations have been adopted in clinical guidelines in Japan and are particularly important for studies involving Asian cohorts. The use of population-specific reference values is essential for the accurate diagnosis and monitoring of respiratory function. The Japanese Respiratory Society (JRS) approach also provides official guidelines [11], offering standardized equations for spirometry reference values that are specifically tailored to the Japanese population. Furthermore, the JRS equations have been utilized as a foundation for developing spirometry-derived lung age equations in previous studies [12].

Table 1 shows the example of lung volume estimation from both Baldwin and JSR techniques. The mean absolute difference between the two sets is 206.50 ml.

TABLE 1: EXAMPLE OF LUNG VOLUME ESTIMATION

subject ID	Lung Volume (VC) Estimation (ml)	
	VC^{BW}: Baldwin	VC^{JSR}: JSR
001	3,169.692	3,376.000
002	2,221.924	2,352.000
003	2,757.888	3,038.000

In this research, the green landmarks are detected and return its coordinates (L_{xi}, L_{yi}) in order to track the chest movement overtime where i represents the index (point) of landmarks. The feature set is collected for each breath cycle. The time steps each breathing cycle was sampled 10 frames per second for a single breathing cycle. The tracked landmarks are collected as six points, ten points , and three points respecively. The number of features is calculated by multiplying the tracked landmarks (L_{xi}, L_{yi}) to the (x, y, z) points. Therefore, the six-points

approach has 6×3 dimensions per frame, the ten-points approach has 10×3 dimensions per frame, and the three-points approach has 3×3 dimensions per frame. The velocity vector (V) is determined on the landmark point between the sampling time frames (L_{xi}, L_{yi}, L_{zi}) and $(L'_{xi}, L'_{yi}, L'_{zi})$ using Euclidean distance as shown in Eq. 4.

$$V_i = \sqrt{((L_{xi} - L'_{xi})^2 + (L_{yi} - L'_{yi})^2 + (L_{zi} - L'_{zi})^2)} \quad (4)$$

where i represents the landmark index, (L_{xi}, L_{yi}, L_{zi}) localizes the landmark point at the current frame, while the next frame is represented as $(L'_{xi}, L'_{yi}, L'_{zi})$.

The feature extraction is applied to these three approachs such as (1) Average velocity vector (V_{avg}) across all frames that gives a central baseline of the breathing depth, (2) standard deviation (V_{SD}) shows how much variation there is in depth or how much the chest is moving, (3) minimum velocity vector (V_{min}) is the closest the green point gets to the camera (chest fully exhaled), (4) maximum velocity vector (V_{max}) is the farthest distance (chest fully inhaled), and (5) Peak-to-peak velocity vector (max - min) (V_{p-p}) is total movement span.

These features capture key aspects of chest movement during breathing, including position trends (mean, min, max), dynamic range (range and amplitude), and variability (standard deviation). Such statistical descriptors are biomechanically meaningful and widely used in respiratory analysis. For instance, a recent work [13] demonstrated that peak-to-peak amplitude of chest movement, derived from smartphone sensors, correlates strongly with tidal volume measured by spirometry. Similarly, a review [14] emphasized that depth-based plethysmography effectively uses statistical features like mean and amplitude to monitor respiratory volume changes non-invasively. In another study [15] used video-based techniques to extract temporal and statistical features from chest motion to classify breathing patterns, including apnea and tachypnea. Together, they provide biomechanically meaningful insights into a person's breathing pattern. For instance, shallow breathers tend to exhibit a low range of motion (low V_{p-p}), while irregular or inconsistent breathing is reflected in a high standard deviation (V_{SD}) of the depth signal.

