Author:
Soleimani, Vahid

Title:
Remote Depth-based Photoplethysmography in Pulmonary Function Testing

Remote Depth-based Photoplethysmography in Pulmonary Function Testing

Vahid Soleimani

A dissertation submitted to the University of Bristol in accordance with the requirements for the degree of Doctor of Philosophy in the Faculty of Engineering, Visual Information Laboratory.

June 2018

32046 words

Abstract

This thesis introduces several novel, noninvasive lung function assessment approaches in which we incorporate computer vision techniques to remotely compute standard clinical Pulmonary Function Testing (PFT) measures.

Using a single depth sensor, a dynamic 3-D model of a subject's chest is reconstructed and used to generate chest volume–time data by estimating the chest volume variation throughout a sequence. Following computation of multiple keypoints and calibration of volume–time data to present real volume of exchanged air, 7 Forced Vital Capacity (FVC) measures and 4 Slow Vital Capacity (SVC) measures are computed. Evaluation on a dataset of 85 patients (529 sequences), attending a respiratory outpatient service for spirometry, shows a high correlation between the proposed depth-based PFT measures and the measures from a spirometer.

Trunk motion during PFT affects the accuracy of these results, so the natural reaction of the subject's body to maximal inhalation and exhalation, must be decoupled from the chest-surface breathing motion. We present an automatic, open source data acquisition and calibration pipeline in which two opposing depth sensors are calibrated and used to reconstruct a well-defined dynamic 3-D model of the trunk during PFT performance. Our proposed method is able to reconstruct dynamic 3-D models with accurate temporal frame synchronisation and spatial registration. Then, we propose a whole body depth-based photoplethysmography (dPPG) approach which allows subjects to perform PFT, as in routine spirometry, without restraining their natural trunk reactions. By decoupling the trunk movement and the chest-surface respiratory motion, dPPG obtains more accurate respiratory volume–time data which improves the accuracy of the estimated PFT measures. A dataset spanning 35 subjects (298 sequences) was collected and used to illustrate the superiority of the proposed dPPG method by comparing its measures to those provided by a spirometer and the single Kinect approach.

Although dPPG is able to improve the PFT measures accuracy to a significant extent, it is not able to filter complex trunk motions, particularly at the *deep forced* inhalation–exhalation stage. To effectively correct trunk motion artifacts further, we propose an active trunk shape modelling approach by which the respiratory volume–time data is computed by performing principal component analysis on temporal 3-D geometrical features, extracted from the chest and posterior shape models in $\mathbb{R}^3$ space. We validate the method's accuracy at the signal level by computing several comparative metrics between the depth-based and spirometer volume–time data. Evaluating on the dPPG PFT dataset (300 PFT sequences), our trunk shape modelling approach outperforms the single Kinect and dPPG methods.

Declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original, except where indicated by special reference in the text, and no part of the dissertation has been submitted for any other academic award.

Any views expressed in the dissertation are those of the author and in no way represent those of the University of Bristol.

The dissertation has not been presented to any other University for examination either in the United Kingdom or overseas.

SIGNED: Vahid Soleimani DATE: 19/07/2018

Acknowledgements

First of all, I would like to thank my supervisor, Professor Majid Mirmehdi, for his scientific support, invaluable guidance and continuous encouragement over these four years. Further, I would like to thank our medical support team, Dr James Dodd, Dr Charles Sharp and Jason Viner for helping with medical justifications, and collaboration in data collection at Southmead Hospital Bristol. I would also like to thank patients from Southmead Hospital Bristol and everybody who participated in my breathing test experiments.

I particularly thank Dr Massimo Camplani and Dr Sion Hannuna for their friendly support and accompany, and consistent encouragements over these years.

I would also like to thank my friends within the Visual Information Laboratory, Shuda Li, Toby Perrett, Teesid Leelasawassuk, Austin Gregg-Smith, Akshaya Thippur, Oliver Moolan-Feroze, Yangdi Xu, Hongping Cai, Adeline Paiement, Jack Greenhalgh, David Hanwell, Osian Haines, Jose Martinez-Carranza, Laurie Bose, Davide Moltisanti, Michael Wray, Erik Bodin, Miguel Lagunez-Fortiz, Eduardo Ruiz Liberos, Luis Contreras Toledo, Samuel Whitehouse and Hazel Doughty. I would also like to thank Dr Neill Campbell, Dr Tilo Burghardt, Dr Carl Henrik Ek, Jen Hawkins and Sarah Rogers.

Finally, my appreciation goes to SPHERE[1], an EPSRC Interdisciplinary Research Centre, for providing the opportunity of doing this research. I would like to thank Professor Ian Craddock, SPHERE director, and my friends within SPHERE office, Dr Lili Tao, Dr Mo Haghighi, Dr Pete Woznowski, Dr Tom Diethe, Ben Miller and Hao Song.

I would like to thank the University of Bristol Alumni and Friends Foundation for funding my PhD.

Last but not least, my special thanks go to my wife, Farnoosh, and my daughter, Avina, without whom it would not be possible to get through these tough four years.

[1] *Sensor Platform for HEalthcare in a Residential Environment.*

Publications

The work described in this thesis has been presented in the following publications:

1. **Vahid Soleimani**, Majid Mirmehdi, Dima Damen and James Dodd, "Body Motion Artifact Correction in Vision-based Respiratory Function Assessment", *Journal Computer Methods in Biomechanics and Biomedical Engineering*, Under review. [**Impact Factor: 1.909**]

2. **Vahid Soleimani**, Majid Mirmehdi, Dima Damen and James Dodd, "Marker-less Active Trunk Shape Modelling for Motion Tolerant Remote Respiratory Assessment", *The 25th IEEE International Conference on Image Processing (ICIP)*, Accepted for publication, 2018.

3. **Vahid Soleimani**, Majid Mirmehdi, Dima Damen, Massimo Camplani, Sion Hannuna, Charles Sharp, James Dodd, "Depth-based Whole Body Photoplethysmography in Remote Pulmonary Function Testing", *IEEE Transactions on Biomedical Engineering*, Vol. 65, No. 6, pp. 1421-1431, 2018. [**Impact Factor: 4.288**]

4. **Vahid Soleimani**, Majid Mirmehdi, Dima Damen, James Dodd, Sion Hannuna, Charles Sharp, Massimo Camplani, Jason Viner, "Remote, Depth-based Lung Function Assessment", *IEEE Transactions on Biomedical Engineering*, Vol. 64, No. 8, pp. 1943-1958, 2017. [**Impact Factor: 4.288**]

5. **Vahid Soleimani**, Majid Mirmehdi, Dima Damen, Sion Hannuna, Massimo Camplani, "3D Data Acquisition and Registration using Two Opposing Kinects", *in Proceeding of International Conference on 3D Vision (3DV)*, pp. 128-137, Stanford, 2016.

6. **Vahid Soleimani**, Majid Mirmehdi, Dima Damen, Sion Hannuna, Massimo Camplani, Jason Viner, James Dodd, "Remote Pulmonary Function Testing using a Depth Sensor", *in Proceeding of IEEE/CAS–EMB Biomedical Circuits and Systems (BioCAS)*, Atlanta, 2015.

Also, the medical significance of this research has been published in:

7. Charles Sharp, **Vahid Soleimani**, Sion Hannuna, Massimo Camplani, Dima Damen, Jason Viner, Majid Mirmehdi, James Dodd, "Towards respiratory assessment using depth measurements from a time-of-flight sensor", *Frontiers in Physiology*, Vol. 8, No. 65, 2017. [**Impact Factor: 4.134**]

8. Charles Sharp, **Vahid Soleimani**, Sion Hannuna, Massimo Camplani, Dima Damen, Jason Viner, Majid Mirmehdi, James Dodd, "Remote pulmonary function testing computer gaming in the respiratory world", *Thorax*, Vol. 70, pp. 117-118, 2015.

To my beloved wife, Farnoosh
and
my sweet daughter, Avina.

Contents

List of Figures

List of Tables

List of Acronyms

API . . . . . Application Programming Interface

ATS/ERS . . . . . American Thoracic Society/European Respiratory Society

BF . . . . . Bilateral Filter

BMI . . . . . Body Mass Index

bpm . . . . . Breath per Minute

BRIEF . . . . . Binary Robust Independent Elementary Features

CC . . . . . Correlation Coefficient

CMOS . . . . . Complementary MetalOxideSemiconductor

COPD . . . . . Chronic Obstructive Pulmonary Disease

dPPG . . . . . depth-based Photoplethysmography

D . . . . . Depth

DTW . . . . . Dynamic Time Warping

EMD . . . . . Empirical Mode Decomposition

ERV . . . . . Expiratory Reserve Volume (SVC Test Measure)

$\boldsymbol{FEF}_{25\%}$ . . . . . Forced Expiratory Flow 25% (FVC Test Measure)

$\boldsymbol{FEF}_{50\%}$ . . . . . Forced Expiratory Flow 50% (FVC Test Measure)

$\boldsymbol{FEF}_{75\%}$ . . . . . Forced Expiratory Flow 75% (FVC Test Measure)

$\boldsymbol{FEF}_{25-75\%}$ . . . . . Forced Expiratory Flow 25-75% (FVC Test Measure)

FEV1 . . . . . Forced Expiratory Volume in 1 Second (FVC Test Measure)

FGOS . . . . . Fiber Grating Optical Sensor

FIR . . . . . Finite Impulse Response

FOV . . . . . Field Of View

FPFH . . . . . Fast Point Feature Histograms

fps . . . . . Frames per Second

FRC . . . . . Functional Residual Capacity

FRD . . . . . Fréchet Distance

FVC . . . . . Forced Vital Capacity

FVC . . . . . Forced Vital Capacity (FVC Test Measure)

GOLD . . . . . Global Initiative for Chronic Obstructive Lung Disease

GPU . . . . . Graphics Processing Unit

H . . . . . Height

HD . . . . . High Definition

IC . . . . . Inspiratory Capacity (SVC Test Measure)

ICA . . . . . Independent Component Analysis

ICP . . . . . Iterative Closest Point

IMF . . . . . Intrinsic Mode Functions

LDVi . . . . . Laser Doppler Vibrometer

LIDAR . . . . . Light Imaging Detection And Ranging

LOOCV . . . . . Leave-One-Out Cross-Validation

LSTM . . . . . Long Short Term Memory

ME . . . . . Main Effort

MISO . . . . . Multi-Input–Single-Output

MoCap . . . . . Motion Capture

MSAC . . . . . M-estimator SAmple Consensus

MSE . . . . . Mean Square Error

NIR . . . . . Near InfraRed

NL_2 . . . . . Normalised L_2 Error

NTP . . . . . Network Time Protocol

OpenNI . . . . . Open Natural Interaction

ORB . . . . . Oriented fast and Rotated BRIEF

PC . . . . . Personal Computer

PC . . . . . Principal Component

PCA . . . . . Principal Component Analysis

PDF . . . . . Probability Distribution Function

PEF . . . . . Peak Expiratory Flow (FVC Test Measure)

PF . . . . . Peak Flow

PFF . . . . . Principal Flow Field

PFT . . . . . Pulmonary Function Testing

PP . . . . . Pulsus Paradoxus

PPG . . . . . Photoplethysmography

PPLV . . . . . Photo-EMF Pulsed Laser Vibrometer

RADAR . . . . . RAdio Detection And Ranging

RGB . . . . . Red, Green, Blue

RGB–D . . . . . Red, Green, Blue – Depth

RM . . . . . Respiratory Monitoring

RMM . . . . . Respiratory Motion Modelling

RMSE . . . . . Root Mean Square Error

ROI . . . . . Region of Interest

RMP . . . . . Respiratory Motion Prediction

RR . . . . . Respiratory Rate

RSV . . . . . Respiratory Syncytial Virus

RV . . . . . Residual Volume

SDK . . . . . Software Development Kit

SIFT . . . . . Scale Invariant Feature Transform

SL . . . . . Structured Light

SPAD . . . . . Single-Photon Avalanche Diode

SpO_2 . . . . . Pulse Oximeter

STD . . . . . Standard Deviation

SURF . . . . . Speeded-Up Robust Features

SVC . . . . . Slow Vital Capacity

TBB . . . . . Threading Building Blocks

TLC . . . . . Total Lung Volume

ToF . . . . . Time of Flight

TV . . . . . Tidal Volume (SVC Test Measure)

TV . . . . . Tidal Volume

VC . . . . . Vital Capacity (SVC Test Measure)

VTK . . . . . Visualization Toolkit

W . . . . . Width

List of Symbols

General Symbols

$\mathcal{A}$	ME Minimum Keypoint
$\mathcal{B}$	ME Maximum Keypoint
$\mathcal{C}$	TV Beginning Keypoint
$\mathcal{D}$	TV End Keypoint
$\mathcal{E}$	ME Begining Keypoint
$\mathcal{F}_i$	TV Local Maxima Keypoints
F_{me}	ME Similarity Measure
F_{tv}	TV Similarity Measure
$\mathcal{G}_i$	TV Local Minima Keypoints
L_2	Absolute Euclidean Norm Distance
n_{me}	Number of ME Training Trials
n_{tv}	Number of TV Training Trials
R^2	Correlation of Determination
$\mathbb{R}^3$	Real-world Coordinate System
R_{aw}	Airway Resistance
$t_{\mathcal{A}}$	Keypoint $\mathcal{A}$ Timestamp
$t_{\mathcal{B}}$	Keypoint $\mathcal{B}$ Timestamp
$t_{\mathcal{F}_i}$	Keypoints $\mathcal{F}_i$ Timestamp
$t_{\mathcal{G}_i}$	Keypoints $\mathcal{G}_i$ Timestamp

τ_{delay} . . . . . Time Delay

$V(t)$ . . . . . Volume–time Data

$\dot{V}(t)$ . . . . . Flow–time Data

$V_{dk}(t)$ . . . . . dPPG Depth-based Volume–time Data

$V_s(t)$ . . . . . Spirometer Volume–time Data

$V_{sk}(t)$ . . . . . Single–Kinect Depth-based Volume–time Data

Symbols from Chapter 2

CC_L . . . . . Lower Correlation Bound

CC_U . . . . . Upper Correlation Bound

$D_t(x, y)$ . . . . . Pixel Displacement at Time t

$F_t(x, y)$ . . . . . Image Flow along Gradient at Time t

$\overrightarrow{G_t}(x, y)$ . . . . . Image Gradient Vector at Time t

I_B . . . . . Binary Image of I_D^m

I_D . . . . . Depth Sequence

I_D^m . . . . . Mean Image of Sequence I_D

I_M . . . . . Identity Matrix

k_1, k_2 . . . . . Intra-subject Normalisation Coefficients

M_x . . . . . median length of row-wise connected components of I_B

M_y . . . . . median length of column-wise connected components I_B

P_{air} . . . . . Atmospheric Pressure

P_{lung} . . . . . Lung Pressure

$P_{lung}^{(max)}$ . . . . . Lung Pressure at the End of Exhalation

$P_{lung}^{(min)}$ . . . . . Lung Pressure at the End of Inhalation

Q_f . . . . . Air Flow

$\mathcal{R}_{chest}$ . . . . . Chest Region Bounding Box

T_L . . . . . Upper Threshold Computed from Histogram of I_D^m

T_U . . . . . Lower Threshold Computed from Histogram of I_D^m

τ . . . . . Respiratory Interval

V_{lung} . . . . . Lung Volume

$\mathbf{V}(x, y)$ . . . . . Flow Field Projection Direction at Pixel (x, y)

Symbols from Chapter 4

$\mathbf{E}_{ij}$ . . . . . Vector jth of Facet S_{T_i}

$\mathbf{F}$ . . . . . Vector Field

G_{σ_r} . . . . . BR Range Gaussian Kernel

G_{σ_s} . . . . . BF Spatial Gaussian Kernel

γ . . . . . Chest Volume Threshold

$\mathbf{n}$ . . . . . Surface Normal Vector

$\mathbf{n}_{ch}$ . . . . . Chest Surface Normal Vector

n_{FS} . . . . . Number of FVC and SVC Training Trials

n_S . . . . . Number of SVC Training Trials

$\mathbf{n}_l$ . . . . . Lateral Surface Normal Vector

$\mathbf{n}_r$ . . . . . Reference Plane Surface Normal Vector

N_t . . . . . Total Number of Chest Surface Triangles

$\mathbf{n}_{T_i}$ . . . . . Normal Vector of S_{T_i}

N_W . . . . . Averaging Window Size

$\mathbf{P}_i$ . . . . . Parametric Representation of S_{T_i}

ρ . . . . . Chest Volume Normalisation Factor

S . . . . . Closed Piecewise Smooth Surface

S_{ch} . . . . . Chest Surface

S_l . . . . . Lateral Surface

S_r . . . . . Reference Plane Surface

S_{T_i} . . . . . Surface of Triangle T_i

σ_r . . . . . BF Range Parameter

σ_s . . . . . BF Spatial Parameter

t_0 . . . . . Timestamp of 'Time Zero' Keypoint

T_i . . . . . ith Triangle of Chest Surface

t_i^x . . . . . Timestamp of ith Element of X

t_i^y . . . . . Timestamp of ith Element of Y

t_{PF} . . . . . Timestamp of PF Keypoint

$V_{in}(k)$ . . . . . Noisy Volume–time Data

$V_k(t)$ . . . . . Kinect Volume–time Data

$\widehat{V}_k^{tv}$ . . . . . Aligned Kinect TV Data

V_{max} . . . . . Global Maximum of Volume–time Data

V_{min} . . . . . Global Minimum of Volume–time Data

$V_{out}(k)$ . . . . . Smoothed Volume–time Data

$\widehat{V}_s^{tv}$ . . . . . Sub-sampled and Aligned Spirometer TV Data

V_t . . . . . Estimated Volume Enclosed by S_{ch}, S_l, S_r at Frame t

W_f . . . . . BF Window

W_p . . . . . BF Normalization Factor

X . . . . . Volume–time Data Minima Set

Y . . . . . Volume–time Data Maxima Set

(ξ_{tv}, ψ_{tv}) . . . . . TV Scaling Factor and Offset

(ξ_{me}, ψ_{me}) . . . . . ME Scaling Factor and Offset

$\left\{\xi_{tv}^{\ell}\right\}_{\ell=1}^{n_{tv}}$ . . . . . TV Training Scaling Factors

$\left\{(\xi_{me}^{\ell}, \psi_{me}^{\ell})\right\}_{\ell=1}^{n_{me}}$ . . . . ME Training Scaling Factors and Offsets

(ξ'_{tv}, ψ'_{tv}) . . . . . Intra-subject Estimated TV Scaling Factor and Offset

(ξ'_{me}, ψ'_{me}) . . . . . Intra-subject Estimated ME Scaling Factor and Offset

3-D Chest-wall Modelling Based Validation

λ_v . . . . . CC of Intra-test PFT Measures

λ'_v . . . . . CC of Intra-subject PFT Measures

μ_v . . . . . Mean of Intra-test PFT Measures' L_2 Error

μ'_v . . . . . Mean of Intra-subject PFT Measures' L_2 Error

Ω_v . . . . . Normalised μ_v

Ω'_v . . . . . Normalised μ'_v

σ_v . . . . . STD of Intra-test PFT Measures' L_2 Error

σ'_v . . . . . STD of Intra-subject PFT Measures' L_2 Error

Chest-wall Averaging Based Validation

λ_m . . . . . CC of Intra-test PFT Measures

λ'_m . . . . . CC of Intra-subject PFT Measures

μ_m . . . . . Mean of Intra-test PFT Measures' L_2 Error

μ'_m . . . . . Mean of Intra-subject PFT Measures' L_2 Error

Ω_m . . . . . Normalised μ_m

Ω'_m . . . . . Normalised μ'_m

σ_m . . . . . STD of Intra-test PFT Measures' L_2 Error

σ'_m . . . . . STD of Intra-subject PFT Measures' L_2 Error

Intra-subject Scaling Factors Analysis

$\texttt{SCE}^{tv}$ . . . . . ξ'_{tv} Error Analysis Measure

$\texttt{SCE}^{me}$ . . . . . ξ'_{me} Error Analysis Measure

$\texttt{SME}^{tv}$ . . . . . F_{tv} Performance Analysis Measure

$\texttt{SME}^{me}$ . . . . . F_{me} Performance Analysis Measure

ξ^o_{tv} . . . . . Original TV Scaling Factor of Test Trial

ξ^o_{me} . . . . . Original ME Scaling Factor of Test Trial

ξ^c_{tv} . . . . . Closest TV Scaling Factor to ξ^o_{tv}

ξ^c_{me} . . . . . Closest ME Scaling Factor to ξ^o_{me}

Symbols from Chapter 5

$\mathbf{M}$ . . . . . Cross Covariance Squar Matrix

$\mathbf{P}$ . . . . . Final Reconstructed Point Cloud

$\mathbf{P}_2$ . . . . . Second Point Cloud

$\mathbf{P'}_2$ . . . . . Transformed $\mathbf{P}_2$

$\mathbf{P}_1^{ref}$ . . . . . Reference Point Cloud

$\mathbf{Q}$ . . . . . Chessboard Crossing Point Set in the Reference Space

$\mathbf{Q}'$ . . . . . Chessboard Crossing Point Set in the Second Chessboard

$\mathbf{Q}_j$ . . . . . jth point in $\mathbf{Q}$

$\mathbf{Q}'_j$ . . . . . jth point in $\mathbf{Q}'$

$\mathbf{Q}_\mu$ . . . . . Centroid of $\mathbf{Q}$

$\mathbf{Q}'_\mu$ . . . . . Centroid of $\mathbf{Q}'$

$\mathbf{R}$ . . . . . Estimated Rotation Matrice

S1-4 . . . . . 4 Sides of a Box

$\mathbf{\Sigma}$ . . . . . Decomposed Diagonal Non-negative Scaling Matrice

$\mathbf{T}$ . . . . . Estimated Translation Matrice

$\mathbf{U}$ . . . . . Decomposed Orthogonal Rotation Matrice

$\mathbf{V}$ . . . . . Decomposed Orthogonal Rotation Matrice

Symbols from Chapter 6

$\left\{F_{me}^{\ell}\right\}_{\ell=1}^{n_{me}}$ . . . . . ME Similarity Measues of $\left\{\left\langle V_{dk}(t)\right\rangle^{\ell}\right\}_{\ell=1}^{n_T}$

$\left\{F_{tv}^{\ell}\right\}_{\ell=1}^{n_{tv}}$ . . . . . TV Similarity Measues of $\left\{\left\langle V_{dk}(t)\right\rangle^{\ell}\right\}_{\ell=1}^{n_T}$

F_{me}^{test} . . . . . ME Similarity Measure of Test Trial

F_{tv}^{test} . . . . . TV Similarity Measure of Test Trial

$\mathcal{L}_{dk}$ . . . . . dPPG Bland-Altam Upper Limit of Agreement

$\mathcal{L}_{sk}$ . . . . . Single–Kinect Bland-Altam Upper Limit of Agreement

λ_{dk} . . . . . CC of dPPG PFT Measures

λ_{sk} . . . . . CC of Single–Kinect PFT Measures

$\mathtt{M}'$ . . . . . Absolute Difference between $\mathtt{M}_x^{tv}$ and $\mathtt{M}_x^{me}$ where $x = dk|sk$

$\mathcal{M}_{dk}$ . . . . . Percentage of dPPG Measures lies in $\mathcal{L}_{dk}-\mathcal{U}_{dk}$

$\mathtt{M}_{dk}^{me}$ . . . . . Mean of dPPG Within-subject ME Scaling Factors

$\mathtt{M}_{dk}^{tv}$ . . . . . Mean of dPPG Within-subject TV Scaling Factors

$\mathcal{M}_{sk}$ . . . . . Percentage of Single–Kinect Measures lies in $\mathcal{L}_{dk}-\mathcal{U}_{dk}$

$\mathtt{M}_{sk}^{me}$ . . . . . Mean of Single–Kinect Within-subject ME Scaling Factors

$\mathtt{M}_{sk}^{tv}$ . . . . . Mean of Single–Kinect Within-subject TV Scaling Factors

μ_{dk} . . . . . Mean of dPPG PFT Measures' L_2 Error

μ_{sk} . . . . . Mean of Single–Kinect PFT Measures' L_2 Error

$\mu_{\mathtt{M}'}$ . . . . . Mean of $\mathtt{M}'$ Across All Subjects

$\mu_{\Sigma'}$ . . . . . Mean of Σ' Across All Subjects

n_T . . . . . Total Number of Training Trials

Ω_{dk} . . . . . Normalised μ_{dk}

Ω_{sk} . . . . . Normalised μ_{sk}

$\Omega_{\mathtt{M}'}$ . . . . . Normalised $\mu_{\mathtt{M}'}$

$\Omega_{\Sigma'}$ . . . . . Normalised $\mu_{\Sigma'}$

$\mathtt{SCE}_{dk}^{me}$ . . . . . dPPG $\langle \xi_{dk}^{me} \rangle^{\jmath_{me}}$ Error Analysis Measure

$\mathtt{SCE}_{dk}^{tv}$ . . . . . dPPG $\langle \xi_{dk}^{tv} \rangle^{\jmath_{tv}}$ Error Analysis Measure

$\mathtt{SCE}_{sk}^{me}$ . . . . . Single–Kinect $\langle \xi_{sk}^{me} \rangle^{\jmath'_{me}}$ Error Analysis Measure

$\mathtt{SCE}_{sk}^{tv}$ . . . . . Single–Kinect $\langle \xi_{sk}^{tv} \rangle^{\jmath'_{tv}}$ Error Analysis Measure

$\mathtt{SME}_{dk}^{me}$ . . . . . dPPG F_{me} Performance Analysis Measure

$\mathtt{SME}_{dk}^{tv}$ . . . . . dPPG F_{tv} Performance Analysis Measure

$\mathtt{SME}_{sk}^{me}$ . . . . . Single–Kinect F_{me} Performance Analysis Measure

$\mathtt{SME}_{sk}^{tv}$ . . . . . Single–Kinect F_{tv} Performance Analysis Measure

σ_{dk} . . . . . STD of dPPG PFT Measures' L_2 Error

$\sigma_{\mathtt{M}'}$ . . . . . STD of $\mathtt{M}'$ Across All Subjects

σ_{sk} . . . . . STD of Single–Kinect PFT Measures' L_2 Error

$\sigma_{\Sigma'}$ . . . . . STD of Σ' Across All Subjects

Σ' . . . . . Absolute Difference between Σ_x^{tv} and Σ_x^{me} where $x = dk|sk$

Σ_{dk}^{me} . . . . . STD of dPPG Within-subject ME Scaling Factors

Σ_{dk}^{tv} . . . . . STD of dPPG Within-subject TV Scaling Factors

Σ_{sk}^{me} . . . . . STD of Single–Kinect Within-subject ME Scaling Factors

Σ_{sk}^{tv} . . . . . STD of Single–Kinect Within-subject TV Scaling Factors

$\mathcal{U}_{dk}$ . . . . . dPPG Bland-Altam Lower Limit of Agreement

$\mathcal{U}_{sk}$ . . . . . Single–Kinect Bland-Altam Lower Limit of Agreement

$V_{ch}(t)$ . . . . . Chest Volume–time Data

$\widehat{V}_{dk}(t)$ . . . . . Detrended and Zero-mean Normalised $V_{dk}(t)$

$\left\{\left\langle V_{dk}(t)\right\rangle^{\ell}\right\}_{\ell=1}^{n_T}$ . . . . . dPPG Training Volume–time Data

$V_{dk}^{test}(t)$ . . . . . dPPG Test Volume–time Data

$\widehat{V}_{dk}^{tv}$ . . . . . TV Segment of $\widehat{V}_{dk}(t)$

$V_{po}(t)$ . . . . . Posterior Wall Volume–time Data

$\widehat{V}_s(t)$ . . . . . Sub-sampled, Detrended and Zero-mean Normalised $V_s(t)$

$V_{sk}^{test}(t)$ . . . . . Single–Kinect Test Volume–time Data

$\widehat{V}_s^{tv}$ . . . . . TV Segment of $\widehat{V}_s(t)$

$(\xi_{dk}^{me}, \psi_{dk}^{me})$ . . . . . dPPG ME Scaling Factor and Offset

$(\xi_{dk}^{tv}, \psi_{dk}^{tv})$ . . . . . dPPG TV Scaling Factor and Offset

$\langle \xi_{dk}^{me}, \psi_{dk}^{me} \rangle^{\jmath_{me}}$ . . . . . dPPG Estimated ME Scaling Factor and Offset

$\langle \xi_{dk}^{tv} \rangle^{\jmath_{tv}}$ . . . . . dPPG Estimated TV Scaling Factor

$(\xi_{sk}^{me}, \psi_{sk}^{me})$ . . . . . Single–Kinect ME Scaling Factor and Offset

$(\xi_{sk}^{tv}, \psi_{sk}^{tv})$ . . . . . Single–Kinect TV Scaling Factor and Offset

$\left\{\langle \xi_{dk}^{me}, \psi_{dk}^{me} \rangle^{\ell}\right\}_{\ell=1}^{n_{me}}$ . . . dPPG ME Training Scaling Factors and Offsets

$\left\{\langle \xi_{dk}^{tv} \rangle^{\ell}\right\}_{\ell=1}^{n_{tv}}$ . . . . . dPPG TV Training Scaling Factors

$\langle \xi_{sk}^{me}, \psi_{sk}^{me} \rangle^{\jmath'_{me}}$ . . . . . Single–kinect estimated ME Scaling Factor and Offset

$\langle \xi_{sk}^{tv} \rangle^{\jmath'_{tv}}$ . . . . . Single–Kinect Estimated TV Scaling Factor

$\langle \xi_{dk}^{me} \rangle^{o}$ . . . . . dPPG Original ME Scaling Factor of Test Trial

$\langle \xi_{dk}^{tv} \rangle^{o}$ . . . . . dPPG Original TV Scaling Factor of Test Trial

$\langle \xi_{dk}^{me} \rangle^{c}$ . . . . . dPPG Closest ME Scaling Factor to $\langle \xi_{dk}^{me} \rangle^{o}$

$\langle \xi_{dk}^{tv} \rangle^{c}$ . . . . . dPPG Closest TV Scaling Factor to $\langle \xi_{dk}^{tv} \rangle^{o}$

$\langle \xi_{sk}^{me} \rangle^{o}$ . . . . . Single–Kinect Original ME Scaling Factor of Test Trial

$\langle \xi_{sk}^{tv} \rangle^{o}$ . . . . . Single–Kinect Original TV Scaling Factor of Test Trial

$\langle \xi_{sk}^{me} \rangle^{c}$ . . . . . Single–Kinect Closest ME Scaling Factor to $\langle \xi_{sk}^{me} \rangle^{o}$

$\langle \xi_{sk}^{tv} \rangle^{c}$ . . . . . Single–Kinect Closest TV Scaling Factor to $\langle \xi_{sk}^{tv} \rangle^{o}$

Symbols from Chapter 7

$\mathbf{C_F}$ . . . . . Covariance Matrix of Feature Matrix $\mathbf{F}$

$\mathbf{F}$ . . . . . Extracted Feature Matrix

$\mathbf{F}_i$ . . . . . Extracted Features for Variable i in $\mathbf{F}$

$\boldsymbol{\mathcal{I}}^t_{ch}$ . . . . . Chest Surface Temporal Interpolant Function

$\boldsymbol{\mathcal{I}}^t_{po}$ . . . . . Posterior Surface Temporal Interpolant Function

λ_i . . . . . The ith Eigenvalue of $\mathbf{C_F}$

$\mu_{\mathbf{F}_i}$ . . . . . Mean Value of Data Samples of Frame i

N_τ . . . . . Number of Independent Variables (Frames) in $\mathbf{F}$

$\mathbf{P}_{ch}$ . . . . . Chest-wall Data Points

$\mathbf{p}_{ij}(t)$ . . . . . 2-D Data points of $\boldsymbol{\mathcal{R}}_{xy}(t)$

$\mathbf{P}_{po}$ . . . . . Posterior-wall Data Points

$\boldsymbol{\mathcal{R}}_{xy}(t)$ . . . . . Temporal 2-D Planar ROI

Υ . . . . . Rank of Covariance Matrix $\mathbf{C_F}$

scl_{me} . . . . . *Main Effort* Scaling Factor

scl_{tv} . . . . . *Tidal Volume* Scaling Factor

t_{max} . . . . . $V_s(t)$ Global Maximum Timestamp

t_{min} . . . . . $V_s(t)$ Global Minimum Timestamp

τ . . . . . PFT Sequence Duration

$V_d(t)$ . . . . . Any Depth-based Volume–time data: $V_{sk}(t)$, $V_{dk}(t)$ or $V_r(t)$

v_i . . . . . The ith Computed Principal Component

$\mathbf{v}_{ij}(t)$ . . . . . Temporal Partial Trunk Volume

$V_r(t)$ . . . . . Trunk Shape Modelling Volume–time Data

Chapter 1

Introduction

Lung function diseases are among the leading causes of death worldwide. In 2015, Chronic Obstructive Pulmonary Disease (COPD) and Asthma affected $\sim$532 million people worldwide, of whom $\sim$ 3.6 million died [78]. COPD describes a long-term progressive lung condition, characterised by airways obstruction which limits respiration airflow and affects normal breathing. It is the fourth major cause of death worldwide and will become the third by killing 4.7 million people in 2020 [7, 118]. In the UK, COPD affects around three million people [12] and it is known as the second largest cause of emergency admission to hospital [8]. In an economic perspective, the annual direct healthcare cost of COPD treatment is estimated as $30 billion in the US [7], €23 billion in the EU [5], and more than £800 million in the UK [11].

Pulmonary Function Testing (PFT) [128] is a group of clinical tests that evaluate human respiratory status and is a vital component of clinical assessment in the investigation of respiratory diseases. Whole–body plethysmography [56] and spirometry [129] are traditional and clinically approved methods for PFT, which require patient co-operation and direct contact with the equipment. In particular, spirometry is known as the *gold standard* approach for COPD diagnosis and assessing its severity [197].

There are also other contact-based techniques and tools which are only used for tidal volume respiratory monitoring and breathing rate estimation, such as: thermistors [98], strain gauge [144], respiration belt [42], inductance plethysmography [54, 86] and impedance pneumotachography [24].

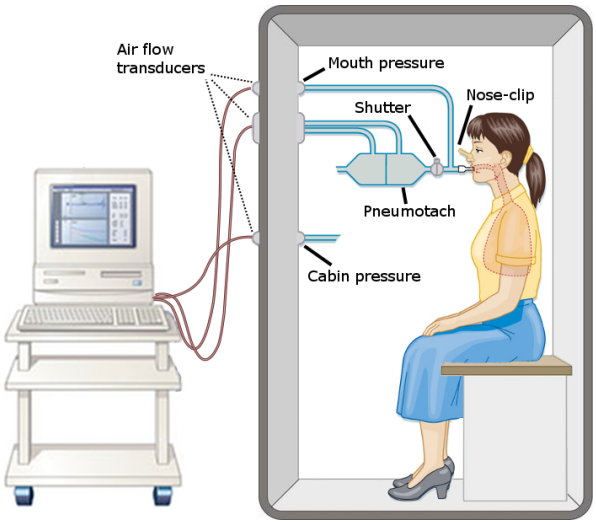


Figure 1.1: *Whole–body plethysmography utilises an air-sealed cabin of known volume and a pneumotach with a mechanically controlled shutter, which enables it to measure Total Lung Volume and Residual Volume of lung, in addition to standard PFT measures. Part of image is reproduced from [103].*

1.1 Body Plethysmography

Plethysmography is a technique to measure the changes in the volume of an organ, limb or body. The word "plethysmography" has been derived from Greek words "plēthysmos" and "graphē" which resepctively mean *increasing* and *writing*. Plethysmography was originally introduced by Bert [34] and Pflüger [146] in 1878 and 1882, which was then developed to volume-constant box whole–body plethysmography by DuBois et al. [62, 63] in 1956. Since then, this technique has been gradually improved to reach its current sophisticated level which incorporates the state-of-the-art technology for PFT.

As seen in Figure 1.1, in whole–body plethysmography patients sit in a cabin and breathe into a pneumotach with a mechanically controlled shutter. As a result of breathing inside an air-sealed cabin with known initial volume and pressure, the unknown volume of lungs is measured using the Boyle-Mariotte law [109]. Thus, in addition to regular PFT measures, plethysmography is able to measure Total Lung Capacity (TLC), Residual Volume (RV) of lungs, airway resistance (R_{aw}), and lung's Functional Residual Capacity (FRC) which is defined as lung's volume at the end of normal expiration [56, 103]. However, whole–body plethysmography is highly expensive and requires bulky equipment, e.g., the cabin, which occupies considerable space. These drawbacks reduce its accessibility and usability.

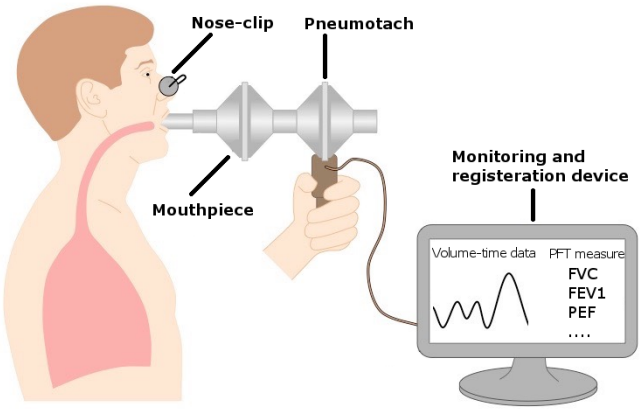


Figure 1.2: *Spirometry is a contact-based clinically approved method to assess lung function by measuring standard PFT measures. Patients breathe into a pneumotach through a mouthpiece while using a nose-clip to prevent air leakage. Image is reproduced from [2].*

1.2 Spirometry

Spirometry is a technique to measure diagnostic information of lungs [129]. The word "spirometry" has been derived from Latin words "spiro" and "meter" which respectively mean *to breathe* and *to measure*. Spirometry was originally invented by the English surgeon John Hutchinson [36] in 1840's in its basic water-based form [184]. He introduced "vital capacity" (VC) as an important lung function measure obtained by his spirometer [185].

In its current form, spirometry is performed by breathing into a contact-based pneumotach through a mouthpiece, in an upright sitting posture while a nose-clip is applied to prevent air leakage [129], as shown in Figure 1.2. Spirometry is the most prevalent PFT method which is broadly used in clinical environments for assessing lung function due to its portability, relative affordability, and accuracy. It is able to provide standard PFT measures, used in diagnosis and assessment of *restrictive* and *obstructive* lung diseases [129].

Forced vital capacity (FVC) and slow vital capacity (SVC) tests are two primary clinical protocols undertaken with a spirometer which vary by the pattern of breathing into the spirometer. Both tests start with a few cycles of normal breathing, called *tidal volume* (TV), followed by a maximal inhalation–exhalation, called *main effort* (ME). The *main effort* inhalation–exhalation is performed at the same speed of normal breathing in the SVC test, whereas it is performed as fast and forcefully as possible in the FVC test.

Table 1.1: *FVC clinical PFT measures.*

Measure	**Name**	**Description**
FVC (litre)	*Forced Vital Capacity*	Maximum amount of air blown out after a maximal inhalation
FEV1 (litre)	*Forced Expiratory Volume*	Volume of air exhaled during the 1st second of forced exhalation
PEF (litre/s)	*Peak Expiratory Flow*	Maximum speed of exhaled air
$FEF_{25\%}$ (litre/s)	*Forced Expiratory Flow 25%*	Flow of exhaled air at 25% of *FVC*
$FEF_{50\%}$ (litre/s)	*Forced Expiratory Flow 50%*	Flow of exhaled air at 50% of *FVC*
$FEF_{75\%}$ (litre/s)	*Forced Expiratory Flow 75%*	Flow of exhaled air at 75% of *FVC*
$FEF_{25-75\%}$ (litre/s)	*Forced Expiratory Flow 25-75%*	Mean forced expiratory flow between 25% and 75% of the *FVC*

FVC and SVC clinical PFT measures are computed using lung volume–time data, i.e., $V(t)$, and flow–time data, i.e., $\dot{V}(t)$, provided by a spirometer. Several FVC measures are obtained within the FVC test among which seven measures, explained in Table 1.1, are more important and frequently used in clinical diagnosis [129, 148]. Four clinical PFT measures, described in Table 1.2, are obtained within the SVC test [129, 148].

Pulmonologists use these PFT measures and their combinations to diagnose *restrictive* lung diseases, e.g., lung fibrosis, and *obstructive* lung diseases, e.g., COPD, Asthma and Bronchiectasis. Global Initiative for Chronic Obstructive Lung Disease (GOLD) [154] introduces *FEV1*, *FVC* and their ratio, i.e., *FEV1/FVC*, as the primary PFT measures used in assessing COPD and its severity. Table 1.3 presents the different levels of obstruction severity based on GOLD classification. In Table 1.3, the predicted *FEV1*

Table 1.2: *SVC clinical PFT measures.*

Measure	**Name**	**Description**
VC (litre)	*Vital Capacity*	Volume change between full inhalation and full exhalation
IC (litre)	*Inspiratory Capacity*	Volume change between a slow, full inhalation and passive end-tidal exhalation
TV (litre)	*Tidal Volume*	Volume of air inhaled and exhaled at rest condition
ERV (litre)	*Expiratory Reserve Volume*	Volume change between passive end-tidal exhalation and complete exhalation

Table 1.3: *GOLD classification of COPD severity [154]. According to GOLD, COPD severity is classified based on the ratio between FEV1 and FVC, i.e., FEV1/FVC < 0.70 and FEV1 measures. The predicted FEV1 is a reference value for a healthy person of the same age, gender, ethnicity and height.*

COPD Class	COPD Severity	*FEV1* (%Predicted)
GOLD 1	Mild	≥ 80
GOLD 2	Moderate	$[50 - 80)$
GOLD 3	Severe	$[30 - 50)$
GOLD 4	Very Severe	< 30

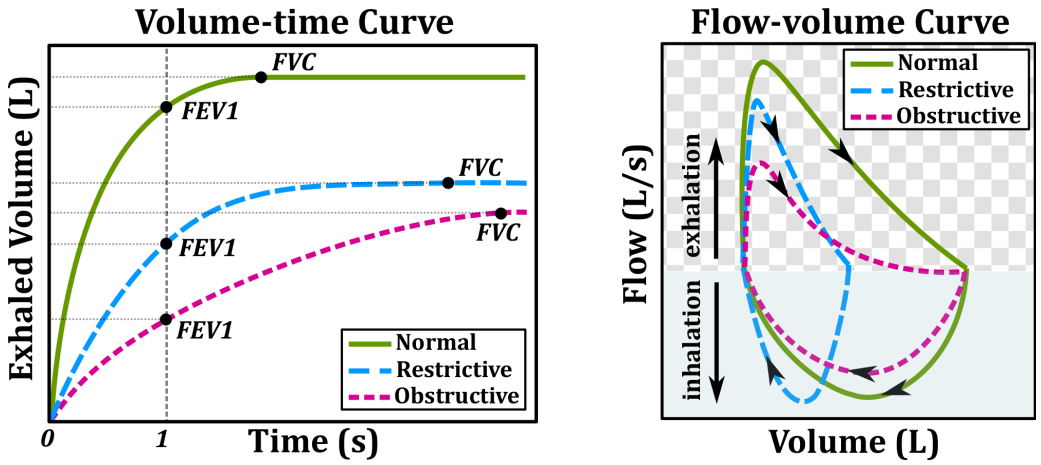


Figure 1.3: *Spirogram patterns of normal, restrictive and obstructive lungs. (a) Comparing the volume–time data, and FVC and FEV1 measures of a forced exhalation. (b) Comparing the flow–volume data of a deep inhalation and a forced and maximal exhalation. Images are reproduced from [72, 99].*

is a reference value for a healthy person of the same age, gender, ethnicity and height.

In addition to PFT measures, pulmonologists also take the advantage of visualised volume–time and flow–volume data (Figure 1.3) for better analysis of lung conditions. For example, Figure 1.3a compares the volume–time spirogram pattern of normal, *restrictive* and *obstructive* lungs and their corresponding *FVC* and *FEV1* measures within a forced exhalation. Similarly, Figure 1.3b presents the flow–volume spirogram pattern of normal, *restrictive* and *obstructive* lungs during a deep inhalation and a forced and maximal exhalation.

1.3 Motivations and Challenges

Spirometry is known as a clinical gold standard approach. However, it cannot be completely error-free as it measures the volume of exchanged air based on the estimation of inhalation and exhalation airflow which can be affected by several factors like environment temperature and humidity. Also, it is highly dependent on patient compliance and their instantaneous effort. Thus, patients are required to repeat PFT several times to achieve at least three consistent-enough results for their measures to be in accordance with ATS/ERS guidelines [128, 129]. Further, spirometry has certain drawbacks which limit its application:

- It is particularly challenging for frail elderly and children, though it is also difficult for other patients to deal with.
- It is a rather expensive approach given the price of a pneumotach and the required disposable accessories, i.e., mouthpiece and nose-clip.
- It is intrusive and unpleasant to wear the nose-clip and mouthpiece, and to be connected to an external device, i.e., pneumotach.
- The spirometer needs to be manually recalibrated on a daily basis to be able to measure accurately.
- It requires the patient's cooperation during the test, hence cognitively impaired patients may find it troublesome to coordinate with.
- Patients with facial muscle weakness and abnormality cannot keep a tight seal around the mouthpiece to prevent air leakage, hence spirometry cannot provide accurate PFT measures for this group of patients [70].
- Cross-infection risk is possible via pneumotach contamination, even though disposable accessories are used [39, 101, 161].
- It requires specialist training.

Remote respiratory sensing has recently become very popular and numerous approaches have been proposed for tidal volume monitoring and respiration rate estimation, without performing PFT [18, 19, 21, 22, 26, 30, 32, 33, 47, 50, 58, 59, 69, 73, 84, 88, 97, 111, 112, 113, 115, 126, 131, 132, 143, 156, 157, 165, 166, 167, 173, 174, 187, 188, 192, 198, 202, 203, 204, 206], based on time-of-flight [143, 157, 192] and structured-light [22, 30, 32, 33, 58, 88, 126, 173, 187, 202, 203, 204, 206] depth sensors, RGB video cameras

[26, 50, 73, 97, 111, 115, 156, 174, 188], motion capture systems [18, 19, 47, 59, 69] and other technologies [21, 84, 112, 113, 131, 132, 165, 166, 167, 198]. These are reviewed in Chapter 2. Among all the existing related studies, we are only aware of the works of Ostadabbas et al. [136, 137] who performed remote *clinical* respiratory assessment. However, they mainly focused on airway resistance estimation using a Kinect and a pulse oximeter [137], and estimated only *FEV1* measure for only 5 healthy subjects in [136].

In this research, we go beyond monitoring of normal respiration and breathing rate estimation, and propose several novel, remote, depth-based pulmonary function testing approaches which estimate 11 standard *clinical* PFT measures using depth measurements obtained from a single or two depth sensors.

Our proposed system has been developed in response to increasing clinical interest in contactless or remote techniques for respiratory assessment. It can be exploited for a wide range of potential applications, such as screening for respiratory diseases, home monitoring, and gating controls for radiological imaging techniques. Our proposed system uses low-cost and easily accessible commodity RGB-D sensors, e.g., Microsoft Kinect V2 [3]. It is easy to setup and does not require calibration on a daily basis. Also, the breathing test performance becomes easier and more pleasant for patients as sealing the nose and using a mouthpiece is no longer required. Due to the remote assessment of the lungs, not only does it cut the costs (pneumotach and disposable accessories), but also it decreases infection risks caused by connecting to a pneumotach. Furthermore, our method requires no specialist training.

In order to remotely acquire breathing data, vision-based respiratory sensing approaches utilise the variation in depth (using RGB–D) or contrast (using RGB) of thoracoabdominal region. Since only chest–surface respiratory motions must be acquired, the subject's trunk movement during the test interferes with the chest–surface motion, and would therefore affect the data acquired by such vision-based remote respiratory sensing approaches [22, 26, 30, 32, 33, 50, 58, 73, 115, 126, 136, 137, 143, 156, 157, 173, 187, 188, 192, 203, 204, 206]. Thus, these approaches need to have the subject's body movement heavily restricted during respiratory monitoring. This would be even more challenging in PFT. As a natural reaction of the human respiratory system, subjects inevitably move their trunk during PFT, especially at the *main effort* deep forced inhalation–exhalation stage. Since constraining such reactive motions would affect the lung function measures, PFT must be performed as in routine spirometry, without restraint. However, when a single depth sensor located in front of the subject is used, decoupling trunk motion and chest-surface respiratory motion would be potentially impossible.

To address the trunk motion problem during PFT performance, an accurate full 3–D dynamic model of the subject's trunk is required. Reconstructing such a model demands: (*i*) accurate alignment of point clouds obtained by distinct depth sensors from different points of view, (*ii*) precise temporal frame synchronization of distinct depth sensors, (*iii*) acquiring body skeletal data to estimate body pose, and (*iv*) *consistent* full frame rate data capturing.

1.4 Contributions

The work presented in this thesis introduces a number of key contributions to the field of remote respiratory assessment:

Remote PFT using a single Kinect: We present a novel remote pulmonary function testing approach in which a single depth sensor is used to compute seven PFT measures of FVC test, i.e., *FVC*, *FEV1*, *PEF*, $FEF_{25\%}$, $FEF_{50\%}$, $FEF_{75\%}$ and $FEF_{25-75\%}$, and four PFT measures of SVC test, i.e., *VC*, *IC*, *TV* and *ERV*. Evaluation on a dataset of 85 patients (529 sequences), attending respiratory outpatient service for spirometry, shows a high correlation between our computed depth-based PFT measures and the measures obtained by a spirometer. Experimental results computed over an unprecedented number of clinical patients confirm that chest–surface motion is linearly related to the changes in the volume of lungs, which establishes the potential toward an accurate, low-cost and remote alternative to cumbersome traditional methods, like spirometry. This work was published in the *Proceedings of the 2015 IEEE Biomedical Circuits and Systems (BioCAS) Conference* [180] and in the *Journal of IEEE Transactions on Biomedical Engineering, Vol. 64, No. 8, pp. 1943-1958, 2017* [182].