3. Results and Evaluation

To evaluate the model, it provides set of target values (VC^{BW} *and* VC^{JSR}): $y = y_1, y_2, \ldots y_n$ as shown in Table 1 and a set of predicted values from our proposed model $VC^{predicted}$: $\hat{y} = \widehat{y_1}, \widehat{y_2}, \ldots, \widehat{y_3}$. The dataset is collected from 10 subjects. The linear regression (LR) and Random Forest Regressor (FR) is applied in order to evaluate the VC (ml). Model performance is assessed using cross-validation, with test set sizes of 20% and 30%. Feature sets are extracted from three different anatomical point-tracking configurations: the six-point approach (F_6), ten-point approach (F_{10}), and three-point approach (F_3). The model evaluation metrics Eq.(5)-(8) include:

Coefficient of Determination (R^2) represents the proportion of variance in the dependent variable that is explained by the independent variables in the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (5)$$

Mean Absolute Error (MAE) is average of absolute differences between predicted and target values. It measures how far predictions are from target values on average.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (6)$$

Mean Squared Error (MSE) is verage of squared differences. It penalizes larger errors more due to squaring. Sensitive to outliers.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (7)$$

Root Mean Squared Error (RMSE) is sqare root of MSE.

$$RMSE = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(y_i - \hat{y}_i)^2} \quad (8)$$

where y_i is an target value, $\hat{y}$ is a predicted value, $\bar{y}$ is mean of target values, and n is number of observations.

Table 2: Model Evaluation

Model	Evaluation metrics			
	R^2	MAE (ml)	MSE (ml^2)	RMSE (ml)
F_6-LR-20	0.60	201.59	46193.07	214.92
F_{10}-LR-20	0.86	168.19	29742.31	172.45
F_3-LR-20	0.88	168.19	29742.31	172.45
F_6-LR-30	0.71	205.09	47691.20	218.38
F_{10}-LR-30	0.82	168.19	29742.31	172.45
F_3-LR-30	0.87	168.19	29742.31	172.45
F_6-RF-20	0.74	172.41	56421.48	237.53
F_{10}-RF-20	0.74	172.62	57421.48	239.62
F_3-RF-20	0.70	207.07	64781.65	254.52
F_6-RF-30	0.75	189.56	54527.95	233.51
F_{10}-RF-30	0.74	172.62	57421.87	239.62
F_3-RF-30	0.67	193.51	72176.43	268.65

Among the tested models as shown in Table 2, Linear Regression (LR) consistently outperformed the others across all feature sets, particularly when applied to the F_3 feature set (three-point tracking). The F_3-LR-20 model, based on a three-point chest tracking configuration and linear regression, produced the most accurate lung volume estimates in this study. The regression equation derived from the model is:

$$\hat{y} = 1140.630 + 129.134\,V_{max} + 178.542\,V_{min} - 8.031\,V_{SD} - 49.408\,V_{p-p} + 78.216\,V_{avg} \quad (9)$$

The Eq.9 indicates that lung volume increases with higher maximum and minimum chest depth values, as well as average velocity vector, suggesting a direct correlation between overall chest expansion and predicted vital capacity ($VC^{predicted}$). Conversely, the model assigns negative coefficients to standard deviation

and peak-to-peak range, which may reflect irregular or unstable breathing patterns associated with lower lung volume accuracy.

4. DISCUSSION AND CONCLUSION

This study explored the use of machine learning regression techniques to estimate lung volume from depth-derived chest motion features. Three feature sets (F_6, F_{10}, and F_3), representing different anatomical point-tracking configurations, were used to extract statistical descriptors such as maximum, minimum, standard deviation, peak-to-peak value, and average velocity vector. We aim to determine which feature configuration and regression method best predicted lung volume measured in milliliters (mL). Table 2 presents the performance of various regression models using different feature sets (F_6, F_{10}, F_3), algorithms (Linear Regression and Random Forest), and test set proportions (20% and 30%). Among all evaluated models, F_3-LR-20, which combines the three-point feature set with linear regression and a 20% test split, achieved the highest coefficient of determination ($R^2 = 0.88$), indicating that it explains 88% of the variance in the target variable. This model also yielded the lowest mean absolute error (MAE = 168.19 mL) and a root mean squared error (RMSE = 172.45 mL), demonstrating strong predictive accuracy and consistency. In contrast, Random Forest models exhibited slightly lower R^2 values and higher RMSEs, suggesting potential overfitting or less effective generalization in this dataset. These results shows that F_3-LR-20 is the optimal model configuration for estimating lung volume based on tracked chest motion points. method for estimating lung volume. The mean absolute difference between the proposed model's predictions and the JSR reference values was 60.00 mL, which is substantially lower than the 169.00 mL difference observed between the Baldwin and JSR equations. This suggests that the proposed model provides a more consistent estimation with respect to the JSR clinical standard. For the future work, the dataset will be expanded to include more diverse subjects and clinical cases to improve model generalizability. In addition, a real-time implementation using embedded systems and validation against gold-standard spirometry will also be pursued to move the system toward clinical deployment.