3–D data acquisition and registration by two opposing depth sensors: An open source, automatic pipeline is introduced in which two static *opposing* RGB–D sensors are calibrated and used to capture depth, RGB, infrared and body joints data (if applicable) of dynamic objects at consistent and full device frame rate of 30 *fps*. The proposed approach is able to reconstruct rigid and dynamic objects to high accuracy, which we evaluated quantitatively on rigid objects and qualitatively on animated subjects. The main contributions of our work are twofold. First, the deployment of only two RGB–D sensors for 3–D data capture, (*a*) minimises the error of the temporal frame alignment and the spatial point cloud registration, (*b*) reduces the system setup and calibration effort, (*c*) lowers system costs and complexity, and (*d*) minimises the overall operation

space. Second, unlike many other previous approaches, which require a considerable overlap between point clouds for registration [100, 106, 130, 155, 171], our method is able to perform temporal and spatial alignment of two non-overlapping point clouds. This work was published in the *Proceedings of 2016 International Conference on 3D Vision* [181].

Depth-based whole–body photoplethysmography (dPPG) in PFT: Using our data acquisition pipeline, we present a novel dPPG approach to decouple trunk movement and chest–surface respiratory motion which achieves more accurate PFT measures by reducing motion artifacts in respiratory volume–time data. Evaluation on a dataset of 35 subjects (298 sequences), and other comparative analysis between the dPPG and the single Kinect approach, all show the superior accuracy of the dPPG approach. The proposed dPPG method remarkably reduces the L_2 error mean and standard deviation of $FEF_{50\%}$, $FEF_{75\%}$, $FEF_{25-75\%}$, *IC*, and *ERV* measures by half, compared to the single Kinect approach. The most significant novelties of this work are that it introduces the concept of motion decoupling into the remote, vision-based respiratory sensing area and achieves accuracy and efficiency in PFT. Unlike all other previous works (including ours), which restrict the subject's movement during the tests, this method allows subjects to perform PFT as routine spirometry procedure without restricting the subject's natural body reactions at the *inhalation–exhalation* stages. In other words, our broad contribution to the state-of-the-art is to facilitate remote respiratory monitoring and diagnosis without unduly constraining patients. This work was published in the *Journal of IEEE Transactions on Biomedical Engineering, Vol. 65, No. 6, pp. 1421-1431, 2018* [183].

Trunk shape modelling for body motion artifacts correction: We present a vision-based trunk-motion tolerant approach to estimate lung volume–time data remotely within FVC and SVC tests. To effectively filter trunk motion during PFT, we compute respiratory motion pattern by extracting temporal geometrical features from the trunk shape within the whole PFT sequence. This considerably improves the accuracy of the depth-based volume–time data compared to our single Kinect and dPPG approaches. By filtering complex trunk motion — which has different patterns in *tidal volume* and *main effort* breathing — our method is able to calibrate the entire volume–time data using only the *tidal volume* scaling factor, whereas the single Kinect and dPPG approaches require separate *tidal volume* and *main effort* scaling factors for calibration. Our evaluation results on the dual-Kinect PFT dataset shows that our active trunk shape modelling approach reduces the average normalised L_2 error of the depth-based volume–time data to 0.05, from 0.454 and 0.136 obtained by the single Kinect

and dPPG approaches, respectively. This work has been accepted for publication in the *2018 IEEE International Conference on Image Processing (ICIP)*. An extended version of this work has been submitted to the Journal of *Computer Methods in Biomechanics and Biomedical Engineering*, (under review).

The medical significance of our research has been published in the *Proceedings of 2015 British Thoracic Society Winter Meeting* [175] and in the *Journal of Frontiers in Physiology, Vol. 8, No. 65, 2017* [176].

1.5 Thesis Outline

This thesis is presented in 8 main chapters. Following this Introduction, in Chapter 2, we present a brief background on depth acquisition methods which is then followed by a review of related literature and state-of-the-art in remote respiratory motion sensing.

In Chapter 3, we outline the collected datasets used in Chapters 4, 6 and 7, and describe their specifications and details. A dataset of 85 patients (529 sequences), attending respiratory outpatient service for spirometry, was collected and used for evaluation of our single Kinect lung function assessment method (Chapter 4). To evaluate our dPPG PFT approach (Chapter 6) and also our active trunk shape modelling method (Chapter 7), we collected a dataset of 35 healthy subjects (300 sequences) in which each subject underwent routine spirometry tests while their body depth data was captured by *two opposing* Kinects using our proposed 3–D data acquisition pipeline.

Chapter 4 describes our depth-based PFT approach which is able to remotely assess patient's lung function using depth measurements obtained from a single RGB–D sensor. After generating a point cloud from scene depth values, we construct a 3–D model of the subject's chest. Then, by estimating the chest volume variation throughout a sequence, we generate volume–time and flow–time data for FVC and SVC spirometry tests. *Tidal volume* and *main effort* sections of volume–time data are analysed and calibrated separately to deal with the effects of subject's torso motion. After automatic extraction of keypoints from the volume–time and flow–time data, seven FVC and four SVC measures are computed and then validated against measures from a spirometer.

In Chapter 5, we explain our proposed 3–D data acquisition pipeline which exploits two *opposing* non-overlapping RGB–D sensors (Kinect V2) for dynamic object reconstruc-

tion. First, the relative pose of the two RGB–D sensors is estimated through a calibration stage, and rigid transformation parameters are computed. These are then used to align and register point clouds obtained from the sensors at frame-level. We validated the proposed system by performing experiments on known-size box objects with the results demonstrating accurate measurements. We also report on dynamic object reconstruction by way of human subjects performing actions like waving hands.

Our novel, remote dPPG PFT approach is presented in Chapter 6. Following spatial and temporal calibration of two opposing RGB-D sensors, a dynamic 3-D model of the subject performing PFT is reconstructed and used to filter trunk motion by subtracting the average depth of the chest-wall from the average depth of the posterior-wall per frame. The retrieved depth-based volume–time data is then calibrated and used to compute 11 clinical PFT measures for FVC and SVC spirometry tests. Experimental validation results and other comparative evaluations between the dPPG and the single Kinect approach, such as Bland-Altman analysis, similarity measures performance, intra-subject error analysis, and statistical analysis of *tidal volume* and *main effort* scaling factors, all confirm the accomplishments achieved by the dPPG approach.

Chapter 7 presents our proposed active trunk shape modelling approach which corrects trunk motion artifacts within spirometry FVC and SVC tests by decoupling subtle chest-surface respiratory motion and trunk motion during the test. After temporal modelling of the trunk shape, generated using two opposing Kinects in a sequence, the chest-surface respiratory pattern is computed by performing PCA on temporal geometrical features extracted from the chest and posterior shapes. Finally, the retrieved depth-based volume–time data is calibrated using scaling factors learnt in an efficient training phase. We evaluated our approach by computing normalised L_2 error, Fréchet distance, dynamic time warping distance and correlation of determination between the depth-based and spirometer volume–time data.

In Chapter 8, we present a summary of the thesis and its contributions, which is then followed by a discussion. Finally, future directions and potential further extensions to the works presented in this thesis are outlined in this chapter.

1.6 Released Dataset, Source Code and Videos

We have released our dual-Kinect PFT dataset, and it is publicly available for other researchers to use. The entire dataset is available for download from the University of Bristol research data repository at:

- A Dataset for Depth-Based Whole–Body Photoplethysmography in Remote Pulmonary Function Testing: http://doi.org/ckrh.

This dataset is composed of 300 PFT sequences performed by 27 male and 8 female subjects. Each PFT sequence contains the subject's body surface depth information and skeletal data. Also, ground truth spirometry data and the results of our works are provided for each sequence.

The source code for our proposed 3–D data acquisition and registration pipeline is publicly available for download at:

- 3–D Data Acquisition and Registration using Opposing Kinects (MIT License): https://github.com/BristolVisualPFT/3D_Data_Acquisition_Registration_Using_Kinects.

Further, videos which visualise the results achieved in our work are available at:

- Remote, Depth-based Lung Function Assessment: https://www.youtube.com/watch?v=AX4BvyoKYYQ.
- 3–D Data Acquisition and Registration using Opposing Kinects: https://www.youtube.com/watch?v=-JKHG3UJG9Q.
- Depth-based, Whole–Body Photoplethysmography in Remote Pulmonary Function Testing: UnderConstruction.

Chapter 2

Background

This chapter gives the background information for the thesis, in which we present a brief overview of depth sensing methods and review the related works in 3-D scene data acquisition and multi-view registration (Section 2.1). Then, we review the relevant literature on remote respiratory motion measurement in respiration monitoring, breathing rate estimation and clinical respiratory assessment (Section 2.2). Finally, we conclude this chapter by presenting a summary of the related works (Section 2.3).

2.1 3-D Data Acquisition and Registration

While a wealth of information can be extracted from a single image, depth (range) data would achieve more accurate 3-D scene modelling, and it is more suitable for geometrical processing. Depth data provides distance information of the scene from the sensor, and can be acquired using passive, i.e., stereo vision [71, 186], and active, i.e., photometric and structured-light (SL) [16, 162], and Time-of-Flight (ToF) [87, 90], approaches.

2.1.1 Depth Sensing

Binocular stereo vision is a technique to estimate depth information using at least two images captured from different viewpoints. Knowing the camera(s) parameters, the depth of a desired point can be estimated by a simple triangulation if the location of

this point is identified in the images, i.e., stereo matching [89, 92]. Although epipolar geometry [67, 209] and image rectification [75, 117] improve stereo matching speed and accuracy, and reduce its complexity, it remains as the main challenge of stereo vision algorithms [66, 168, 190, 195]. Thus, as a more robust depth sensing method, active stereo vision has emerged in the form of SL [169, 196] and ToF [87, 90] techniques.

Using a geometrical formulation similar to passive stereo vision, SL methods substitute the optical-based correspondence matching with an active pinhole modelling in which a projector/illuminator projects the correspondences as SL patterns for more reliable stereo matching [207]. Since the type of the projected light pattern has a huge impact on the accuracy of the estimated depth map [163], various patterns, e.g., 1-D binary [95, 196] and gray [64, 107] stripes, binary and gray coded grid [83, 114], color coded grid [138, 145, 194] and color coded dots [61, 142], have been explored in the literature. Nevertheless, SL depth sensing has also its limitations. First, it is restricted to in-house use only, due to the destructive effect of direct sunlight on the projected pattern [124]. Second, the effective depth measurement distance is limited to several meters due to the projected pattern visibility contraint [102]. Third, similar to other correspondence matching techniques, SL suffers from *shadowing effect* in which the distance between the camera and the projector affects the object's side views visibility, and no depth data can be acquired for these regions [80]. This is more critical for the objects closer to the sensor. Microsoft Kinect [3, 85] and Asus Xtion Pro [1] are two well-known consumer-grade SL depth sensors with embedded RGB video camera for visualisation. Table 2.1 presents more hardware/software details for these sensors [85, 102, 164, 210]. Interference has been reported for simultaneous use of multiple Kinect or Asus sensors [31, 46].

Unlike passive and SL depth sensing methods, ToF sensors do not employ geometrical triangulation for depth measurements. These sensors estimate the depth based on the radio detection and ranging (RADAR) principle in which the round-trip time of a light pulse is measured to determine the distance between the sensor and the collision point. Depending on the deployed distance measurement technique, ToF sensors can be divided into two broad categories, i.e., *pulsed-light* and *continuous-wave* based technologies. The first category employs high power single-photon avalanche diodes (SPAD) which is used for outdoor long-distance (from a few meter to several Kilometres) measuring sensors like light imaging detection and ranging (LIDAR) cameras. In contrast, in the second category a sinusoidally modulated light signal is emitted using a regular NIR diode, by which the travelling distance is computed by measuring the phase shift between the emitted and the received signals, which limits its application to indoor and short-

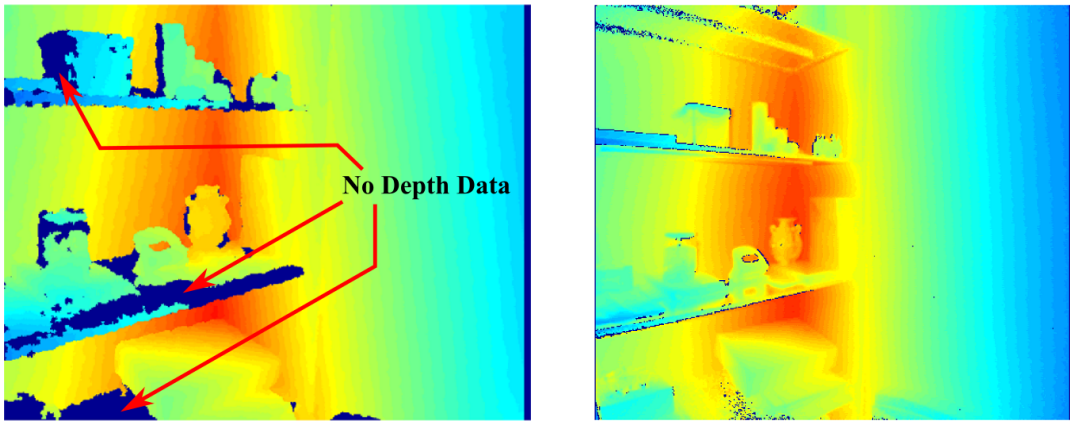
distance (from a few centimetres to several meters) measurements like Microsoft Kinect V2 [90, 207]. Compared to SL Kinect and Asus, Kinect V2 is better in terms of price and technical specifications (see Table 2.1) [90, 108, 139, 164]. While shadowing effects are inevitable in SL depth sensors, they are minimised in ToF sensors as the transmitter and receiver can be located within minimum distance of each other or ideally co-positioned. Figure 2.1 shows samples of depth maps (color-coded) acquired by a SL Kinect and a ToF Kinect V2. As seen, the depth map of the SL Kinect contains numerous areas with no depth data (in dark blue) compared to the ToF Kinect V2 because of shadowing effects and other noises. Also, literature reported no interference for simultaneous use of multiple Kinect V2 [106]. We therefore use this sensor in our remote PFT approaches.

2.1.2 3-D Scene Reconstruction

There are many existing works on the registration of multi-view range images obtained by photometric stereo and SL techniques, such as [16, 65, 159, 162, 200], with some summative works, e.g., in [74, 172, 178]. We limit this review to methods using affordable commodity RGB-D sensors, such as the Kinect, for multi-view 3-D reconstruction and registration using single and multiple RGB-D sensors.

Table 2.1: *Technical details and comparison of Microsoft Kinect, Asus Xtion Pro and Microsoft Kinect V2 [90, 108, 139, 164].*

Specification	**Microsoft Kinect**	**Asus Xtion Pro**	**Microsoft Kinect V2**
Depth Sensing Technology	Structured-light	Structured-light	ToF
Released Date	2010	2012	2014
Depth Resolution (pixels)	640×480 11-bit	640×480 11-bit	512×424 13-bit
RGB Resolution (pixels)	1280×1024	1280×1024	1920×1080
RGB/Depth frame rate	30*fps*	30*fps*	30*fps*
Min−Max depth distance	0.8 − 3.5m	0.8 − 4.0m	0.5 − 4.5m
Horizontal×Vertical FOV	57°×43°	58°×45°	70°×60°
USB Interface	2.0	2.0 / 3.0	3.0
Dimensions (W×D×H)	280×76×76 mm	180×35×50 mm	250×66×67 mm
Weight	550g	120g	970g
Power Consumption (W)	2.5 (Power Supply)	< 2.5 (USB)	15 (Power Supply)
Software Driver	Microsoft SDK 1.8	OpenNI 2.0	Microsoft SDK 2.0
Skeleton Joint Defined	20	20	26
Full Skeletons Tracked	2	2	6
Individual PC per Sensor	No	No	Yes



(a) *SL Kinect* **(b)** *ToF Kinect V2*

Figure 2.1: *Comparing depth maps (color-coded) of (a) SL Kinect, and (b) ToF Kinect V2. SL Kinect has more areas with no depth data than ToF Kinect V2 due to shadowing effect and other noises. Images are taken from [199].*

2.1.2.1 Single RGB-D Sensor 3-D Reconstruction

Approaches which apply a single capturing device, either use a moving sensor on a path around the object or the object rotates for a fixed position sensor. These approaches apply point matching algorithms, mainly Iterative Closest Point (ICP) [35, 208] and other adapted variants [150], to register point clouds by minimising the distance between continuously detected corresponding keypoints in consecutive keyframes. These corresponding keypoints can be determined using uniform sampling of point clouds, general 2-D features, e.g., Scale Invariant Feature Transform (SIFT) [119] and Speeded Up Robust Features (SURF) [27], or depth features specifically designed for 3-D registration, e.g., Fast Point Feature Histograms [160].

Some approaches [94, 96, 134, 135, 212] have recently been proposed for reconstruction of non-rigid objects and scenes using a single RGB-D sensor. Izadi et al. [96] and Newcombe et al. [134] introduced Kinect-Fusion as a real-time 3-D reconstruction approach using a moving Kinect. They presented a new GPU pipeline which allows for real-time camera tracking, surface reconstruction, and rendering. However, these methods expect a static scene during reconstruction. In [212], Zollhöfer et al. first acquired an initial template using Kinect-Fusion of [96], for which the object needed to be static for ∼1 minute. Next, in the non-rigid reconstruction phase, for each frame they roughly aligned the template to the input data and then fitted the non-rigid surface using a new efficient GPU-based Gauss-Newton solver, which minimised the fitting energy function. Newcombe et al. [135] presented a real-time Dynamic-Fusion technique for tracking surfaces

and dynamic reconstruction of non-rigid objects. Each live depth frame was fused into a canonical space using an estimated volumetric warp field, which removed the scene motion, and a truncated signed distance function volume reconstruction was obtained. However, since they omitted the RGB stream and also not utilised global features, their method fails to track surfaces in specific types of topological changes (for example closed to open hands) and it is also prone to drift. In another work, Innmann et al. [94] proposed a similar method to [135], in which they tried to address these issues. In addition to the dense depth correspondences, they applied global sparse color-based SIFT feature correspondences which allows them to better deal with drifts and improve tracking.

Although these single RGB-D sensor approaches yield highly impressive results, they are not able to capture changes that simultaneously happen in those parts of the object that are not within the field of view. Further, they require a substantial overlap in the depth data of consecutive frames, which enables the point matching algorithm, e.g., ICP, to better estimate the point cloud registration parameters. Finally, they can be restricted in speed and size of deformations, e.g., fast and large deformations.

2.1.2.2 Multiple RGB-D Sensor 3-D Reconstruction

In these approaches, multiple static RGB-D sensors are used to simultaneously capture the scene from different points of view. To be able to find the sensors' relative pose, they need to be calibrated individually and together using optical and/or geometric techniques. In the former, the point clouds from each RGB-D camera are aligned using the correspondences established by calibration objects or visual makers, e.g., in [25, 28, 60, 100, 106]. In the latter, rigid transformations that help align the point clouds are computed using geometric-based optimisation methods like ICP and bundle adjustment [193], e.g., in [155, 171]. The geometric methods can also be used as a refinement step for precise alignment of roughly aligned point clouds, e.g., in [100, 106, 130].

Avetisyan et al. [25] employed an optical tracking system to increase depth measurement accuracy of three inward and circularly-located RGB-D sensors. A tracked chessboard was moved inside the capture space by which an individual lookup table was created for each sensor. This lookup table consisted of the chessboard crossing points locations in the tracker coordinate system and their corresponding locations in the sensor coordinate system. The lookup table was then used to correct the sensors' depth measurement accuracy during scene reconstruction. The depth sensors and the optical tracking system

were calibrated using a single rigid transformation.

Beck and Froehlich [28] proposed a volumetric method to calibrate multiple RGB-D sensors by transforming each depth sensor space into a normalized volume space, performing a reference sampling and interpolation. A chessboard was placed in various locations of the capturing volume and captured by the sensors while simultaneously being tracked by a motion capture (MoCap) system. Real-world location of the chessboard crossing points in both RGB-D sensors and motion capture system were used to fill a 3-D lookup table. After performing an interpolation to fill empty cells, the table was used in the reconstruction stage. Approximately 2000 reference samples are required for calibrating a volume of $1.5 \times 1.8 \times 1.5\,\mathrm{m}^3$, which takes $20-30$ minutes to be performed.

Deng et al. [60] localized rigid transformation parameters to improve registration accuracy of point clouds obtained from two Kinects. A 3-D grid of translation and rotation parameters was generated using the established correspondence points obtained from a moving chessboard, and then interpolated. The authors reported improvements in their point cloud registration accuracy compared to the global rigid transformation approaches. However, their method demands a huge amount of overlap in the two sensors capturing space. Their local registration results in geometrical distortion in the final reconstructed point cloud.

Kowalski et al. [106] presented a 3-D data acquisition system, using four Kinect V2 sensors, in which they manually calibrated their system to align the point clouds in a two-step procedure involving rough estimation and refinement. Their qualitative-only results showed good performance for general static and dynamic object reconstruction. However, their calibration stage is cumbersome requiring self-designed markers, manual labelling of markers' location, and sufficient overlap between the sensors. Figure 2.2 shows the 3-D reconstructed models of a subject playing guitar. Spatial registration errors are observable especially on the face, hands and fingers which have more visual details. Also legs 3-D data has not been captured even by using 4 Kinects.

Both optical and geometric techniques were used in [100] to register point clouds from multiple Kinects. First, the relative pose of the Kinects was approximated using a customised calibration box with 2-D visual markers attached on each side. After the point clouds were roughly aligned, they applied an adapted version of Kinect-Fusion [96] in an extra refinement step to create the final point cloud. Similar to Kowalski et al. [106], relative position of markers had to be computed manually.

Figure 2.2: *3-D reconstructed models of a subject playing guitar using 4 Kinects by Kowalski et al. [106]. Spatial registration errors are observable especially on the face, hands and fingers which have more visual details. Also legs 3-D data has not been captured even by using 4 Kinects. Images are taken from [106].*

Miller et al. [130] proposed an unsupervised method to estimate the rigid transform parameters of two overlapping RGB-D sensors, without any initial calibration. First, a moving foreground object was detected in the scene captured by both sensors and the point clouds were roughly aligned by using the centroid of moving foreground objects. In a refinement stage, they tuned the estimation by optimising an energy function which used a nearest-neighbour penalty across all frames. However, this penalty had negative effects where there was not sufficient overlap between the point clouds.

Using a geometric registration approach, Rafighi et al. [155] and Seifi et al. [171] exploited image-based features to align overlapping point clouds obtained from two RGB-D sensors. Matching keypoints were detected from corresponding RGB images of both sensors using SURF and ORB [158] feature descriptors. They were then refined to reject the incorrectly matched keypoints. Finally, after identifying the corresponding location of matching keypoints in depth space, rigid transformation parameters were estimated. However, like all geometric approaches, such as [100, 130], there is a dependency on the availability of good features and a considerable amount of overlap in the sensors' capturing space.

In our proposed optical data acquisition and registration approach in Chapter 5, we perform a 1-step, fast and accurate calibration of two opposing Kinect V2 sensors by using three double-sided chessboards at different depths in the scene and only a pair of infrared/depth images taken each sensor. Our method does not involve any refinement or interpolation steps.

2.2 Remote Respiratory Measurement

In this section, we review the related literature in remote sensing for respiration monitoring, breathing rate estimation, and clinical respiratory assessment. Remote respiratory sensing has recently become an active research field and attracted many researchers, particularly since the availability of commodity RGB-D sensors, e.g., Microsoft SL and ToF Kinects and Asus Xtion Pro, and also high-resolution RGB video cameras. While there is a considerable number of works which have investigated noncontact respiration monitoring and breathing rate estimation, e.g., [22, 26, 30, 32, 33, 50, 58, 73, 88, 97, 110, 111, 115, 126, 149, 156, 157, 173, 174, 187, 187, 188, 192, 202, 203, 204, 206], we are aware of only two works [136, 137] that have performed remote clinical respiratory assessment. Categorising the related works to "respiration monitoring and breathing rate estimation" (Section 2.2.1) and "clinical respiratory assessment" (Section 2.2.2), we review each group based on the applied data acquisition sensor, i.e., RGB video camera, ToF and SL depth sensors, and other types of optical/non-optical sensors.

2.2.1 Respiration Monitoring and Breathing Rate Estimation

2.2.1.1 RGB Video Camera

Many researchers have employed RGB video cameras for respiratory monitoring and breathing rate estimation due to its easy accessibility, low price and minimum hardware requirement. We review these works by categorising them based on their approach in acquiring respiratory data, i.e., *illumination-based* [26, 156] and *motion-based* [50, 73, 97, 110, 111, 115, 174, 188] approaches.

Illumination-based Respiratory Sensing – To remotely acquire breathing data, Bartula et al. [26] generating a 1-D row-wise profile for each captured frame, by which the breathing-related motions were temporally detected by computing the lag of the maximum correlation between consecutive frames' profile. The respiratory signal was then obtained by numerically integrating the estimated displacements. A two-step post-processing stage was developed to improve the signal reliability. In the first step, each frame was divided into "moving" or "stationary" smaller blocks using an adaptive thresholding approach, from which a motion signal was computed as the ratio of "moving" blocks to the total number of blocks, and used to eliminate the signal non-respiratory

intervals. In the second step, a breath-to-breath classifier was used to extract only the reliable breathing cycles. Extracting a feature vector from each breath cycle, a pre-trained decision tree was exploited to classify the valid breath cycles. They reported an initial 85% accuracy for 5 healthy subjects, which was improved to 89% and 95% by performing the first and the second post-processing steps, respectively.

Using RGB data from a smartphone camera, Reyes et al. [156] estimated the chest-wall movement signal in the phone by averaging the chest-wall pixels intensity of R, G, and B channels. The breathing rate and tidal volume were then estimated off-line on a PC. The breathing rate was computed using the signal energy density, and the tidal volume was estimated from the calibrated breathing signal. The calibration was performed by subject-specific scaling factors, computed in a training phase using the spirometer data. They reported an average RMSE of 0.414 ± 0.178 bpm for the estimated breathing rate and 0.182 ± 0.107 litres for the peak-to-peak tidal volume, for 15 subjects. Each subject underwent a 2-minute spontaneous breathing half of which was used for the training.

Motion-based Respiratory Sensing – To conquer the illumination variation drawbacks noted in [26, 156], motion estimation methods like optical flow have been employed. Frigola et al. [73] quantified the breathing-related body movements by measuring the variation of the image gradient vector $\overrightarrow{G_t}$ in every pixel as $D_t(x,y) = |\overrightarrow{G_t}(x,y) - \overrightarrow{G_{t-\tau}}(x,y)|$. Initialising τ as the half of the respiratory rate, and then adjusting it by an iterative Kalman filtering approach, they reduced the effects of unrelated body movement. The breathing signal was computed as the number of pixels contributed in D_t in each frame. After smoothing the signal, the breathing rate was extracted by performing a peak analysis within every 30-second window. Not quantitatively evaluated, authors visually compared their breathing rates to a inductive plethysmograph. Using the same approach, Tan et al. [188] reduced the sensitivity to the unrelated body movements by quantising the movement using the grayscale images themselves rather than their gradient vectors and empirically choosing $\tau = 0.5$. The breathing rate was computed from the derivative of the smoothed respiratory signal. To better track the respiratory motion, subjects were asked to wear a tight and stripy t-shirt which increased the number of detectable pixels on the thoracoabdominal region. The assessment was performed qualitatively by comparing the estimated breathing rate to the visually observed ones.

Taking the advantages of [73, 188] methods, Chatterjee et al. [50] estimated the respiratory rate by applying the optical flow along the image gradient defined as flow field $F_t(x,y)$. Since $F_t(x,y)$ was very small and noisy for textureless and flat regions (Figures 2.3a and 2.3b), it was projected along the direction $\mathbf{V}(x,y)$ to increase its effective-

(a) (b) (c)

Figure 2.3: *(a) Inhalation and (b) exhalation visualisation of flow field. (c) Visualisation of PFF. Images are taken from [50].*

ness. The direction for which the projection was maximised for the first 250 frames of a sequence was introduced as the principal flow field *PFF* (Figure 2.3c). The respiratory signal was then generated for the remaining frames using the computed *PFF*.

Instead of considering the whole image for measuring the respiratory motion as in [50, 73, 188], Shao et al. [174] and Lin et al. [115] localised the respiratory ROIs to speed up their motion-based respiratory detection approaches. Applying the shoulders' edge information, Shao et al. [174] generated the respiratory signal by measuring the shoulders breathing-related vertical movements in two tiny ROIs (40×40 pixel) manually selected from top of the shoulders. The validation was performed on 10 healthy subjects for which 0.93 correlation was reported between the estimated breathing rate measurements and the ground truth readings. In contrast to Shao et al. [174]'s manual ROI selection, Lin et al. [115] applied AdaBoost algorithm refined by a skin-color filtering to automatically select the subject's upper-body from which the respiratory ROI was detected using Haar-like features. The respiratory signal was computed by using a basic optical flow approach [91] from which the breathing frequency was extracted a zero-crossing method. Evaluating on 296 sequences from 8 males and 3 females in various scenarios, the correlation range $0.82 - 0.98$ was reported against a respiratory belt.

In another motion-based approach, Li et al. [110, 111] estimated the respiratory rate of sleeping subjects inside a dark room using an infrared camera. Dividing each frame of the breathing sequence into a grid of cells, 7 keypoints were detected in each cell by Harris corner detector, and then tracked over the whole sequence using the Lucas-Kanade optical flow method [120]. After computing each keypoint trajectory over a 30-second sliding window, non-useful trajectories were eliminated using thresholding approach. After applying PCA on all trajectories of every sliding window, among 5 main principal components the component with the highest periodical resemblance to

the previous sliding window, was used for extracting the breathing rate. A dataset of 75 breathing sequences were collected from 10 females and 7 males in five different sleeping postures, i.e., supine, left-side, right-side, prone, and supine with the torso covered by a blanket. An average RMSE error of 1.12 ± 1.38 bpm was reported against readings from a respiratory belt for the whole dataset.

In a similar approach to Li et al. [110, 111]'s work, Janssen et al. [97] proposed a method to detect breathing-related motions in a sequence. By applying a dense optical flow [44] on each captured frame, a motion matrix was generated and factorised into individual motion trajectories among which the most temporally correlated trajectories were chosen. Then, the trajectory matrix was partitioned into separate spatial regions, and each partition was scored based on its contribution to respiratory motion. Finally, by applying an adaptive thresholding approach on the histogram of partitions score, the respiratory regions were specified and used for generating the respiratory signal. The evaluation was performed on 148 breathing sequences from 4 adults and 2 neonates, with an error of 1.4 bpm for adults and 92.55% accuracy for neonates reported.

There are other RGB camera based works of note [51, 52, 121, 149, 152, 189, 201, 211] with similar approaches and methodologies. We have summarised the methods and materials of these works in Section 2.3.

2.2.1.2 SL and ToF Depth Sensor Approaches

Taking the advantage of depth sensors range measurement, depth-based respiratory sensing approaches generally compute the respiratory time series by measuring the distance of the respiratory-related ROIs (mostly on thoracoabdominal area) from the sensor or chest-wall volume estimation, for *tidal volume estimation* [88, 173, 192], *respiratory motion modelling* [187], *respiration monitoring* [22, 33, 58, 143, 202, 203, 204, 206], or *breathing rate measurement* [30, 32, 126, 157]. Thus, any non-respiratory movement, e.g., user's body motion, interferes with the respiratory motion and affects most of these approaches and their results.

Tidal Volume Estimation – Applying multiple depth sensors, Harte et al. [88] presented a chest-wall motion analysis approach using depth measurements acquired by four SL Kinects. After temporal synchronisation and spatial calibration of Kinects, respiratory time series of a sequence were obtained by estimating the volume of the reconstructed torso in each frame. They used off-the-shelf commercial 3-D software, i.e., Geomagic

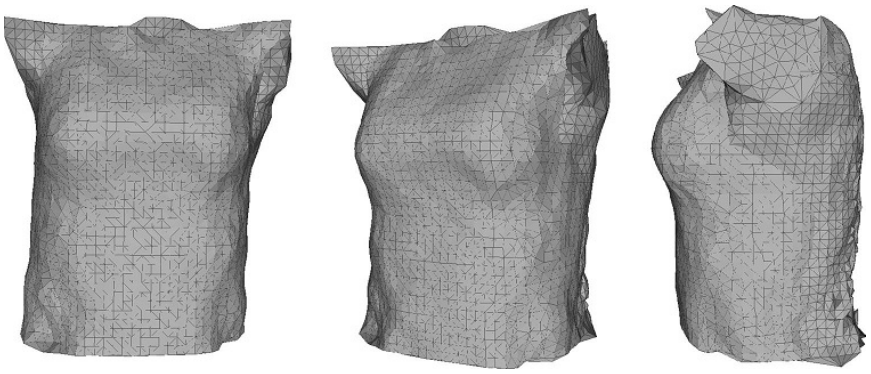


Figure 2.4: *Torso reconstruction using Harte et al. [88] multiple-Kinects approach. The participant has kept up arms away from the torso for Kinects to capture the lateral sides. Images are taken from [88].*

Studio 2012 and Mesh Doctor plug-in [6], for (a) computing rigid transformation and spatial calibration of Kinects, (b) isolating the torso point cloud and filtering its noise per subject, and (c) hole filling and 3-D meshing of the torso surface. To be able to construct a complete 3-D model of the torso during the test, subjects were asked to keep their arms away from their torso (see Figure 2.4) such that Kinects could capture the torso lateral sides (under arms), which is not compatible with spirometry ATS/ERS guidelines [128]. Due to their inability in precise inter-Kinect synchronisation, they reported errors in the dynamic trunk reconstruction. According to the authors, these errors resulted in obtaining breathing time series with inconsistent sampling rates and sudden sharp changes in the estimated volume. Their measurements on a static torso-type resuscitation mannequin with ground truth volume 22.751 litres, showed 0.1 litres error. Also, the minimum and maximum correlation for 40-second tidal breathing of 13 healthy subjects and 9 patients against a spirometer was reported as 0.86 and 0.92. However, no explanation was provided about the torso volume calibration (to present the exchanged volume of air) for validating against the spirometer.

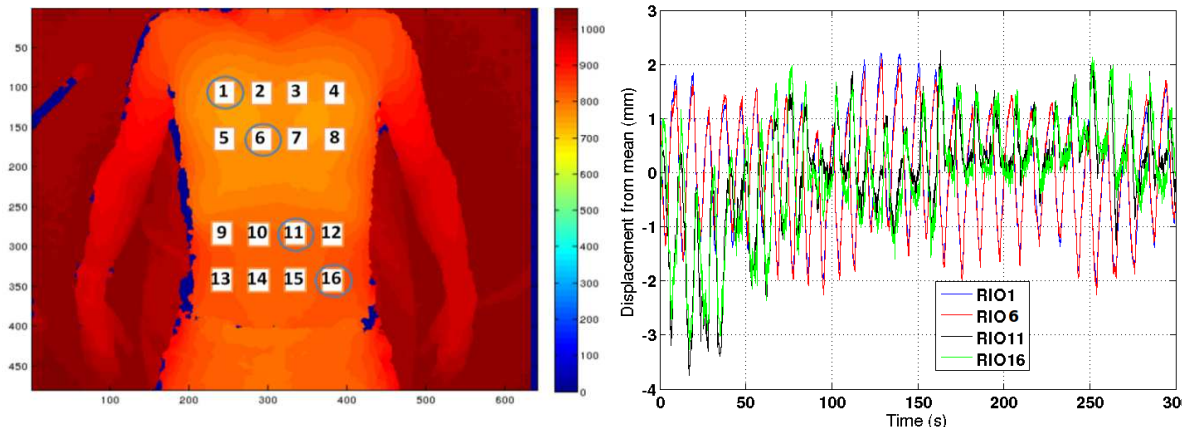
Transue et al. [192] proposed a real-time iso-surface reconstruction algorithm for tidal volume estimation by generating an omni-direction temporal model of the chest-wall. Acquiring the depth data using a ToF Kinect V2, a water-tight 3-D model of the subject's chest-wall was generated and used to estimate the respiratory signal. To present the real exchanged air volume, this signal was calibrated by a non-linear correlation function estimated by a Bayesian neural network [122]. Evaluating on 4 healthy subjects, each subject performed 20 trials of 20-second spontaneous breathing in standing posture. Similar to [88], they constrained participants to keep up their arms away from the trunk to avoid occluding the chest-wall lateral sides, as seen in Figure 2.5. Part of the collected data was used for training their Bayesian network for each participant and the remaining

Figure 2.5: *Subject's standing posture during respiration monitoring [192]. Participants are constrained to keep up their arms away from the trunk to avoid occluding the chest-wall lateral sides. Images are taken from [10, 192].*

was used for testing. Validating against a spirometer, a range of 92.2% to 94.19% accuracy was reported for the estimated tidal volume across their dataset. Moreover, the tidal volume mean error within a 0.2-second window was reported as 0.079, 0.075, 0.067 and 0.055 litres for four subjects, respectively.

Proposing a multi-input–single-output approach, Seppanen et al. [173] used optimally trained filter banks to compute 5 different models by which the airflow data was generated from the chest and abdomen respiratory motion. Evaluation was performed on 8 healthy subjects in 5-minute breathing experiments, composed of 2 free rate breathing periods at the beginning and the end, and 3 controlled breathing periods with 0.2 Hz, 0.15 Hz and 0.33 Hz breathing frequencies. The training was performed on the spirometer and Kinect first free rate breathing period. An average correlation of 0.930, 0.932, 0.909, 0.928 and 0.881 were reported for each of their 5 models for test data of all subjects.

Respiratory Motion Modelling – Tahavori et al. [187] proposed a markerless approach for the chest-wall respiratory motion modelling and drift detection. Using depth data acquired by a SL Kinect mounted on the top of subject's chest-wall, they analysed the respiratory motion data of 20 healthy subjects in three breathing sessions using single-ROI and Multi-ROI approaches. In the single-ROI approach, applying PCA on the temporal data of the thoracoabdominal area showed that the first principal component describes more than 70% of respiratory motion variance in this region. In the multi-ROI approach, the average depth values of 16 ROIs of size 16×16 pixels (see Figure 2.6a), were computed over the whole sequence. PCA analysis of these values showed that the first principal component describes more than 98% and 90% of respiratory variance in the chest and abdomen regions for two participants, respectively. Figure 2.6b presents the respiratory motions of four ROIs specified on the chest and abdomen of the partic-



(a) (b)

Figure 2.6: *(a) Respiratory motion modelling using* 16 *specified ROIs. (b) Respiratory motions of ROIs 1, 6, 11 and 16. Images are taken from [187].*

ipant shown in Figure 2.6a. Also, the authors visualised respiratory motion patterns of the chest and abdomen regions using Gaussian kernel density estimation to detect any potential temporal drift between the chest and abdomen respiratory components.

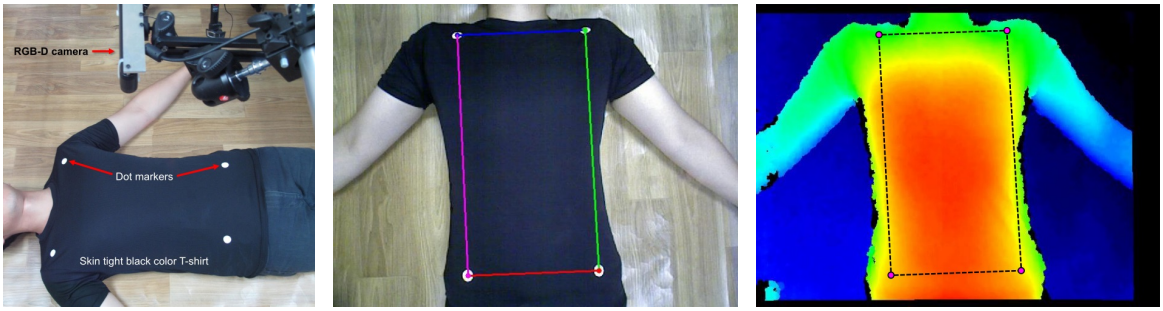
Respiration Monitoring – Using depth measurements from a SL Kinect, Aoki et al. [22] computed respiratory waveform by estimating the volume of thoracoabdominal 3-D model reconstructed in each frame of the sequence. Evaluating on 4 healthy male subjects performing 3-minute spontaneous breathing, an average 0.98 correlation was reported against an expiratory gas analyser. De Boer et al. [58] computed the respiratory signal by estimating the subject's chest-wall volume similar to [22]. To improve the correspondence matching accuracy, they performed an active stereo vision using *two video cameras* and a regular data projector. By projecting a visible chessboard pattern on the subject's body, the crossing points detected by both cameras were used to estimate the scene depth map to generate the chest-wall 3-D model. Undetected correspondence points were corrected by a cubic interpolation based algorithm. An average correlation of 0.91 was reported against the respiration volume data obtained from a spirometer for 40 healthy subjects.

Similarly, by calibrating a near infrared illuminator with a video camera, Bernal et al. [33] developed their self-made SL depth sensor for measuring the chest-wall respiratory motion. After establishing the correspondences using the projected light pattern, the scene depth map was reconstructed by performing triangulation on every pair of correspondences. The respiratory signal was then computed by measuring the chest-wall average depth variation in each frame of the breathing sequence. Without evaluating against

any ground truth, they only reported qualitative results for four different respiratory patterns from an unknown number of subjects performing the test in supine position.

Penne et al. [143] investigated the accuracy of a ToF prototype depth sensor for real-time respiratory monitoring. After manually selecting abdominal and thoracic surfaces, the respiratory waveforms were generated by temporally measuring the Euclidean distance between the fitted plane of these surfaces and the underneath clinical bed surface. The average correlation on 13 subjects was reported as 0.85 and 0.91 for the abdominal and thoracic waveforms against the ANZAI belt, respectively. Not satisfied with the Kinect's depth measurement precision, Xia et al. [204] proposed a motion amplification approach in which they used a flat rigid board on top of the subject's chest (in supine position) to decompose the respiratory motion into the vertical and horizontal displacements. By reducing the angle between the sensor principal axis and the translation surface to 5°, the vertical respiratory motion precision was improved from 1 cm to 1 mm at a 2 m distance. Using this configuration, the respiratory signal was generated from an ROI manually selected from the middle of the board. The correlation range $0.958 - 0.978$ was reported for 4 subjects against an ANZAI belt for a 5-minutes normal breathing.

Wijenayake and Park [202, 203] measured external respiratory motion by applying PCA on the depth measurements acquired by an Asus SL depth sensor. Lying in supine position at 85 cm under the sensor (Figure 2.7a), the thoracoabdominal ROI was selected using the location of 4 markers attached on the subject's t-shirt (Figures 2.7b and 2.7c). After filling the holes of the first 100 frames' depth data (training data) and performing a bilateral filter to remove the depth measurement noise, a respiratory motion model was generated by applying PCA and choosing the first 3 principal components. This model was then used for reconstructing the depth data of the remaining frames. According to



(a) **(b)** **(c)**

Figure 2.7: *(a) System setup in [203]. (b)-(c) The thoracoabdominal detected ROI in RGB and depth images using the markers. Images are taken from [203].*

the authors, reconstructing the test depth frames using the obtained model eliminates the spatial and temporal noise. The respiratory signal was finally generated by estimating the thoracoabdominal 3-D model volume similar to [22, 58]. Evaluated on a dataset of 15 breathing sequences (unknown length) from 10 healthy males, they reported an average of 0.97 correlation against a spirometer for the whole dataset.

In [206], a regional respiratory measurement method was developed by Yu et al. for acquiring the left and right thorax, and abdomen breathing data using a SL Kinect. To be able to measure the absolute size of the user's chest-wall, the relationship between the pixel size and acquisition distance was established in a calibration stage. After performing a viewpoint transformation to correct the Kinect's non-perpendicular viewpoint, the respiratory waveforms were estimated using the computed length per pixel information and the depth of each pixel within the left and right thorax and abdomen ROIs. These ROIs were automatically detected using a predefined chest-wall mask which was adjusted for each subject using their processus xiphoideus. Evaluated on 192 spontaneous breathing sequences from 12 healthy subjects, an average correlation of 0.96 was reported against a spirometer for the estimating respiratory volume.

Breathing Rate Measurement – Benetazzo et al. [30] presented a noncontact breathing rate estimation method and extensively evaluated its performance under various operational and experimental scenarios. Using a SL depth sensor, they automatically detected the chest-wall ROI using the sensor body joint data and generated the respiratory signal using the average depth value of the chest-wall ROI per frame. After smoothing and removing the signal baseline using a weighted averaging filter, the respiratory rate was computed by finding the local optima of the first derivative of the respiratory signal. Validating against spirometer measurements, they reported the correlation range $0.89-0.98$ and the average maximum error 0.53 bpm for 195 spontaneous respiratory tests collected from 5 healthy subjects.

Bernacchia et al. [32] and Rihana et al. [157] applied a similar averaging-based methods to estimate breathing rate from depth data. Bernacchia et al. [32] manually selected three ROIs from the abdomen, thorax and neck regions, from which the respiratory signal was computed by performing Independent Component Analysis (ICA) and wavelet decomposition on the temporal average depth values of these ROIs. They reported an average 0.96 correlation for 40-second spontaneous breathing from 10 healthy subjects against spirometer results. In a similar approach, Rihana et al. [157] investigated the depth measurement accuracy of the ToF Kinect V2 in measuring respiratory rates with varied breathing frequency. The respiratory waveform was generated by measuring the average

depth values of a manually selected ROI on the thoracic surface through the sequence. Fifty 1-minute duration breathing sequences with varied frequencies were collected from 10 healthy subjects. The spectral coherence between the estimated breathing rates and the readings from a respiratory belt showed a range of $0.29 - 0.85$ correlation for 0.3 Hz down to .17 Hz breathing frequencies. According to the authors reported results, the correlation is lower for the higher breathing frequencies.

To improve the breathing rate monitoring accuracy similar to [204], Martinez and Stiefelhagen [126] configured the infrared projector of a SL Kinect to 1280×1024 pixels at 9.1 *fps* and installed a Nikon zoom lens only to the infrared projector by which they increased the size of the projected dots to improve the infrared dots tracking accuracy. Tracking the projected dots within consecutive frames, PCA analysis was performed on the trajectory of all dots over the last 30 seconds. Next, the first 16 principal components were used to estimate the respiratory rate by performing an autoregressive spectral analysis. They evaluated their method on 9 healthy subjects each of which performed 2-minute spontaneous breathing in supine position. An average of 0.99 correlation was reported against an inductance plethysmograph.

2.2.1.3 Other Types of Sensors

Other optical and non-optical sensors have also been investigated for estimating breathing rate and respiratory monitoring and modelling. Prior to public accessibility of depth sensors, many researches applied other types of marker-based [17, 18, 19, 47, 59, 69] and markerless [21, 165, 166, 167, 198] optical sensors for respiratory sensing. Other non-optical but still contactless methods like micro Doppler passive radar [112, 113], Doppler multi-radar [84], Doppler ultrasound [131] and ultrasonic proximity [132] sensors have also been explored for respiratory motion sensing.

Marker-based Optical Methods – Numerous marker-based MoCap methods [17, 47, 59, 69] have investigated the correlation between the chest-wall motion and the exchanged airflow. A maximum of 100 markers were placed on the subject's trunk by which the chest-wall volume was extracted and validated against ground truth measurements.

In a different marker-based MoCap approach, Alnowam et al. [18, 19] used two sets of 8 active markers to acquire the respiratory motion of the chest and abdomen surfaces for the respiratory motion prediction (RMP) [18] and the respiratory motion modelling (RMM) [18, 19]. In [18], after applying PCA for reducing the feature space, a probability

density function (PDF) model was used to predict the chest and abdomen respiratory motions. The PDF model was continuously trained over a 10-second sliding window of the acquired respiratory motion amplitude and velocity, and then used to predict the next respiratory samples amplitude. A joint Gaussian density function was also used to model the respiratory motion using the depth measurements variation. Evaluation was performed on 20 subjects each of which performed 10-minute spontaneous breathing in supine position. An average RMSE of 0.78 mm was reported for the predicted respiratory motion over the whole dataset. In a similar approach, Alnowam et al. [19] used a Gaussian kernel probability density estimation method to analyse inter and intra subject variation of the in-phase and anti-phase respiratory motion patterns.

However, MoCap approaches have major drawbacks that limit their application. First, MoCap systems require a cumbersome calibration of the acquisition space. This calibration should be carried out before each data acquisition session which usually takes around 15 minutes. Second, while markers are unpleasant for the user, placing and removing a sufficient amount of markers on the subject's body takes considerable time and effort. Third, MoCap systems are highly expensive and require a lot of space. Thus, these approaches are almost outdated and have not been used recently.

Markerless Optical Methods – A limited number of works have used fiber grating optical sensor (FGOS) [21, 165] and laser-based instruments [105, 166, 167, 198] for detecting respiration abnormality and breathing monitoring.

Sato and Nakajima [165] proposed a remote breathing abnormality detection method using a fiber grating optical sensor. Projecting the luminescent spots by the fiber grating projector, the location of spots was detected by a 3-D vision sensor and used for estimating the thoracoabdominal peak-to-peak volume change. Validating against a spirometer, 0.96, 0.96, 0.89 and 0.91 correlations were reported for supine, prone, right-side and left-side breathing postures, respectively. Using a similar data acquisition device and methodology, Aoki et al. [21] performed 400 all-night monitoring of 65 elderlies for detecting respiratory abnormalities during sleep.