ACKNOWLEDGEMENT

This work was jointly supported by College of Arts, Media and Technology, Chiang Mai University and JSPS KAKENHI Grant Number 23K11950.

REFERENCES

[1] J. Wenger et al., "Standardisation of the measurement of lung volumes", *European Respiratory Journal*, Vol. 26(3): pp. 511-522, 2005

[2] C. Crimi, P. Impellizzeri, R. Campisi, S. Nolasco, A. Spanevello, N. Crimi, "Practical considerations for spirometry during the COVID-19 outbreak: Literature review and insights", *Pulmonology*, Vol. 27. Issue 5, pp. 438-447, 2021

[3] C. Romano, E. Schena, S. Silvestri, C. Massaroni, "Non-Contact Respiratory Monitoring Using an RGB Camera for Real-World Applications", *Sensors*, Vol. 21, No. 15, 5126, 2021

[4] M. Eugenia P. Reyes, J. D. Palmero, J. L. Diaz, E. Aragon, A. Taboada-Crispi, "Computer Vision-Based Estimation of Respiration Signals", in proc. VIII Latin American Conference on Biomedical Engineering and XLII National Conference on Biomedical Engineering, pp 252-261, Mexico, 2019

[5] B. A. Reyes, N. Reljin, Y. Kong, Y. Nam and K. H. Chon, "Tidal Volume and Instantaneous Respiration Rate Estimation using a Volumetric Surrogate Signal Acquired via a Smartphone Camera", *IEEE Journal of Biomedical and Health Informatics*, Vol. 21, No. 3, pp. 764-777, May 2017

[6] H. E. Boussaki, R. Latif, A. Saddik, Z. E. Khadiri, H. E. Boujaoui, "Non-contact Respiratory Rate Monitoring Based on the Principal Component Analysis", International Journal of Advanced Computer Science and Applications, Vol. 14, No. 9, 2023

[7] Baldwin, "Vital Capacity Calculation," Osaka Medical and Pharmaceutical University, [Online]. Available:https://www.ompu.ac.jp/udeps/in1/res/calc/VC.html.

[8] Helgeson SA, Quicksall ZS, Johnson PW, et al. "Estimation of Static Lung Volumes and Capacities From Spirometry Using Machine Learning: Algorithm Development and Validation" JMIR AI. 2025;1(1)

[9] Fan Z, Zhao S, Wang L, et al, "Comparison between functional lung volume measurement and segment counting for predicting postoperative pulmonary function after pulmonary resection in lung cancer patients", BMC Pulm Med. 2023;23:6.

[10] The Japanese Respiratory Society (JRS), "Reference values for spirometry," [Online]. Available: https://www.jrs.or.jp/modules/en/index.php?content_id=1.

[11] Shimizu Y, Dobashi K, Sato K, et al, "Novel equations for spirometry-derived lung age using data from a large Japanese population: the JRS Lung Age Study", Respir Investig. 2015;53(1):1-8.

[12] Japanese Respiratory Society, "Lung function testing: the Official Guideline of the Japanese Respiratory Society". Respirology. 2004;9(Suppl 1):S1-S46.

[13] Reyes, B. A., Reljin, N., & Chon, K. H., "Tidal Volume and Instantaneous Respiration Rate Estimation Using a Smartphone" IEEE Journal of Biomedical and Health Informatics, 21(2), 452–460.

[14] Alsufyani, A., Park, E. J., & Kulić, D., "Depth-Based Measurement of Respiratory Volumes: A Review". Sensors, 22(24), 9787. PMCID: PMC9785978

[15] Yang, Z., et al., "Respiratory Feature Extraction for Contactless Breathing Pattern Recognition Using a Single Digital Camera", Sensors, 23(6), 3035. DOI:10.3390/s23063035