Scalise et al. [166, 167] exploited an optical laser Doppler vibrometer (LDVi) device for remote respiratory rate measurement. In [167], the distance variation of 10 pre-defined points of interest on the subject's neck, chest and abdomen were measured by emitting laser beams from the LDVi. After discretising the acquired analogue reflection signal, the respiratory movement signal was computed by applying a 12-level wavelet decomposition. An average absolute error of 13 miliseconds was reported for the estimated respiratory

rate of 33 subjects. In [166], they used a similar system for remote monitoring of other vital signs, e.g., heart rate, besides the breathing rate. While Scalise et al. [166, 167]'s approach required the user to be bare-chest with the reflective material applied for the laser diffraction, Wang et al. [198] used a photo-EMF pulsed laser vibrometer (PPLV) without requiring a strong laser diffraction which allowed the estimation of the subject vital signs in their normal clothing.

Non-optical Remote Methods – Among the other remote works of note, Li et al. [112] developed a micro Doppler radar system to measure the chest surface respiratory motion using passive radar signal energy harvesting. A minimum 0.60 and maximum 0.85 correlations was reported for the distances of 100 and 20 cm, respectively. In [113], they extended their work to a two-stage signal processing framework in which they extracted the respiratory rate using an in-home wireless energy transmitter. To improve the acquired respiratory motion accuracy, Gu and Li [84] used a multi-radar system comprising two radar sensors with four antenna patches located at a distance of 50cm from the chest-wall. An average RMSE of 0.88, 2.3, 0.945 and 1.88 mm were reported for the estimated respiratory motion in natural, anger, tenderness and speaking conditions, respectively.

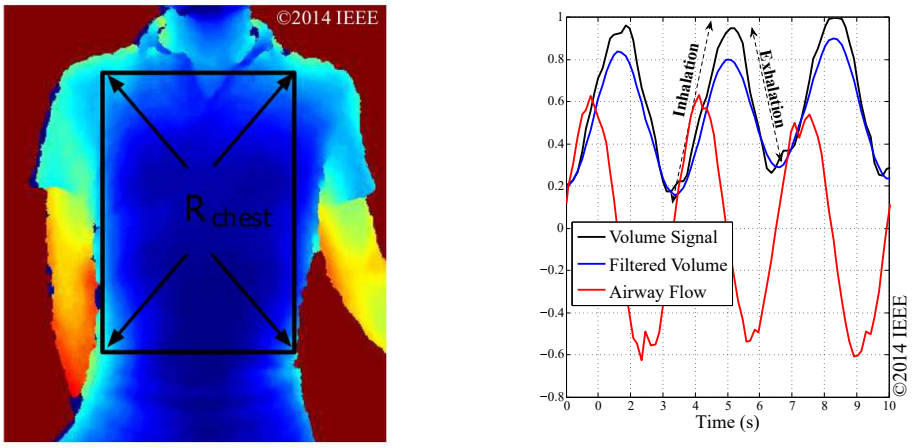
Exploring the ultrasonic sensors accuracy for respiratory motion detection [131, 132], Min et al. [132] estimated the breathing rate from the chest-wall respiratory motion acquired by an ultrasonic proximity sensor after correcting the unwanted high frequency signal drops by an envelop detection method. Evaluated on 10 male subjects, they reported an average of 0.93 and 0.98 correlations for the with and without clothing scenarios.

2.2.2 Clinical Respiratory Assessment

As the most relevant works to ours, we know of only the works [136, 137] of Ostadabbas et al. who also performed remote clinical respiratory assessment but mainly in airway resistance estimation. In [136], they proposed a passive quantitative method to estimate airway resistance using depth data from a SL Kinect. Using the mechanical model of lung, the airway resistance R_{aw} [81] was computed as

$$R_{aw} = \frac{P_{air} - P_{lung}}{Q_f}, \tag{2.1}$$

where P_{lung} and Q_f indicate lung pressure and airflow, respectively, and P_{air} denotes the atmospheric pressure, which is constant at a given altitude. Considering a piecewise



(a) **(b)**

Figure 2.8: *(a) Color-coded depth image of a subject performing a breathing experiment with the chest region ($\mathcal{R}_{chest}$) manually specified. (b) Normalised volume and flow data over a period of time, obtained from the depth data of $\mathcal{R}_{chest}$. Images are taken from [136].*

linear relationship between the pressure of lung (P_{lung}) and its volume (V_{lung}) [38], i.e., $P_{lung} \propto V_{lung} = k_1 \cdot V_{lung}$, they estimated the airway resistance over time t as

$$R_{aw} = \frac{P_{air} - k_1 \cdot V_{lung}(t)}{\frac{\Delta V_{lung}(t)}{\Delta t}}, \tag{2.2}$$

where V_{lung} is the normalised volume of the lung, estimated by an averaging method from the depth information of the chest region $\mathcal{R}_{chest}$ (Figure 2.8a). Using temporal chest volume and flow data obtained from the Kinect (Figure 2.8b), unknown parameters R_{aw} and k_1 (2.2) were computed by solving

$$\begin{bmatrix} \frac{\Delta V_{lung}(t)}{\Delta t} & V_{lung}(t) \end{bmatrix} \begin{bmatrix} R_{aw} \\ k_1 \end{bmatrix} = P_{air} \cdot \mathbf{I}_M, \tag{2.3}$$

using an overdetermined least squares technique, where $\mathbf{I}_M$ denotes the identity matrix with the same size as the temporal input data. By solving (2.3) separately for the inhalation and exhalation periods, two distinct subject-specific k_1 values were obtained for these periods that are supposed not to changed in the future experiments.

They evaluated their method on 5 healthy participants in forced and spontaneous breathing experiments. To induce varied airway resistance in the forced breathing experiments, three custom mouthpieces were made by putting five plastic straws for “Mild” obstruction, three straws for “Moderate” obstruction and one straw for “Severe” obstruction.

Table 2.2: *Average intra-subject k_1 of the forced and spontaneous breathing. Table is reproduced from [136] © 2014 IEEE.*

Calibration Factor	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
Forced k_1	0.85	0.91	0.87	0.81	0.88
Spontaneous k_1	0.94	0.95	0.73	1.03	0.88

Each subject performed 15 trials of forced tests using the three custom mouthpieces (5 times each), and 5 trials without mouthpiece. The exhalation volume and flow data of one trial was used to compute k_1 by solving (2.3). The computed k_1 was then used to extract airway resistance R_{aw} in the other test trials. Their results showed a semi-linear increase in the computed R_{aw} value as the obstruction, emulated by the mouthpieces, had increased. In the spontaneous breathing experiment, a similar procedure was followed and R_{aw} was computed for one minute spontaneous breathing periods, which again showed a semi-linear increase as the obstruction had increased. However, the intra-subject k_1 values computed within the forced breathing experiments were different with the ones computed within the spontaneous breathing experiments. Table 2.2 reports the averaged k_1 for the 5 subjects in the forced and spontaneous breathing experiments.

Using the exhalation volume data, Ostadabbas et al. [136] computed the *FEV1* measure for the forced breathing tests. Table 2.3 reports the average correlation of the computed *FEV1* measures against the measures obtained from a spirometer for each subject. This table also reports the lower and upper bounds of the correlation, i.e., CC_L and CC_U, for each subject. However, no explanation was given about the calibration of *FEV1* measure as this measure was computed from the volume data which is normalised to the chest volume, i.e., data values lie between 0 and 1.

In both experiments, they restricted subjects' trunk movement by asking them to press their back against a wall during the tests. They also asked subjects to manually mark the timestamps for inhalation and exhalation by moving a wireless mouse on the wall during the test.

Table 2.3: *Average intra-subject correlation coefficient (CC) of computed FEV1 measure against spirometry, along with the lower and upper correlation bounds (CC_L/CC_U) for each subject. Table is reproduced from [136] © 2014 IEEE.*

Measure	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
CC	0.95	0.75	0.90	0.86	0.96
CC_L/CC_U	0.88/0.98	0.45/0.89	0.77/0.96	0.64/0.94	0.90/0.98

By expanding their previous study in [136], Ostadabbas et al. proposed a noncontact airway resistance monitoring system [137] in which they used a Microsoft Kinect V2 along with a pulse oximeter (SpO_2) to passively estimate the severity of airway obstruction as mild, moderate or severe. They exploited the same mechanical model of lung in [136] to estimate airway resistance using lung pressure P_{lung} and airflow Q_f, however, they improved their approach in two aspects.

First, while they still estimated the airflow using depth measurements from a depth sensor similar to [136], instead of manually selecting the chest region in [136], they applied a depth-based thresholding approach to automatically detect a subject's chest bounding box. First, a mean image I_D^m is computed by averaging depth sequence I_D, captured from a subject who remained motionless for about one minute. By thresholding I_D^m using T_L and T_U, computed from the histogram of I_D^m, a binary image I_B is computed. This is then followed by a row-wise and a column-wise connected components filtering of I_B, in which M_x and M_y are median lengths of the row-wise and column-wise connected components, respectively. This is finally accomplished by performing an anatomical adjustment to justify the chest bounding box to the subject's torso height (see Figure 2.9 flowchart).

Second, instead of estimating lung pressure as $P_{lung} = k_1 V_{lung}$, in this work they used a photoplethysmograpghy-based surrogate model

$$PP \propto \Delta P_{lung} = k_2 \left(P_{lung}^{(max)} - P_{lung}^{(min)} \right), \tag{2.4}$$

where PP (Pulsus Paradoxus) is the intrathoracic pressure changes in cardiac output and can be estimated from PPG signal acquired by the SpO_2. $P_{lung}^{(min)}$ and $P_{lung}^{(max)}$ are the lung pressure at the end of inhalation and exhalation, respectively. They evaluated their method on two groups of healthy participants and patients in two different phases,

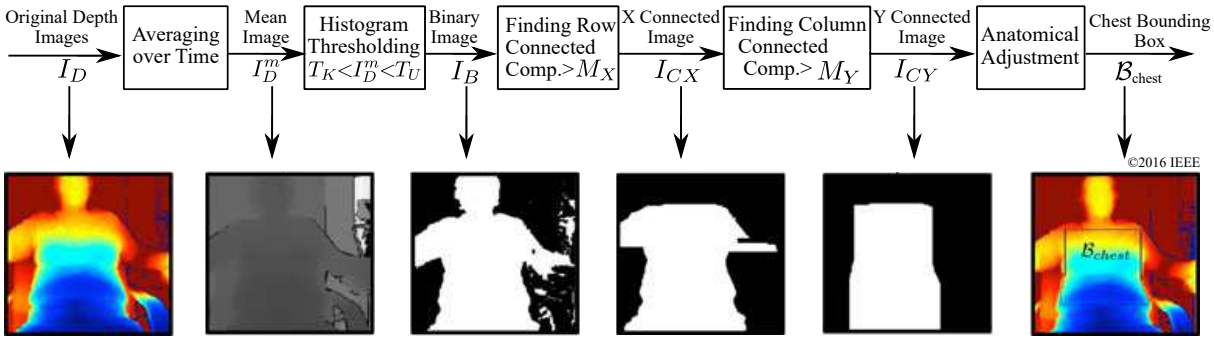


Figure 2.9: *Automatic chest bounding box detection, proposed by Ostadabbas et al. [137]. Image is reproduced from [137].*

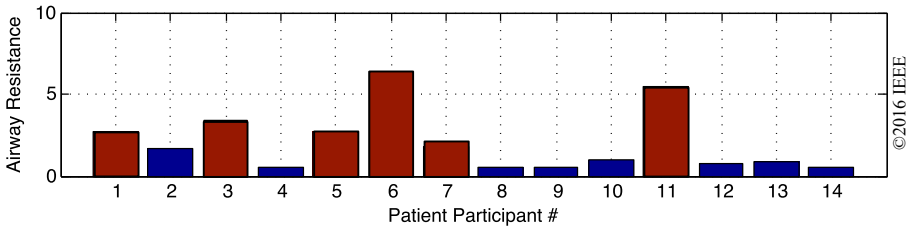


Figure 2.10: *Airway resistance prediction for* 14 *patients in [137]. Red bars specify patients predicted with airway obstruction. Image is taken from [137].*

respectively. In the first phase, they predicted the airway obstruction of 14 healthy subjects who simulated airway obstruction by breathing through the three customised mouthpieces. Each subject performed a spontaneous breathing trial of 1-minute duration by each of the mouthpieces. They reported an average 76.2±13.8% accuracy in predicting airway obstruction as "Mild", "Moderate" and "Severe" in the phase I.

In the second phase, 14 patients participated in an 8-minute spontaneous breathing session in which their overall airway resistance was computed by averaging the airway resistance estimated per breath, over the whole breathing period. Figure 2.10 shows the airway obstruction estimated for these patients. As seen, their proposed method predicted 6 patients to have obstructed airways (red highlighted), whereas only 4 of them were among the 5 patients who were clinically diagnosed with airway obstruction according to the spirometry PFT measures. Thus, 2 patients were incorrectly detected to have airway obstruction and one subject with airway obstruction was not detected. No explanation was given about the threshold used for their classification. They also reported 0.77 correlation and an average error of 0.07±0.06 litres between their computed tidal volume and the plethysmography. However, they did not provide the details of their tidal volume calibration as they computed the tidal volume from the normalised chest volume data.

Since their airflow estimation method in [137] was prone to subject's trunk movement, to minimise such artifacts during the test, healthy participants performed the test in supine position with the Kinect mounted above their chest. However, as this was difficult for patients, they performed the test in sitting position while they were still advised to remain as motionless as possible during the test. Also, a custom-built casing was designed for the pulse oximeter to reduce the finger motion artifacts during PP estimation [137].

2.3 Summary

In this chapter, we presented a background on active depth sensing and multi-view 3-D data acquisition and registration methods. This was then followed by reviewing the state-of-the-art and related literature in remote respiration monitoring, breathing rate estimation and clinical respiratory assessment. Categorising by the acquisition sensor type, a summary of the reviewed works are presented in Tables 2.4 (RGB video camera), 2.5 (SL and ToF depth sensor), and 2.6 (other optical and non-optical sensors).

For these works, type of the computed respiratory measure(s), a summary of their approach as well as the total number of subjects and breathing tests are provided. Comparatively, there are more works on RGB video cameras due to their affordability, availability and the minimum hardware requirements. However, depth sensor based approaches are more robust to illumination variation and the other measurement and environmental noises, and consequently achieve more accurate results.

Among all the existing related studies, we are only aware of Ostadabbas et al.'s works [136, 137] which performed real clinical respiratory assessment, i.e., airway resistance and *FEV1* measure estimation. Inspired by this, we deploy computer vision and signal

Table 2.4: *Summary of related works on respiratory monitoring (RM), and measuring TV and respiratory rate (RR) by RGB video camera.*

Authors	Date	Measure	Approach	Subjects	Tests
Frigola et al. [73]	2002	RM	Image gradient vector subtraction / Kalman filter	–	–
Wiesner & Yaniv [201]	2007	RM	Abdomen Marker-based tracking / PCA analysis	1	5
Tan et al. [188]	2010	RM	Grayscale image subtraction	2	–
Poh et al. [149]	2011	RR	R,G,B illumination processing / ICA analysis	12	12
Bartula et al. [26]	2013	RM	image projection / cross-correlation / breath classification	5	5
Zhao et al. [211]	2013	RR	R,G,B illumination processing / ICA analysis	15	15
Shao et al. [174]	2014	RM	Manual 40×40 pixels ROI / Image gradient analysis	10	10
Li et al. [110]	2014	RR	Feature points / trajectory generation / PCA & Averaging	5	21
Lukac et al. [121]	2014	RM	Optical flow / Spectral analysis	1	1
Tarassenko et al. [189]	2014	RR	Skin detection / Spectral analysis	46	133
Chatterjee et al. [50]	2016	RR	Modified Optical flow	31	186
Lin et al. [115]	2016	RR	Automatic ROI / Vertical optical flow	8	296
Reyes et al. [156]	2016	*TV*, RR	Manual ROI / R,G,B illumination averaging	15	15
Janssen et al. [97]	2016	RM	Chest detection / motion trajectory selection & scoring	6	148
Chatterjee et al. [52]	2016	RR	Optical flow / Fourier analysis	31	186
Chatterjee et al. [51]	2016	RM	Modified Optical flow	33	198
Li et al. [111]	2017	RR	Feature points / trajectory creation / periodicity analysis	17	75
Prathosh et al. [152]	2017	RM	Singal processing / Spectral analysis	31	186

Table 2.5: *Summary of related works on respiratory monitoring (RM), respiratory motion modelling (RMM), and measuring airway resistance (R_{aw}), FEV1, TV, VC and respiratory rate (RR) by SL and ToF depth sensors.*

Authors	Date	Measure	Approach	Subjects	Tests
Penne et al. [143]	2008	RM	Manual chest & abdomen ROI / Depth averaging	13	13
De Boer et al. [58]	2010	RM	Automatic chest-wall ROI / volume estimation	40	–
Xia et al. [204]	2012	RM	Increase depth sensing accuracy by a translation surface	4	4
Aoki et al. [22]	2012	RM	Automatic chest-wall ROI / Volume estimation	4	4
Yu et al. [206]	2012	RM	Thorax mask generation / Volume estimation	12	192
Martinez et al. [126]	2012	RR	Use optical lens to improve depth sensing accuracy	9	9
Ostadabbas et al. [136]	2014	R_{aw},*FEV1*	Manual chest ROI / Depth averaging	5	75
Tahavori et al. [187]	2014	RMM	Manual multi ROI / Depth averaging / PCA	20	60
Benetazzo et al. [30]	2014	RR	Automatic chest-wall ROI / depth averaging	5	195
Bernal et al. [33]	2014	RM	Automatic chest-wall ROI / Volume estimation	2	2
Bernacchia et al. [32]	2014	RR	Manual multi ROI / Depth averaging / ICA	10	10
Seppanen et al. [173]	2015	RM	Automatic ROI / Multi-Input–Single-Output FIR filter	8	8
Wijenayake & Park [202]	2016	RM	Marker-based ROI / PCA modeling / Volume estimation	5	5
Harte et al. [88]	2016	*TV*, *VC*	Trunk reconstruction / Volume estimation	22	66
Ostadabbas et al. [137]	2016	R_{aw}	Automatic chest-wall ROI / Depth averaging / SpO_2	28	56
Transue et al. [192]	2016	*TV*	Automatic chest-wall ROI / Volume estimation	4	80
Rihana et al. [157]	2016	RR	Manual chest ROI / Depth averaging	10	50
Wijenayake & Park [203]	2017	RM	Marker-based ROI / PCA modeling / Volume estimation	10	10

Table 2.6: *Summary of related works on respiratory monitoring (RM), respiratory motion modelling (RMM), respiratory motion prediction (RMP), and measuring TV, VC and respiratory rate (RR) by other sensors.*

Authors	Date	Measure	Approach	Subjects	Tests
Ferrigno et al. [69]	1994	RM	MoCap-based chest-wall volume estimation	12	120
Cala et al. [47]	1996	*TV*	MoCap-based chest-wall volume estimation	2	6
Dellaca et al. [59]	2001	*TV*, *VC*	MoCap-based chest-wall volume estimation	8	–
Aoki et al. [21]	2003	RM	optical fiber grating sensor / Volume estimation	65	400
Sato & Nakajima [165]	2005	RM	optical fiber grating sensor / Volume estimation	1	1
Wang et al. [198]	2007	RR	PPLV / Distance variation measurement	2	2
Min et al. [131]	2007	RM	Doppler ultrasound / Distance variation measurement	1	6
Scalise et al. [166]	2008	RR	LDVi / Distance variation measurement	5	5
Alnowami et al. [18]	2010	RMM,RMP	MoCap / PCA / PDF & KDE estimation	20	20
Alnowami et al. [19]	2010	RMM	MoCap / PCA / PDF estimation	10	30
Scalise et al. [167]	2010	RR	LDVi / Distance variation measurement	33	33
Min et al. [132]	2010	RM	Ultrasonic sensor / Distance variation measurement	10	20
Gu & Li [84]	2015	RM	Doppler multi-radar signal analysis	10	10
Li et al. [112]	2016	RR	Doppler passive radar signal analysis	1	5
Li et al. [113]	2018	RR	Doppler passive radar + in-home wireless signal analysis	3	9

processing algorithms and techniques to estimate primary clinical PFT measures of FVC and SVC breathing tests using various novel approaches.

Acknowledged by all of the existing related works, a subject's body movement during the respiratory monitoring is the main challenge for vision-based respiratory sensing approaches as it interferes with the breathing-related chest-wall motion. Thus, all of these approaches, except for the 4-Kinect based approach by Harte et al. [88] and marker-based approaches [47, 59, 69], restricted their participants trunk motion during the respiration monitoring or just investigated the sensitivity of their approach for limited body motion [30, 115]. Performing PFT using a single depth sensor (Chapter 4), we have also experienced the subject's body movement artifacts. This is even more critical in PFT as subjects would hardly be able to avoid trunk movement during deep forced inhalation–exhalation (ATS/ERS guidelines [128, 129]) since it is an involuntary reaction of the human respiratory system when required to maximally inhale and exhale. Further, restraining this movement can prevent subjects from performing their best breathing effort and would therefore affect their lung function measures. Inspired by this, we present a 3-D data acquisition and registration pipeline using only two opposing Kinects by which we are able to decouple trunk motion and subtle respiratory motion, and significantly improve the accuracy of the estimated PFT measures.

Chapter 3

Data

In this chapter, we introduce the PFT datasets that we collected and used throughout this study. To the best of the author's knowledge, no vision-based PFT dataset deemed fit for evaluating our remote PFT approaches was available at the time of doing this research. Thus, we collected two depth-based PFT datasets using *a single* and *two opposing* depth sensor(s). We collected these datasets based on ATS/ERS spirometry guidelines [128, 129]. To find out the optimal distance of the subject from the depth sensor in PFT data acquisition, we carried out a depth measurement noise analysis which is explained in Section 3.1. In Sections 3.2 and 3.3, we present the details our single-Kinect and dual-Kinect PFT datasets. A summary is presented in Section 3.4.

3.1 Depth Measurement Noise Analysis

Kinect depth estimation suffers from measurement noise caused by the depth sensor technology. At the time when we wished to use the state-of-the-art depth measurement sensor, i.e., Microsoft Kinect V2, there was no public data on its noise characteristics. Thus, we performed a planar depth measurement noise analysis experiment to find the optimal distance range between the sensor and the subject for PFT respiratory data acquisition.

In this experiment, we estimated the sensor measurement error by placing the Kinect at various distances — from 60 cm to 500 cm at 20 cm intervals — in front of a white wall under normal room temperature and lighting conditions, with the sensor's optical axis

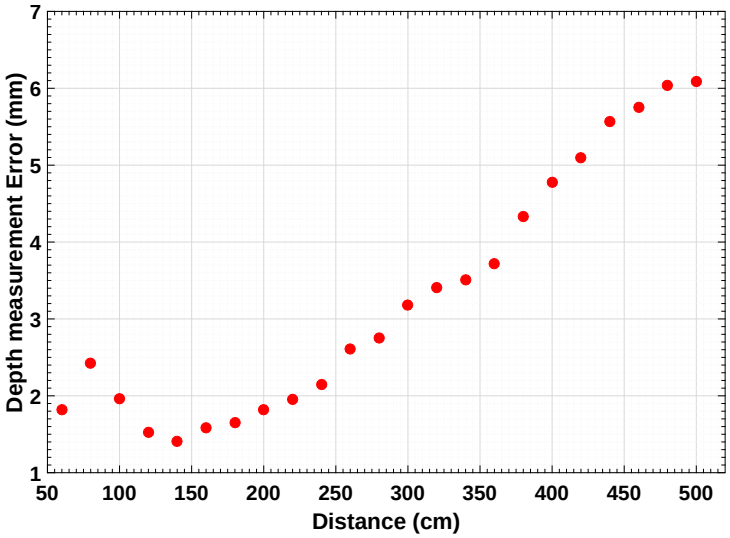


Figure 3.1: *Planar surface noise analysis of depth measurement within a distance range of* $60 - 500\,cm$.

approximately perpendicular to the wall. The sensor's optical axis perpendicularity was checked by measuring the distance of the leftmost and the rightmost edge of the sensor's casing from the wall. At each position, a sequence of 200 frames were recorded and 15 K depth values were randomly sampled from a constant-size patch, and their STD was computed. The patch was selected from the centre of the sensor's viewpoint to minimise the sensor's lens distortion effects, and its size was approximately determined based on the average size of a subject's chest-wall area.

The results of our experiment show that the depth measurement noise increases between ∼60 and ∼80 cm, with the maximum error of 2.4 mm at the distance of ∼80 cm. The error then decreases and drops to its minimum of 1.4 mm at ∼150 cm, and then starts increasing again. Accordingly, we carried out all our experiments with the Kinect placed at ∼150 cm from the subject for which the measurement error is 1.4 mm. Considering the chest-wall respiratory motion range, i.e., ±2.5 cm for the *tidal volume* breathing and ±10 cm for the *main effort* deep *inhalation–exhalation* approximately, the effect of this error on the breathing depth measurements is minimal and limited to jitter noises in the obtained volume-time data which can be easily addressed by applying a smoothing filter as explained in Section 4.1.4. Figure 3.1 illustrates the results of our noise analysis experiment in which the error (in millimetres) has been plotted against the sensor distance to the wall (in centimetres). The presented error analysis shows a non-linear behaviour similar to that of general ToF depth sensors [104, 116]. A similar noise curve was reported by Breuer et al. [41].

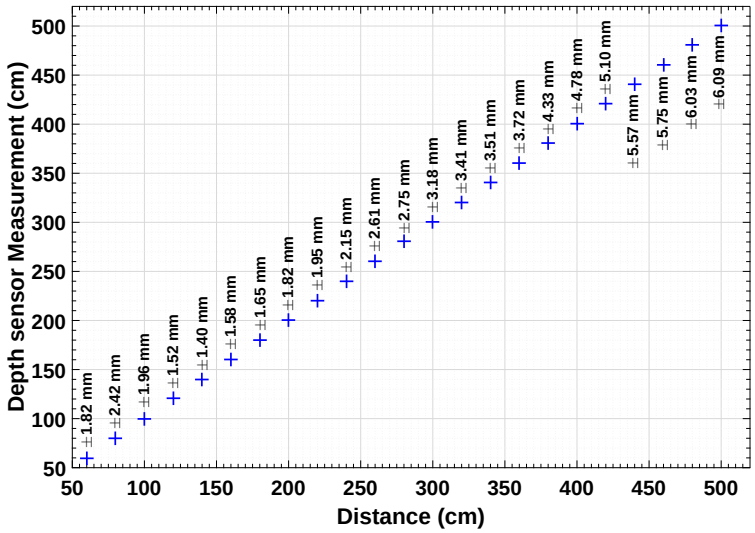


Figure 3.2: *Kinect depth measurement precision validation within a distance range of* $60 - 500\,cm$*. The depth measurement error for each distance is shown in millimetres*

We also evaluated the Kinect depth measurement precision by comparing the real distance of the sensor from the wall (groundtruth) to the average depth value of 15 K randomly sampled pixels. Figure 3.2 shows the comparison of the measured distance against the groundtruth (in centimetres) with the error shown at each distance (in millimetres). However, noise may vary under different environmental lighting and temperature conditions and also depends on the sensor temperature itself. These factors therefore require the optimal distance to be recomputed for the environment the device is to be used in.

3.2 Single-Kinect PFT Dataset

To evaluate our remote PFT approach in Chapter 4, we need a vision-based PFT dataset that contains:

- Participants' chest-wall depth data acquired during FVC / SVC manoeuvre.
- Participants' respiratory volume–time and flow–time data, as well as PFT measures, acquired by a ground truth device for the system training and validation.

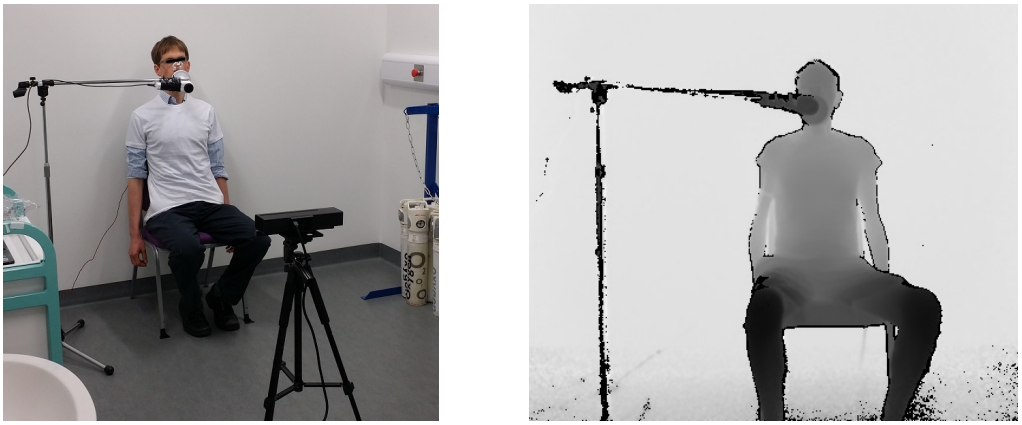
More importantly, FVC and SVC spirometry tests must be perfomed in accordance with ATS/ERS guidelines [128, 129]. According to these guidelines, a minimum of three FVC / SVC manoeuvres with an acceptable repeatability criterion is required for the test

to be considered as successful. For the FVC test, this criterion is to obtain three *FVC* and *FEV1* measures for which their difference with their corresponding measures across the three manoeuvres is < 0.150 litres. Also, the forced exhalation duration must be longer than 6 seconds (3 seconds for subjects aged < 10). Similarly, for the SVC test, the acceptable repeatability criterion is to obtain three *VC* measures for which their difference across the three manoeuvres is < 0.150 litres. The exhalation in the *SVC* manoeuvre is performed in a relaxed manner, and can take a couple of seconds based on the subject's lung capacity. If the acceptable repeatability is not achieved in three manoeuvres, performing additional manoeuvres is required. As per ATS/ERS guidelines [128, 129], the participant must perform the FVC / SVC manoeuvre in an upright sitting posture, using a standard clinically approved spirometer with a nose clip applied to avoid air leakage. The spirometer must be capable of measuring a minimum volume of 8 litres over a minimum 30-second duration. The spirometer must be calibrated before each PFT session using a standard 3-litre syringe. To minimise the infection risk, disposable mouthpiece and nose clip must be used. To avoid air leakage, the mouthpiece must be completely sealed around the lips.

To the best of the author's knowledge, such a PFT dataset was not available at the time of doing this research. Thus, we collected a vision-based PFT dataset based on ATS/ERS guidelines [128, 129] which covers all of the data requirements for evaluating our proposed single-Kinect PFT approach in Chapter 4. Prior to the data collection, an NHS ethical approval was granted from the North West England Research Ethics Committee (reference-15/NW/0040) for recruiting patients at Southmead Hospital Bristol Respiratory Clinic.

As noted previously in Chapter 2, many of the related works minimised the participant's body movement artifacts by performing the data acquisition in the supine position. Although this would restrict the subject's body movement, it is not in accordance with ATS/ERS guidelines [128, 129]. Moreover, it would be difficult for fragile COPD patients to accomplish the deep forced inhalation–exhalation correctly in supine position. Thus, for collecting our dataset, we asked patients to perform FVC / SVC manoeuvres in a sitting posture to be in accordance with the guidelines. Patients sat upright on a chair, facing the Kinect placed at a distance of 1.5 m away from the subject and at a height of 0.6 m, as seen in Figure 3.3a.

The subject's frontal upper-body depth data and pose information were acquired in each PFT performance using our open source single Kinect data acquisition software [9]. Figure 3.3b shows a typical depth image of a subject performing PFT. The subject



(a) **(b)**

Figure 3.3: *(a) Data acquisition setup for collecting the single-Kinect PFT dataset. (b) A typical Kinect depth image of a participant performing PFT.*

was asked to put on a reasonably tight t-shirt to help improve the tracking accuracy of the chest-wall motion. To be able to train and also validate our method, the respiratory volume–time and flow–time data was simultaneously acquired using a clinically approved spirometer, i.e., 'HDpft 1000 High Definition'. This spirometer provides raw volume-time and flow-time data at 200 Hz sampling rate for FVC and 50 Hz for SVC by which the clinical PFT measures are computed using the spirometer software, i.e., "nSight Health". For validation, we compare our results against PFT measures directly taken from the spirometer software.

A total of 590 PFT sequences, i.e., 287 FVC manoeuvres and 303 SVC manoeuvres, were collected from 85 patients attending the respiratory clinic at Southmead Hospital Bristol with a range of lung pathologies as they underwent their routine spirometry test. The data collection spanned several months between March and July of 2015. The 36 male and 49 female patients were of average age of 61.7 ± 14.5 years, height of 166.2 ± 8.2 cm, weight of 77.9 ± 20.5 Kg and BMI of 28.1 ± 6.4 Kg/cm^2. From the total of 590 sequences recorded, 61 cannot be used, among which 15 are due to patient inability to complete the spirometry manoeuvre, 8 are due to hardware failure and 38 are due to significant body movement during the manoeuvre. The evaluation is therefore carried out on the remaining 529 sequences (260 FVC and 269 SVC) from 85 patients. Note that sequences are removed from the dataset only where their volume–time data has been greatly affected by the body movement such that they cannot be processed for automatic computation of the required keypoints, scaling factor computation and generalisation, and PFT measures computation.

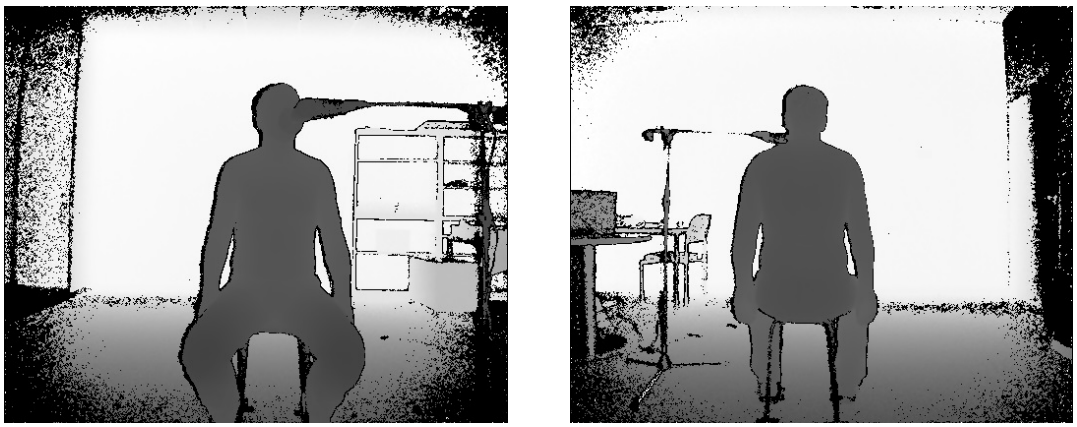
Figure 3.4: *Data acquisition setup for collecting the depth-based PFT dataset using two opposing Kinects.*

Data from this study is unavailable for sharing due to insufficient consent from the study participants.

3.3 Dual-Kinect PFT Dataset

We collected this dual-Kinect dataset for evaluating our remote PFT approaches in Chapters 6 and 7 in which we reduce the subject's body movement artifacts during PFT performance. Employing our open source 3-D data acquisition pipeline [4], we acquire depth and body pose data of the participant's trunk during FVC / SVC manoeuvres using *two opposing* ToF Kinects. Each of the Kinects was placed at a distance of ∼1.5 m away from the subject to minimise the depth measurement noise, and at a height of 0.6 m. Figure 3.4 illustrates the dual-Kinect data collection setup. Each subject was asked to wear a reasonably tight t-shirt and sit up straight on a backless chair. Participants were neither restricted nor advised to be stationary during the PFT manoeuvres, and the tests were performed as routine FVC / SVC spirometry test. The ground truth respiratory data, i.e., volume–time and flow–time data as well as PFT measures, was simultaneously collected using the clinical spirometer device — mounted on a clamp — along with the software that we previously used in our single-Kinect data collection. Figures 3.5a and 3.5b show the depth images of a subject acquired by the front and back Kinects, respectively, during an FVC spirometry manoeuvre.

Thirty five healthy subjects (8 females and 27 males) were recruited for participating



(a) (b)

Figure 3.5: *Sample depth images of a subject performing a PFT spirometry manoeuvre from (a) the front Kinect, and (b) the back Kinect viewpoints.*

in this study. Subjects were of average age of 30.3±5.3 years, height of 175.4±8.0 cm, weight of 73.6±11.1 Kg and BMI of 23.9±3.0 Kg/cm^2. Ethical approval was obtained for this study from the University of Bristol Research Ethics Committee (Reference 56124), and each participant signed a written consent form.

Using the proposed experimental setup, a total of 156 FVC and 150 SVC sequences were collected from the 35 subjects. Similar to the previous dataset, every test was performed based on ATS/ERS guidelines [128, 129]. Thus, each subject underwent several FVC and SVC spirometry tests (at least three) to achieve consistent PFT measures. Most of the participants therefore had to perform extra manoeuvres to ensure consistency.

From the total of 306 PFT manoeuvres, the data for 5 sequences are omitted due to the spirometer (2 sequences) and Kinect (3 sequences) failures, and 1 sequence's data was removed as a subject occluded their chest with their hands during the manoeuvre. Our proposed depth-based whole-body photoplethysmography approach (Chapter 6) fails to process the data from 2 sequences due to the subject's complex body motion patterns in these manoeuvres. Thus, in our proposed method in Chapter 6, the volume–time data of all the other 298 sequences are successfully analysed, and their PFT measures are computed and considered in the experimental analysis. However, our active trunk shape modelling approach (Chapter 7), successfully analyses all 300 sequences' volume–time data and computes their keypoints.

We have released our dual-Kinect PFT dataset and it is publicly available for download at: http://doi.org/ckrh.

Table 3.1: *Summary of our single-Kinect and dual-Kinect PFT datasets.*

Specifications	**Single-Kinect Dataset**	**Dual-Kinect Dataset**
Total Subjects (females, males)	85 (49, 36)	35 (8, 27)
Smokes / Quit / Never	10 / 36 / 39	5 / 6 / 24
Total Acquired Sequences (FVC, SVC)	590 (287, 303)	306 (156, 150)
Age (mean ± STD)	61.7±14.5 years old	30.3±5.3 years old
Height (mean ± STD)	166.2±8.2 cm	175.4±8.0 cm
Weight (mean ± STD)	77.9±20.5 Kg	73.6±11.1 Kg
BMI (mean ± STD)	28.1±6.4 Kg/cm^2	23.9±3.0 Kg/cm^2
Total Used Sequences (FVC, SVC)	529 (260, 269)	300 (155, 145)
Excluded Sequences – Hardware Failure	8	5
Excluded Sequences – Motion Artifacts	38	1 (occluding chest-wall)
Excluded Sequences – Spirometry Failure	15	—

3.4 Summary

In this chapter, we first explained our depth measurement noise analysis study which showed the optimal distance between the Kinect and subject as ∼1.5 m. Then, we presented the specifications of the two datasets collected, i.e., the single-Kinect and the dual-Kinect PFT datasets. Table 3.1 presents a summary of the specifications and details of the two datasets. Thirty eight sequences were excluded from the single-Kinect dataset as our algorithms failed to extract their keypoints automatically for scaling factor computation and generalisation, and PFT measures computation due the destructive effects of the body motion. Also, one sequence was excluded from the dual-Kinect dataset as the subject occluded the chest-wall by their hand during the test.

Chapter 4

Remote PFT Using a Single Kinect

In this chapter, we present our single-Kinect remote PFT approach. After identifying the subject's chest-wall ROI automatically, and reconstructing its dynamic 3-D model in each frame of a PFT sequence, Kinect volume–time data is computed by measuring the chest-wall temporal volume variation (Section 4.1). Since this data presents the chest-wall volume variation rather than real volume of exchanged air, it must be calibrated by linearly scaling the chest volume to the lung volume using subject-specific (intra-subject) scaling factors. To deal with the subject's trunk movement, the volume–time data is automatically split into *tidal volume* and *main effort* parts using several computed keypoints, and each part is analysed separately (Section 4.2). In a learning phase, we compute *tidal volume* and *main effort* scaling factors from training PFT manoeuvres by performing linear regression analysis on every Kinect/spirometer volume–time data pair. Then, we develop a scaling factor generalisation method to minimise the effect of a subject's trunk movement when calibrating a Kinect test volume–time data (Section 4.3). Finally, using the extracted keypoints, we compute 7 FVC and 4 SVC measures from the Kinect calibrated data (Section 4.4). We evaluate our remote approach on the single-Kinect dataset which contains 529 sequences (260 FVC and 269 SVC) collected from 85 actual patients. The experimental results show a high correlation between our Kinect PFT measures and the ground truth measures provided by the spirometer software. Further, we evaluate the scaling factor learning method, and PFT measures' reproducibility and stability (Section 4.5). We conclude this chapter by presenting a summary of our method and its achievements, and directions for future works (Section 4.6).

Figure 4.1 presents an overview of the our remote PFT approach.

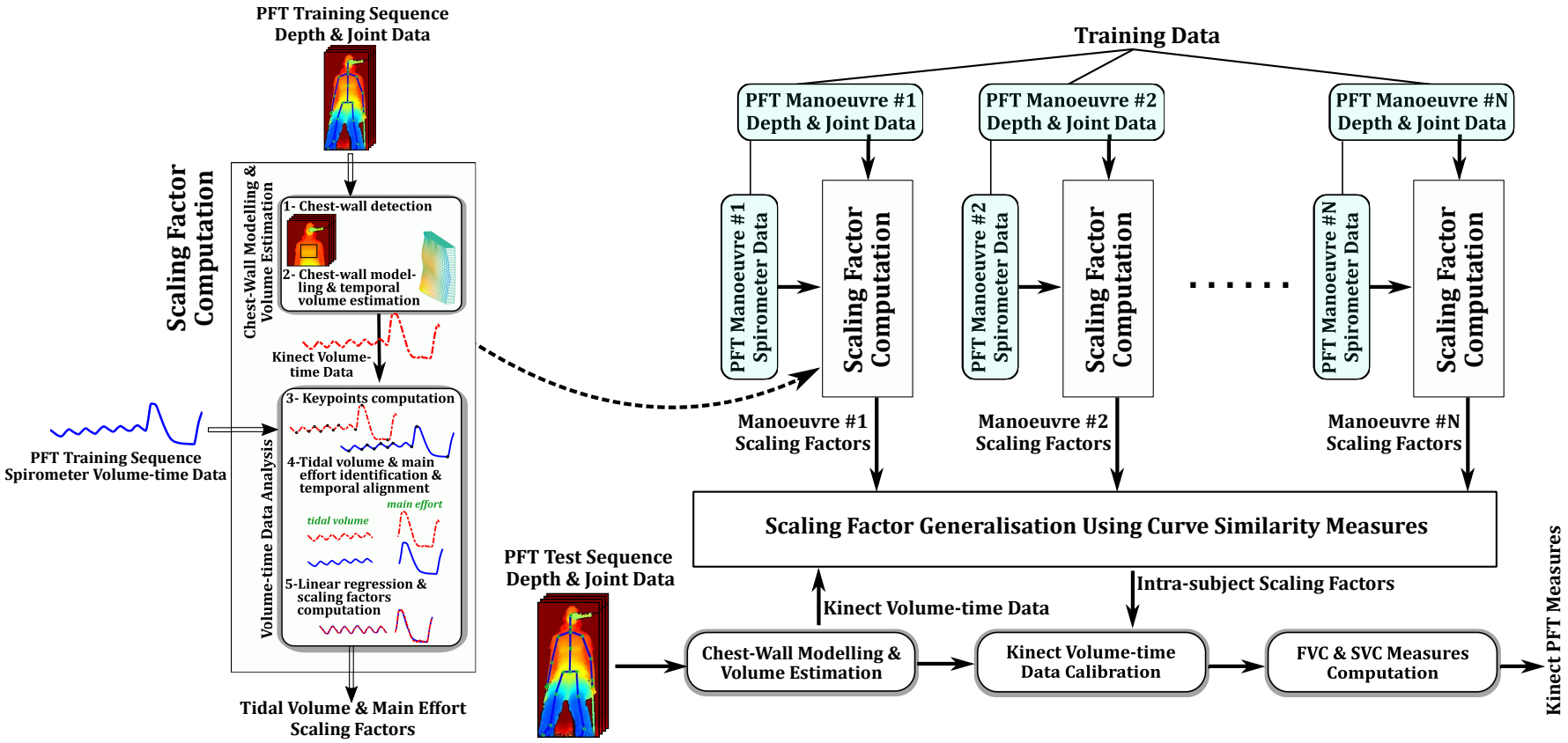


Figure 4.1: *A schematic of the proposed method. The calibration scaling factors are computed for each training pair of Kinect/spirometer volume–time data from which the optimum scaling factors for calibrating the Kinect test PFT sequence are chosen using our scaling factor generalisation method.*

4.1 Chest-wall Modelling and Volume Estimation

4.1.1 Depth Measurement Noise Filtering

To filter the sensor's depth measurement noise previously noted in Section 3.1, we use a bilateral smoothing filter due its ability in eliminating high frequency noises while preserving edges in depth data [140]. We apply

$$BF[I]_p = \frac{1}{W_f} \sum_{q \in S} G_{\sigma_s}\big(\|p-q\|\big) G_{\sigma_r}\big(|I_p - I_q|\big) I_q, \tag{4.1}$$

to every frame of the PFT captured sequence, where W_p is the normalization factor, G_{σ_s} is a spatial Gaussian kernel, G_{σ_r} is a range Gaussian kernel, p and q are the locations of the central and neighbouring pixels, $\|p-q\|$ is the Euclidean distance between pixel locations p and q, and I is the image to be filtered. The range parameter σ_r of the bilateral filter is determined to be 1.5, which is approximately equal to the STD of distance measurements obtained by the Kinect at the chosen distance of $\sim 150cm$. In particular, this value is selected as in [48], according to the level of noise at this distance, to optimize the performance of the range component of the bilateral filter. For the spatial filter, we select $W_f = 13$ which guarantees a good trade-off between precision and processing speed, also reported by Camplani et al. in a similar filtering approach [48]. Consequently, $\sigma_s = W_f/6$, such that the the significant part of the Gaussian kernel (up to $3\sigma_s$) is completely included within the selected window W_f [71].

4.1.2 Chest-wall 3-D Modelling

To segment a subject's chest-wall point cloud from the captured scene in the filtered depth image, a chest-wall ROI mask is automatically computed using Kinect skeleton joints data defined by *ShoulderRight*, *ShoulderLeft*, *SpineShoulder* and *SpineMid* as seen in Figure 4.2a. While the mask width and height could be simply specified as $|ShoulderRight.x - ShoulderLeft.x|$ and $1.5\times|SpineMid.y - SpineShoulder.y|$, respectively, it is not always a reliable approach because the mask might wrongly cover the non-breathing-related regions as shown in Figure 4.2a. This happens due to the inconsistent skeleton joints position across subjects. To address this issue, we have developed Algorithm 4.1 by which the mask is computed by tracking the top of the shoulders and further processing of the joints. Figure 4.2b shows the final obtained mask.

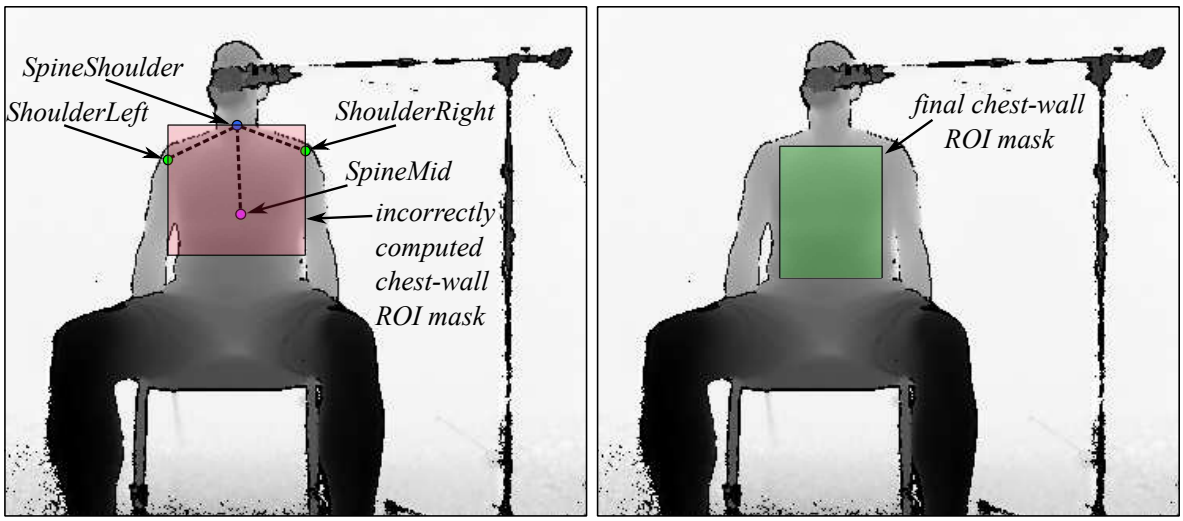


Figure 4.2: *(a) Skeleton joints used in computing the chest-wall ROI mask, i.e., ShoulderRight, ShoulderLeft, SpineShoulder and SpineMid. As seen, using only the original skeleton joints without any further processing, the chest-wall ROI mask is incorrectly computed. (b) The correctly computed chest-wall ROI mask using Algorithm 4.1.*

input: $depthImage, SoulderRight, ShoulderLeft, SpineShoulder, SpineMid$
output: $chestRectROI$

initialise $chestWidth = \frac{2}{3} \times [ShoulderRight.x - ShoulderLeft.x]$;
initialise $chestHeight = \frac{3}{2} \times [SpineShoulder.y - SpineMid.y]$;
initialise $ShRx = SpineShoulder.x - \frac{1}{2}chestWidth$;
initialise $ShLx = SpineShoulder.x + \frac{1}{2}chestWidth$;
initialise $ShRy = ShoulderRight.y$;
initialise $ShLy = ShoulderLeft.y$;
while $\left(|depthImage(ShLx, ShLy) - depthImage(ShLx, ShLy - 1)| < 10\,cm\right)$ **do**
    $ShLy = ShLy - 1$;
while $\left(|depthImage(ShRx, ShRy) - depthImage(ShRx, ShRy - 1)| < 10\,cm\right)$ **do**
    $ShRy = ShRy - 1$;
$chestRectROI.x = ShLx$;
$chestRectROI.y = Max(ShLy, ShRy) + \frac{1}{5} \times [SpineShoulder.y - SpineMid.y] + 1$;
$chestRectROI.width = chestWidth$;
$chestRectROI.height = chestHeight$;

Algorithm 4.1: *Chest-wall ROI mask computation algorithm.*

Since the skeleton joints are unstable and noisy in consecutive frames due to the partial occlusion of the face by the spirometer, the mask size and position would be very noisy when computed using the joints data of each frame. We address this issue by computing the mask from the subject's joint data at the first frame of the sequence.

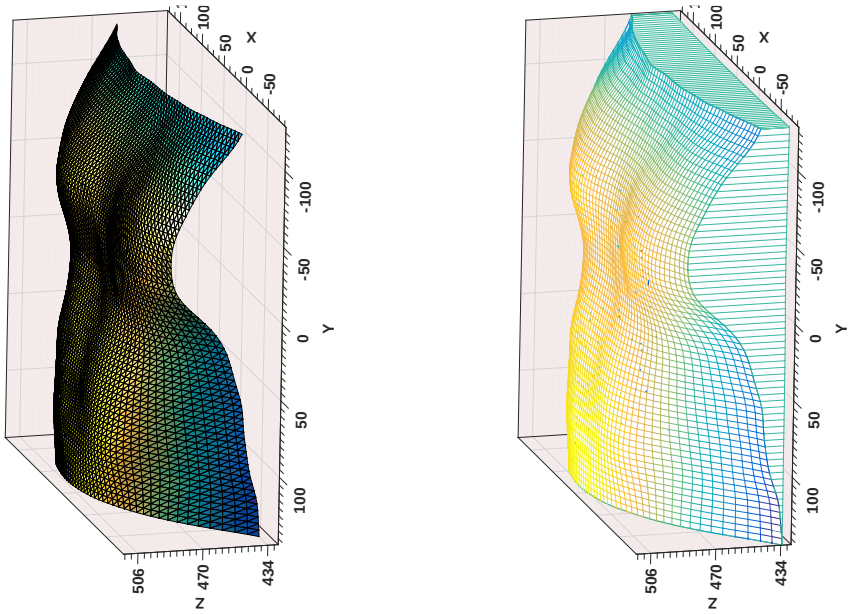


Figure 4.3: *(a) Reconstructed chest-wall 3-D model of a subject wearing a tight t-shirt. (b) Chest-wall surface confined by a reference plane and lateral sides.*

It is worth noting that Microsoft Kinect V2 obtains more consistent skeleton joints, compared to its predecessor, when the subject is seated and partially occluded. The chest-wall surface is then reconstructed by applying a 2-D Delaunay triangulation [57] on the point cloud. Figure 4.3a shows a sample of a reconstructed chest-wall 3-D model.

4.1.3 Chest-wall Volume Estimation

In [136], Ostadabbas et al. estimated the chest-wall volume by obtaining the mean depth value of this region. Given the 2.5-D data, we propose a method to estimate the chest-wall volume by computing the volume between the generated chest-wall 3-D model and a reference plane at a predefined distance from the sensor. Our approach is sufficient to compute the volume–time data $V(t)$ by measuring variations in the estimated volume, based on the assumption that body movements are minimal during PFT and can be ignored. The more accurately the chest-wall respiratory motion is tracked, the more precise the computed volume–time data would be. Ideally, the bare-chest would provide the best results, but as this was not permitted by the granted ethical approval, participants were asked to put on a reasonably tight t-shirt. Enclosing the reconstructed chest-wall surface by surrounding lateral sides and a reference plane (Figure 4.3b), its volume is

estimated using the divergence theorem,

$$\iiint_V (\nabla \cdot \mathbf{F})\, dV = \oiint_S (\mathbf{F} \cdot \mathbf{n})\, dS, \tag{4.2}$$

where V is a compact solid region which is limited by a closed piecewise smooth surface S, $\mathbf{F}$ is a vector field with continuous differentiable components, and $\mathbf{n}$ is an outward pointing normal. Thus, the enclosed volume at time point t, i.e., V_t, is estimated by computing $\oiint_S (\mathbf{F} \cdot \mathbf{n})\, dS$ for each sub-surface of the enclosed chest-wall volume, i.e., *chest surface*, *lateral surface* and *reference plane*, as

$$V_t = \oiint_S (\mathbf{F} \cdot \mathbf{n})\, dS = \oiint_{S_{ch}} (\mathbf{F} \cdot \mathbf{n}_{ch})\, dS_{ch} + \oiint_{S_l} (\mathbf{F} \cdot \mathbf{n}_l)\, dS_l + \oiint_{S_r} (\mathbf{F} \cdot \mathbf{n}_r)\, dS_r. \tag{4.3}$$

Since we want only the volume to be computed, i.e., $\iiint_V dV$, then $(\nabla \cdot \mathbf{F}) = 1$. To facilitate the computation, the solid region is transformed in the coordinate system such that the reference plane is placed on the xy-plane and the chest-wall surface displacement is in the z axis (as seen in Figure 4.3). With this arrangement, by choosing $\mathbf{F} = z\vec{\mathbf{k}}$, the only part of (4.3) which needs to be computed is $\oiint_{S_{ch}} (\mathbf{F} \cdot \mathbf{n}_{ch})\, dS_{ch}$. Also, by considering the whole chest-wall surface as a set of tiny triangles T_i (obtained by triangulation), we compute the chest-wall volume by summing all these partial triangles as

$$V_t = \oiint_{S_{ch}} (\mathbf{F} \cdot \mathbf{n}_{ch})\, dS_{ch} = \oiint_{S_{ch}} (z\vec{\mathbf{k}} \cdot \mathbf{n}_{ch})\, dS_{ch} = \sum_{i=1}^{N_t} \oiint_{S_{T_i}} (z\vec{\mathbf{k}} \cdot \mathbf{n}_{T_i})\, dS_{T_i}, \tag{4.4}$$

where S_{T_i} denotes the surface of triangle T_i, $\mathbf{n}_{T_i}$ is the normal vector of S_{T_i} and N_t is the total number of triangles. To compute the surface derivative of each triangle, i.e., dS_{T_i}, for further simplification of (4.4), each triangle surface S_{T_i} is modelled using the parametric surface representation

$$\mathbf{P}_i(u_i, v_i) = p_{i0} + u_i\mathbf{E}_{i1} + v_i\mathbf{E}_{i2}, \qquad u_i \geq 0, v_i \geq 0, u_i + v_i \leq 1. \tag{4.5}$$

where $\mathbf{E}_{ij}$ are two vectors inside S_{T_i} defined as

$$\mathbf{E}_{ij} = p_{ij} - p_{i0} = (x_{ij} - x_{i0}, y_{ij} - y_{i0}, z_{ij} - z_{i0}), \qquad 1 \leq j \leq 2, \tag{4.6}$$

and $p_{ij} = \{(x_{ij}, y_{ij}, z_{ij}),\ 0 \leq j \leq 2\}$ are the vertices of triangle T_i defined in counter-clockwise order. Thus, given the infinitesimal measure of the surface area as

$$dS_{T_i} = \left| \frac{\partial \mathbf{P}_i(u_i, v_i)}{\partial u_i} \times \frac{\partial \mathbf{P}_i(u_i, v_i)}{\partial v_i} \right| du dv = \left| \mathbf{E}_{i1} \times \mathbf{E}_{i2} \right| du_i dv_i, \tag{4.7}$$

the final volume of the chest-wall surface at time point t is estimated as

$$V_t = \sum_{i=1}^{N_t} \oiint_{u_i v_i} (z\vec{\mathbf{k}} \cdot \mathbf{n}_{T_i}) \left| \mathbf{E}_{i1} \times \mathbf{E}_{i2} \right| du_i dv_i = \sum_{i=1}^{N_t} \left(\frac{1}{6} \sum_{j=1}^{3} z_{ij} \times \begin{vmatrix} x_{i1} & y_{i1} & 1 \\ x_{i2} & y_{i2} & 1 \\ x_{i3} & y_{i3} & 1 \end{vmatrix} \right). \quad (4.8)$$

Similar to other related works, e.g., [30, 32, 45, 136, 204], we also estimate the uncalibrated chest-wall volume by computing the average distance of all the pixels located in the chest-wall region. The chest-wall averaging method is simple to implement and fast to compute. We report results using both the chest-wall 3-D modelling and averaging methods.

4.1.4 Kinect Volume–time Data Smoothing

The Kinect volume–time data is obtained for each sequence by estimating the chest-wall volume as a function of time. Smoothing the data, in one form or another, is routinely applied in all other works, for example in [68, 125, 137, 141]. Here, although the bilateral filter is applied to each frame of the depth sequence, the obtained volume–time data still remains considerably noisy (see Figure 4.4) as the chest-wall volume is estimated temporally in a very limited breathing motion, i.e., $\pm 2.5cm$, approximately. Thus, we

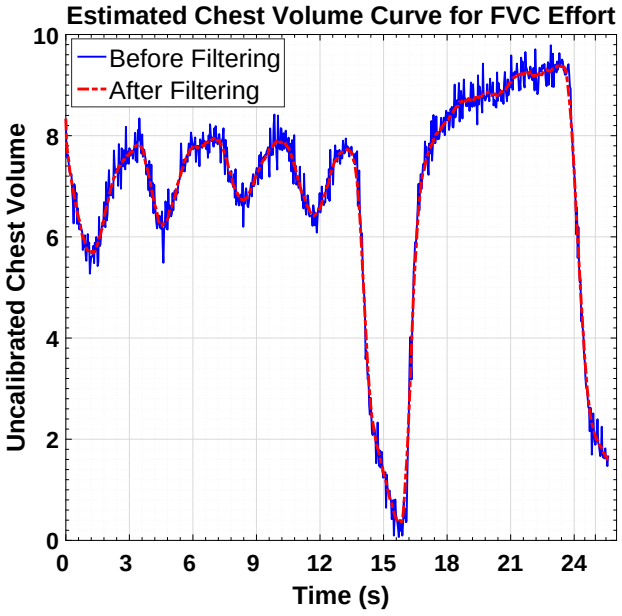


Figure 4.4: *Kinect volume–time data before and after smoothing by a moving averaging filter.*

use a non-causal moving average filter, which is a low-pass finite impulse response (FIR) filter [179], to eliminate the high frequency noise in the Kinect volume–time data as

$$V(t) = \frac{1}{N_W} \sum_{i=-(N_W-1)/2}^{(N_W-1)/2} V_{in}(t-i), \tag{4.9}$$

where $V_{in}(t)$ and $V(t)$ are the input and filtered volume–time data, respectively, and N_W is the averaging window size, which is selected as $N_W = 15$ based on the filter cut-off frequency of 1 Hz [133]. The cut-off frequency is chosen according to the range of respiratory rates (frequency) for healthy adults at $12-20$ bpm ($0.2-0.34$ Hz) [77], elderly at $16-25$ bpm ($0.27-0.42$ Hz) [127], and those with severe pulmonary disorders at 36 bpm (0.6 Hz) [55]. The computed range of respiratory rates for the 85 patients of our dataset, at $8-32$ bpm ($0.13-0.53$ Hz), satisfies the chosen cut-off frequency of 1 Hz.

4.2 Volume–time Data Analysis

All PFT manoeuvres start with a few cycles of normal breathing, called *tidal volume*, followed by the intended lung function test, called *main effort*. Since the Kinect volume–time data measures the chest-wall volume in cubic metres (m^3) relative to an arbitrary plane, as opposed to the spirometer's air volume measured in litres, we need to linearly scale the y-axis in the volume–time data (using computed scaling factors) to enable the correlation of the computed PFT measures. Note that this is not to imply that the depth sensor truly measures *lung volume*, but chest-wall volume is a proxy for the amount of air within the lungs that we show is linearly related to air flow as measured by the spirometer.

4.2.1 Keypoint Computation

Several keypoints are automatically computed from the volume–time data in order to,

- identify *tidal volume* and the *main effort* parts of volume–time data,
- establish scaling factors, and
- compute PFT measures.

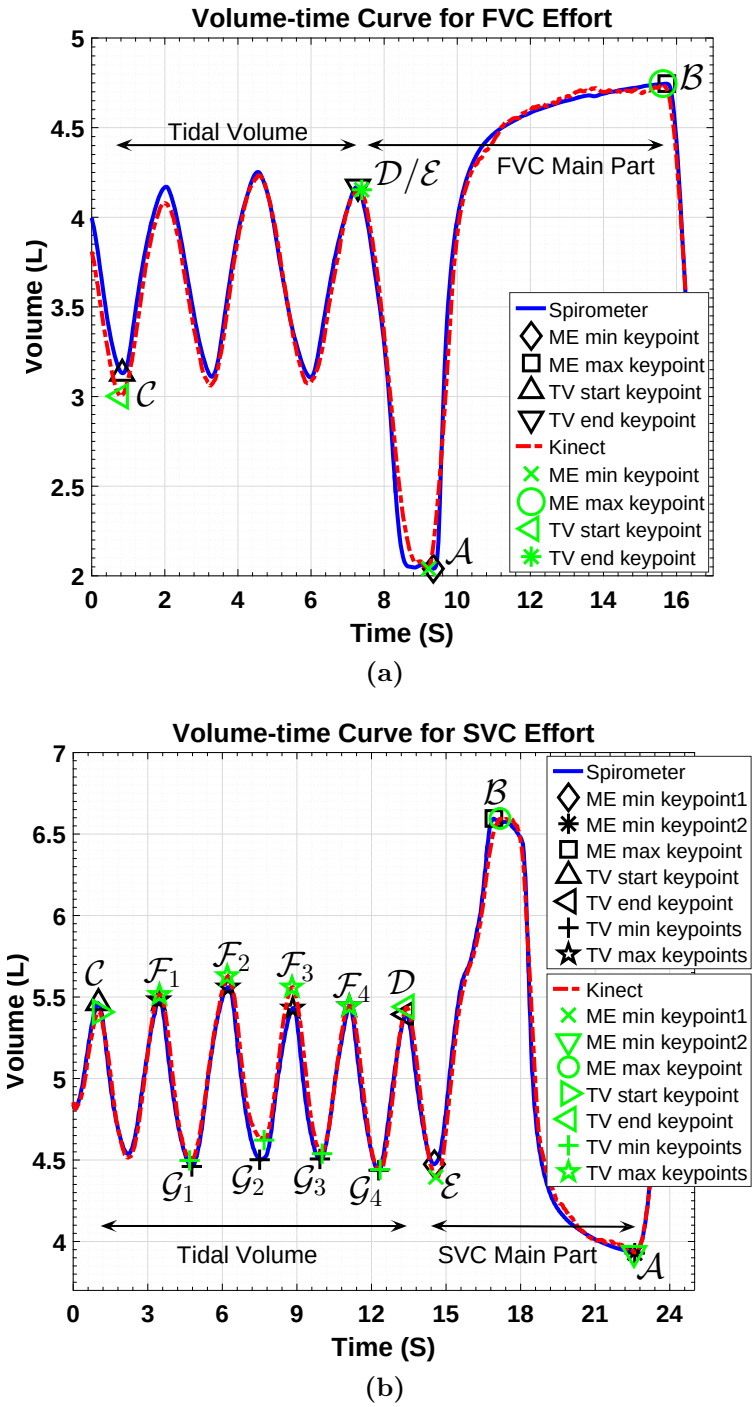


Figure 4.5: *Kinect and spirometer volume–time data and their corresponding keypoints for (a) an FVC, and (b) an SVC manoeuvre.*

Five keypoints are required for separating *tidal volume* and *main effort* in the FVC and SVC volume–time data $V(t)$, which are named as $\{\mathcal{C}, \mathcal{D}\}$ (beginning and end of *tidal volume*) and $\{\mathcal{E}, \mathcal{A}, \mathcal{B}\}$ (beginning to the end of *main effort*), as illustrated in Figure 4.5.

In order to compute the keypoints correctly, first we need to find the extrema which iden-

tify respiratory cycles during the PFT manoeuvre. Since the curve can be noisy (e.g., because of chest movement and coughing), local minima or maxima may be incorrectly selected. To prevent this, the difference between two consecutive turning points presenting local extrema, needs to be greater than a threshold γ. Considering V_{min} and V_{max} as the smallest and largest estimated chest-wall volume in a sequence (global minimum and maximum), $[V_{max} - V_{min}]$ indicates the maximum volume of exchanged air that occurs during *main effort*. A fraction of this exchanged volume is defined as γ to identify local extrema, i.e., $\gamma = \frac{1}{\rho}[V_{max} - V_{min}]$, where ρ is defined as the ratio of the greatest exhaled air during *main effort* (6.8 litres) to the smallest exhaled air during *tidal volume* (0.35 litres) among all sequences, which is $\rho \approx 20$.

Note that SVC volume–time data presents inhalation and exhalation in the opposite direction to the FVC, as seen in Figure 4.5. This means, while an increase in FVC volume–time data corresponds to exhalation, it indicates inhalation in the SVC data. This is similar to the volume–time data obtained from the spirometer.

FVC keypoints – Since lungs always contain a residual air volume, the amount of exhaled air volume in deep expiration is greater than inhaled air in a deep inspiration. Hence, keypoints $\mathcal{A}$ and $\mathcal{B}$, indicating the beginning and end of deep expiration respectively, are more detectable than other points, as seen in Figure 4.5a. These keypoints, timestamped $t_{\mathcal{A}}$ and $t_{\mathcal{B}}$ respectively, are computed as a pair of consecutive minimum and maximum points with the largest change in volume between them during expiration as

$$[t_{\mathcal{A}}, t_{\mathcal{B}}] = \underset{t_i^x, t_i^y}{\arg\min} \left\{ x(t_i^x) - y(t_i^y) \right\}, \qquad 1 = i \ldots n, \qquad \forall t_i^x, t_i^y \ni \; t_i^y > t_1^x, \tag{4.10}$$

where X and Y are sets of extrema computed as minima and maxima, $x(.) \in X$ and $y(.) \in Y$, t_i^x and t_i^y are each minimum and maximum corresponding timestamps, and n is computed as

$$n = \min(|X|, |\bigcup_{i=1}^{|Y|} y(t_i^y)|), \qquad t_i^y > t_1^x. \tag{4.11}$$

The local maximum directly before $t_{\mathcal{A}}$ was selected as $\mathcal{E}$ (and also $\mathcal{D}$ as in FVC, keypoints $\mathcal{D}$ and $\mathcal{E}$ are coincident in $V(t)$). The first extremum of the curve was selected as $\mathcal{C}$.

In addition to the volume–time data, we also use the flow–time data to compute some

FVC measures. The flow is defined as the rate of changing volume as

$$\dot{V}(t) = \frac{\partial V}{\partial t}. \tag{4.12}$$

FVC peak flow and time zero – To compute some FVC test measures, such as *FEV1*, we also need to compute the *Peak Flow* (*PF*) point and 'time zero' t_0 (Figure 4.6). *PF* is the point at t_{PF} with the maximum air flow speed during *main effort* exhalation,

$$t_{PF} = \underset{t\in[t_{\mathcal{A}},t_{\mathcal{B}}]}{\arg\max}\{\frac{\partial}{\partial t}\big(V(t)\big)\}. \tag{4.13}$$

Since *FEV1* is a timed PFT measure, instead of using keypoint $\mathcal{A}$ (timestamped $t_{\mathcal{A}}$), a starting 'time zero' t_0 keypoint is used for computing *FEV1*. This is because keypoint $\mathcal{A}$ is easily affected by hesitant or delayed exhalation in the *main effort* manoeuvre leading to an incorrect and decreased *FEV1* value. After subtracting $V(t_{\mathcal{A}})$ from the estimated volume, t_0 is computed using a standard clinical approach, i.e., back-extrapolation [129],

$$t_0 = t_{PF} - \big[V(t_{PF}) - V(t_{\mathcal{A}})\big] \times \left[\frac{\partial}{\partial t}\big(V(t)\big)\Big|_{t=t_{PF}}\right]^{-1}. \tag{4.14}$$

Figure 4.6 shows the t_0 and *PF* keypoints, and also the back-extrapolation method [129]. As seen, there is a notable difference between the t_0 and $t_{\mathcal{A}}$ as the start of the exhalation.

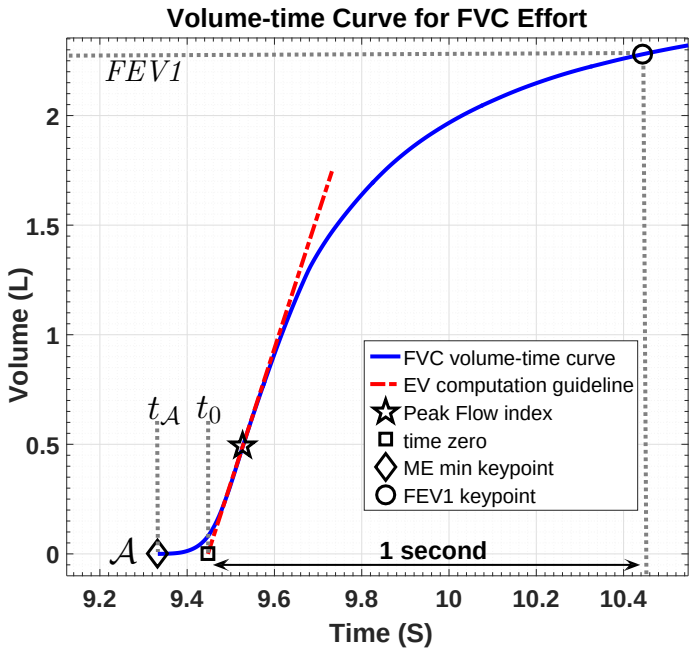


Figure 4.6: *'Time zero' and peak flow keypoints in FVC volume–time data.*

SVC keypoints – In the SVC test, we extract $\{\mathcal{C}, \mathcal{D}\}$ and $\{\mathcal{E}, \mathcal{B}, \mathcal{A}\}$ keypoints for partitioning the volume–time data into the *tidal volume* and *main effort* parts, respectively, as shown in Figure 4.5b. Similar to the FVC keypoints extraction method, to be able to find other keypoints, we first compute $\{\mathcal{B}, \mathcal{A}\}$ timestamps as,

$$[t_{\mathcal{B}}, t_{\mathcal{A}}] = \underset{t_i^y, t_i^x}{\arg\max} \big\{ y(t_i^y) - x(t_i^x) \big\}, \qquad i = 1 \dots m, \qquad \forall t_i^y, t_i^x \ni t_i^x > t_1^y, \tag{4.15}$$

where notations are similar to (4.10) and m is computed as

$$m = \min(|\bigcup_{i=1}^{|X|} x(t_i^x)|, |Y|), \qquad t_i^x > t_1^y \ . \tag{4.16}$$

Here, in the volume–time data $V(t)$, inhalation in SVC shows as exhalation in FVC. Thus, we still use the exhalation part of the *main effort*, which is more reliable, to extract $\mathcal{B}$ and $\mathcal{A}$, similar to the FVC test. Keypoint $\mathcal{E}$ marks the beginning of inhalation in *main effort* and is determined as the local minimum directly before $t_{\mathcal{B}}$. Like FVC, keypoint $\mathcal{C}$ is chosen as the first extremum of the curve, and $\mathcal{D}$ is the local maximum directly before $t_{\mathcal{B}}$. For computing SVC measures, four maxima keypoints $\{\mathcal{F}_i\}_{i=1}^{4}$ and four minima keypoints $\{\mathcal{G}_i\}_{i=1}^{4}$ from the *tidal volume* part are also extracted (see Figure 4.5b).

4.2.2 Tidal Volume Analysis and Calibration

To compute PFT measures, our method first needs to learn the calibration scaling factors per subject (intra-subject) by which it would be able to calibrate the Kinect volume–time data of a test sequence and compute its PFT measures without using any spirometer data. For this, we

(i) temporally align the Kinect and spirometer training data,

(ii) compute intra-subject scaling factors, and

(iii) use them to calibrate a Kinect test data.

We perform the alignment and scaling of the *tidal volume* and *main effort* parts separately, to take into consideration any inevitable trunk movement when subjects take a deep inhalation, followed by a maximal exhalation.

After selecting the *tidal volume* part of the Kinect and spirometer volume–time data using the $\mathcal{C}$ & $\mathcal{D}$ keypoints, we perform some pre-processing operations on these two

sub-signals to allow them to be directly processed and compared. The spirometer sub-signal is sampled at the Kinect sampling rate of 30 Hz. Both signals are normalized to zero mean. Finally, the two sub-signals are synchronized by computing the optimal time delay using windowed cross-correlation,

$$\tau_{delay} = \arg\max_{\tau} \left(\sum_{-\infty}^{+\infty} V_k^*(t) V_s(t+\tau) \right), \tag{4.17}$$

where $V_k^*(t)$ and $V_s(t)$ denote the complex conjugate of the Kinect normalized *tidal volume* and spirometer sub-sampled and normalized *tidal volume* data, respectively.

The *tidal volume* scaling factor can be computed using only a pair of consecutive minimum and maximum points, however this is not very reliable. We model it with a first degree polynomial,

$$\widehat{V}_s = \xi_{tv} \cdot \widehat{V}_k + \psi_{tv}, \tag{4.18}$$

where $\widehat{V}_s$ and $\widehat{V}_k$ are sub-sampled and aligned Kinect and spirometer *tidal volume* data, ψ_{tv} is the offset between the Kinect and spirometer *tidal volume* parts, and ξ_{tv} presents the *tidal volume* scaling factor. Since the Kinect and spirometer *tidal volume* parts are mean-zero normalized, then $\psi_{tv} \approx 0$.

However, in many cases, this approach is insufficient to deal with an incremental or decremental trend in the data that can appear in one or both of the Kinect and the spirometer data. Figure 4.7a shows example Kinect and spirometer *tidal volume* data

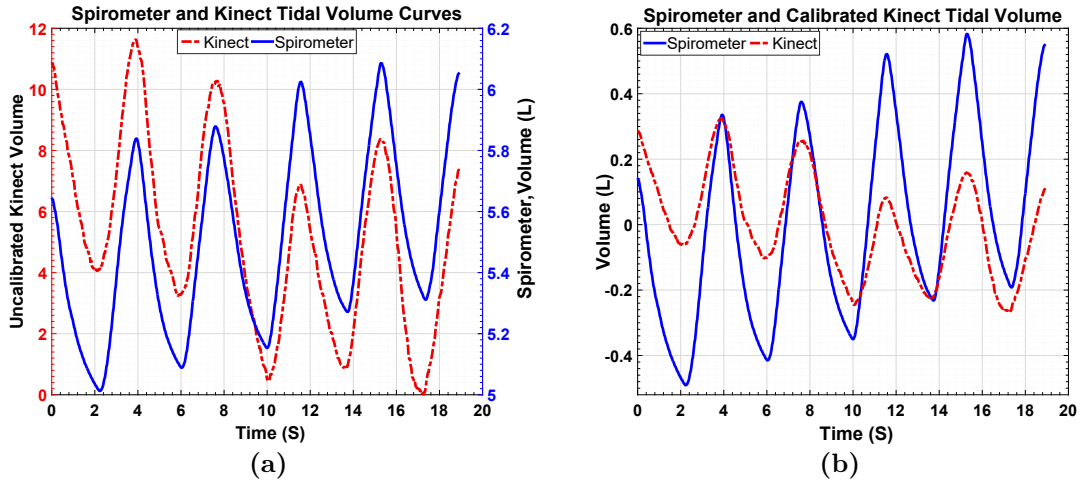


Figure 4.7: *(a) Existing trends in the Kinect and spirometer tidal volume data. (b) The Kinect data is incorrectly calibrated due to the existing trend.*

each plotted on a different scale, with the left y-axis for the uncalibrated Kinect volume and the right y-axis for the spirometer volume (litres). Both volume–time data exhibit such a trend which makes the extraction of a correct scaling factor (or an alignment process) a cantankerous task. Figure 4.7b shows the calibrated Kinect *tidal volume* data before the detrending process. As seen, the Kinect data has been incorrectly calibrated due to the existing trends which affect the computation of *tidal volume* scaling factors. This trend might occur due to one or more reasons: the use of a nasal Oxygen mask by patients during the test (which affects only the spirometer data), lung hyperinflation, or the subject's trunk movements.

A simple approach to modelling the trend to help eliminate it, would be linear regression. However, we find this to be insufficient due to the non-linear nature of the trend, thus we apply Empirical Mode Decomposition (EMD) [93] to estimate the trend more accurately. EMD is an adaptive method to decompose a non-linear and non-stationary signal in the time domain into its individual components, i.e., Intrinsic Mode Functions (IMFs) and a residual r,

$$s(t) = \sum_{j=1}^{l} IMF_j(t) + r(t), \tag{4.19}$$

from which no more IMFs can be extracted.

Figure 4.8a present the first three IMFs and the residual of a *tidal volume* data where the residual displays the signal trend (solid pink). Figure 4.8b shows the modified *tidal volume* data after subtracting the residual. Figure 4.8c shows the Kinect and spirometer *tidal volume* data with their trend estimated and removed by EMD, and the Kinect

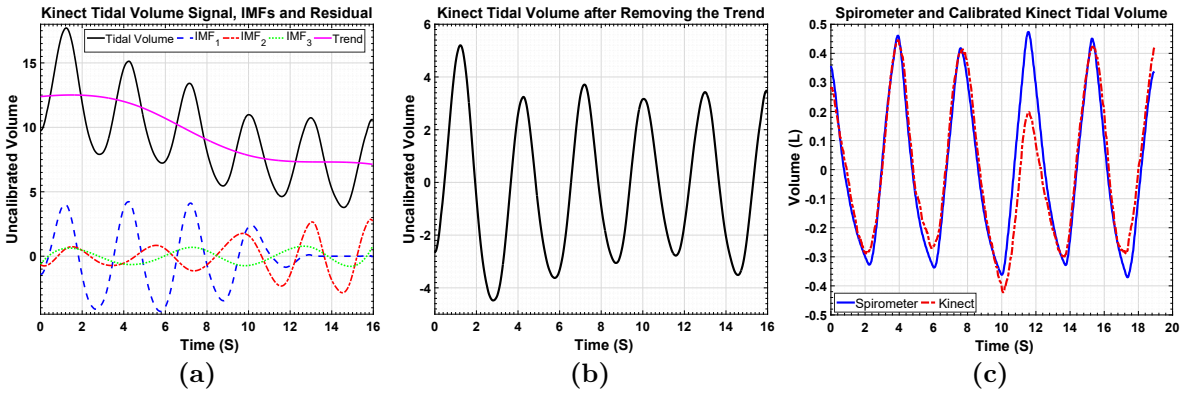


Figure 4.8: *(a) Original Kinect tidal volume data, IMFs and the residual signal. (b) Tidal volume data after removing the trend. (c) Correct calibration of Kinect tidal volume after removing the trend.*

volume–time data has been calibrated using the correct *tidal volume* scaling factor. Note we use the modified *tidal volume* data (detrended by EMD) only for temporal alignment and computing scaling factors while other analysis are performed on the original Kinect and spirometer data.

4.2.3 Main Effort Analysis and Calibration

As stated in Section 4.2.2, the Kinect and spirometer volume–time data are aligned only using their *tidal volume* parts to avoid errors arising from the subject's upper-body movement during *main effort*. Then, the *main effort* scaling factor (ξ_{me}) is obtained by solving

$$\widehat{V}_s = \xi_{me} \cdot \widehat{V}_k + \psi_{me}, \tag{4.20}$$

using only the $\mathcal{A}$ & $\mathcal{B}$ keypoints on each signal as they are less affected by motion artifacts and thus more reliable. Unlike in the *tidal volume* calibration process where ψ_{tv} is zero, here, ψ_{me} correlates with the upper-body movement and appears as an offset along the y-axis. However, in scenarios where subjects are stationary during the whole test (e.g., see Figure 4.5b), then $\psi_{me} \approx 0$, and there is no offset between the *tidal volume* and *main effort* parts.

We calibrate the *tidal volume* and *main effort* parts individually, and generate two calibrated Kinect volume–time data. For the first, the whole Kinect volume–time data is scaled by multiplying by the *tidal volume* scaling factor ξ_{tv}. Then, it is vertically aligned with the spirometer tidal volume part by making both the Kinect and spirometer tidal volumes zero-mean, as shown in Figure 4.9a. For the second, the whole Kinect volume–time data is scaled by multiplying by the *main effort* scaling factor ξ_{me}, and vertically aligned with the spirometer *tidal volume* part by adding the *main effort* offset ψ_{me} to all Kinect volume–time data, as shown in Figure 4.9b.

4.3 Scaling Factor Generalisation

So far we have shown that we can compute PFT measures from the Kinect volume–time and flow–time data which are calibrated using scaling factors computed from the corresponding spirometer volume-time data. We refer to this as an 'intra-test' procedure.

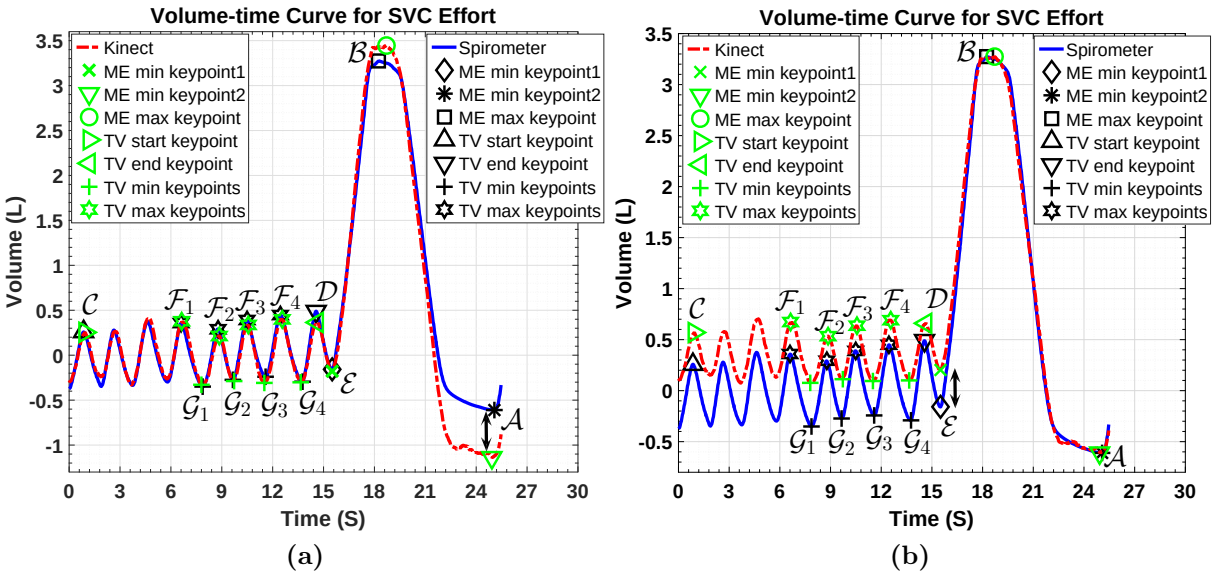


Figure 4.9: *Calibrating and aligning Kinect volume–time data using (a) tidal volume part, and (b) main effort part.*

However, we need to remove this dependency, so we can compute measures for a new trial[1] using only Kinect volume–time data, i.e., a more practical 'intra-subject' procedure.

While our experiments show that there is a linear relation between a subject's chest-wall motion and the volume of exchanged air, the coefficient of this relationship varies from one subject to another, since it depends on the subject's physical specifications, age, gender, race and health condition. In theory, this relationship should remain unchanged for a subject who performs a test several times (even on different days) with the same system configuration. However, in practice, this is only true for the *tidal volume* scaling factors, but not for the *main effort* scaling factor due to the subject's trunk motion. To overcome this problem, we analyse the *tidal volume* and *main effort* stages separately and compute scaling factors for them individually. Since there is no significant movement during *tidal volume*, it should be possible to detect trunk movement during *main effort* by comparing scaling factors ξ_{tv} and ξ_{me}. However, even when ξ_{tv} and ξ_{me} are very similar (i.e., $\xi_{tv}/\xi_{me} \approx 1$), which implies there is no trunk motion, the Kinect volume–time data might still be affected by trunk movements. This can be categorised in two ways:

(i) Backward motion at the beginning of deep inhalation (between $\mathcal{E}$ and $\mathcal{A}$ keypoints) for FVC and SVC tests, and

(ii) Forward lean at the beginning or middle of the deep and fast exhalation (after $\mathcal{A}$ in

[1] A trial refers to each performance of the FVC/SVC test by each subject.

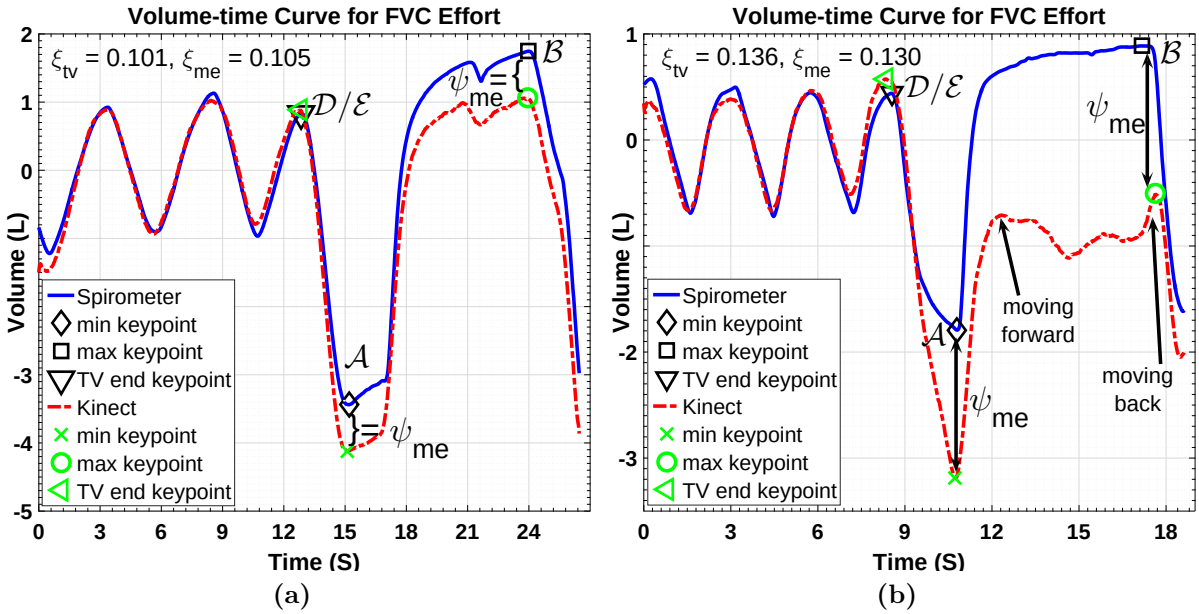


Figure 4.10: *Two different types of trunk motion patterns during main effort breathing. Even very close values of ξ_{tv} & ξ_{me} (specified in top of the figures) does not indicate the subject's trunk movement status, i.e., moving versus motionless, during the main effort breathing.*

both tests), and then a move back at the end of exhalation such that it compensates the first forward lean – which might be also accompanied by the motion pattern in (i) as well.

Figures 4.10a and 4.10b present two examples of volume–time data related to categories (i) and (ii), and their scaling factors. The effects of similar motion artifacts on chest-wall volume estimation have also been reported in [137, 206], previously.

The similarity of the motion patterns of trunk movements across different trials of a subject allows us to estimate the best matching scaling factors for calibrating the Kinect volume–time data of a new trial. This means that unless there is unexpected body movement, we can train our system to learn the *tidal volume* and *main effort* scaling factors for each subject, which enables us to compute PFT measures directly from the Kinect volume–time data without using spirometer data when testing.

Training phase – We use training data, provided as pairs of corresponding Kinect and spirometer volume–time data from training trials, to compute training *tidal volume* scaling factors $\{\xi_{tv}^{\ell}\}_{\ell=1}^{n_{tv}}$ and training *main effort* scaling factors and offsets $\{(\xi_{me}^{\ell}, \psi_{me}^{\ell})\}_{\ell=1}^{n_{me}}$, as explained in Sections 4.2.2 and 4.2.3. n_{tv} and n_{me} are the number of *tidal volume* and *main effort* training trials, respectively.

Testing phase – We calibrate the Kinect volume–time data of a test trial, by applying the best matching scaling factors and offsets learned from the training phase. Our analysis shows that similar intra-subject Kinect volume–time data can be calibrated using similar scaling factors and offsets as they present similar chest-wall movement which is a proxy for similar inhalation and exhalation air flow measured by the spirometer. Thus, to calibrate a test Kinect volume–time data, we find the best matching scaling factors and offsets among the training scaling factors using the curve similarity measures

$$F_{tv} = \frac{1}{4}\sum_{i=1}^{4}\Big[V_k(t_{\mathcal{F}i}) - V_k(t_{\mathcal{G}i})\Big], \quad \text{and} \tag{4.21}$$

$$F_{me} = \Big[V_k(t_{\mathcal{B}}) - V_k(t_{\mathcal{A}})\Big], \tag{4.22}$$

for *tidal volume* and *main effort* parts, respectively, where $V_k(t)$ is the original Kinect volume–time data, and $t_{\mathcal{A}}$, $t_{\mathcal{B}}$, $t_{\mathcal{F}_i}$ and $t_{\mathcal{G}_i}$ are automatically computed keypoint timestamps, as introduced in Section 4.2.1.

For the FVC test, the estimated *main effort* scaling factor ξ'_{me} is computed as

$$\xi'_{me} = \xi^k_{me} \quad \ni \quad k = \underset{j\in[1..n_{FS}]}{\arg\min}\Big\{\big|F^{test}_{me} - F^{j}_{me}\big|\Big\}, \tag{4.23}$$

where F^{test}_{me} denotes the *main effort* curve similarity measure extracted from the Kinect test volume–time data in (4.22), F^{j}_{me} is the same measure for the jth Kinect training volume–time data, j denotes different trials, $\{\xi^{\ell}_{me}\}_{\ell=1}^{n_{FS}}$ states the training *main effort* scaling factors, and n_{FS} is the total number of training FVC and SVC trials for this subject. Since vital capacity, $\big|V_s(t_{\mathcal{A}}) - V_s(t_{\mathcal{B}})\big|$, is equal for FVC and SVC tests (notwithstanding the reproducibility measurement error), we also use training SVC trials to estimate the best matching scaling factors for the FVC test trial. As no measure is computed from the *tidal volume* section in FVC tests, F_{tv} is not extracted and therefore, ξ'_{tv} is not computed.

Similarly, for the SVC test, the estimated *tidal volume* scaling factor ξ'_{tv} and the estimated *main effort* scaling factor and offset (ξ'_{me}, ψ'_{me}), are computed as

$$\xi'_{tv} = \xi^k_{tv} \quad \ni \quad k = \underset{j\in[1..n_S]}{\arg\min}\Big\{\big|F^{test}_{tv} - F^{j}_{tv}\big|\Big\}, \quad \text{and} \tag{4.24}$$

$$(\xi'_{me}, \psi'_{me}) = (\xi^k_{me}, \psi^k_{me}) \quad \ni \quad k = \underset{j\in[1..n_{FS}]}{\arg\min} \left\{ \left| F^{test}_{me} - F^j_{me} \right| \right\}, \tag{4.25}$$

respectively, where n_S is the total number of only SVC training trials. The *tidal volume* part of the FVC data is not used for the estimation of ξ'_{tv} because the *tidal volume* breathing cycles are too short in FVC and are not reliable for computing F_{tv} and consequently the *tidal volume* scaling factor. Note that in all FVC and SVC tests, $\psi'_{tv} \approx 0$.

The proposed scaling factor generalization approach is evaluated using leave-one-out cross-validation (LOOCV), which repeatedly takes one trial as the *test* and the rest as the training data. LOOCV is a more suitable approach than k-fold cross-validation or other conventional validation methods, due to the limited number of FVC and SVC trials for each subject.

After calibrating the Kinect volume–time data of the *test* trial using the estimated *tidal volume* and *main effort* scaling factors and offsets, PFT clinical measures are computed.

4.4 Clinical PFT Measures Computation

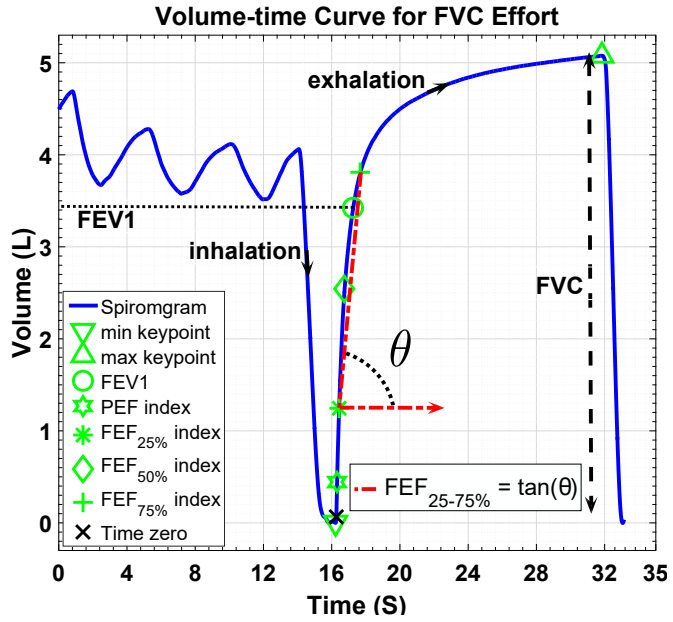
4.4.1 FVC Measures

Within an FVC spirometry test, several clinical measures are provided by the spirometer software. In addition to these numerical measures, there are two common 'qualitative' presentations of lung function test, i.e., volume–time curve (Figures 4.11a) and flow-volume curve (Figure 4.11b), that pulmonologists often use to visually diagnose problems in the patient's breathing function.

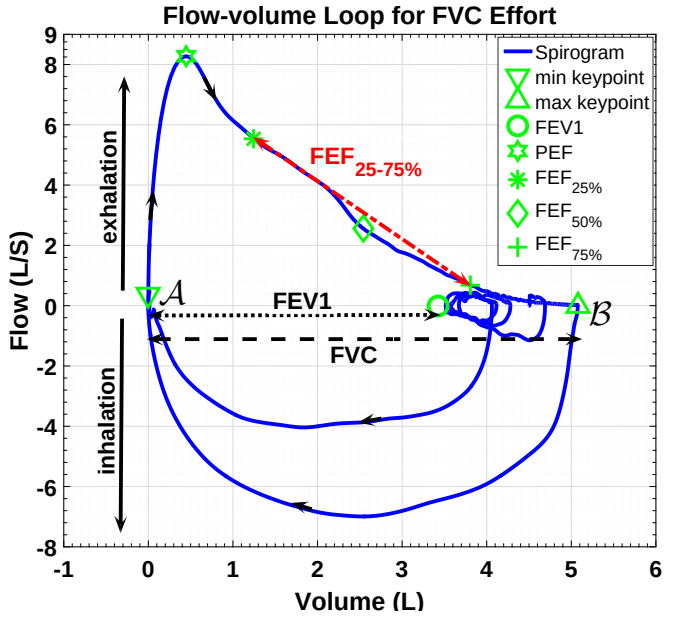
The 7 most significant FVC measures that we compute using the Kinect FVC volume–time and flow–time data are:

(i) *FVC* as the maximum amount of air in litres blown out after a maximal inhalation, determined as the volume change between keypoints $\mathcal{A}$ & $\mathcal{B}$, i.e.,
$FVC = \left[V(t_{\mathcal{B}}) - V(t_{\mathcal{A}})\right]$,

(ii) *FEV1* (Forced Expiratory Volume) as the volume of air forcibly expired in 1 second starting from 'time zero' (4.14), i.e., $FEV1 = \left[V(t_0 + 1) - V(t_0)\right]$,



(a)



(b)

Figure 4.11: *Clinical FVC measures illustrated on (a) a sample volume–time data, and (b) its corresponding flow–volume data extracted from our single-Kinect PFT dataset.*

(iii) PEF (Peak Expiratory Flow) as the maximum speed of exhaled air, i.e.,

$$PEF = \dot{V}(t_{PF}),$$

(iv) $FEF_{25\%}$ (Forced Expiratory Flow as flow of exhaled air at 25% of FVC, i.e.,

$$FEF_{25\%} = \dot{V}(t_{0.25FVC}),$$

(v) $FEF_{50\%}$ as flow of exhaled air at 50% of FVC, i.e., $FEF_{50\%} = \dot{V}(t_{0.5FVC})$,

(vi) $FEF_{75\%}$ as flow of exhaled air at 75% of FVC, i.e., $FEF_{75\%} = \dot{V}(t_{0.75FVC})$, and

(vii) $FEF_{25-75\%}$ as the mean forced expiratory flow between 25% and 75% of the FVC, computed as

$$FEF_{25-75\%} = \frac{0.75FVC - 0.25FVC}{t\,(FEF_{25\%}) - t\,(FEF_{75\%})}. \tag{4.26}$$

FVC, $FEV1$ and $FEF_{25-75\%}$ measures are illustrated in Figures 4.11a and 4.11b, and PEF, $FEF_{25\%}$, $FEF_{50\%}$ and $FEF_{75\%}$ measures are marked on the flow–volume curve in Figure 4.11b. Note that since the last four measures are computed using volume–time and flow–time data, only their corresponding locations are marked as 'index' on volume–time curve in Figure 4.11a using their timestamps.

4.4.2 SVC Measures

Within an SVC test, four clinical measures and only one 'qualitative' presentation of lung function, i.e., the volume–time curve (Figure 4.12), are provided by the spirometer software, which we compute on the Kinect volume–time data:

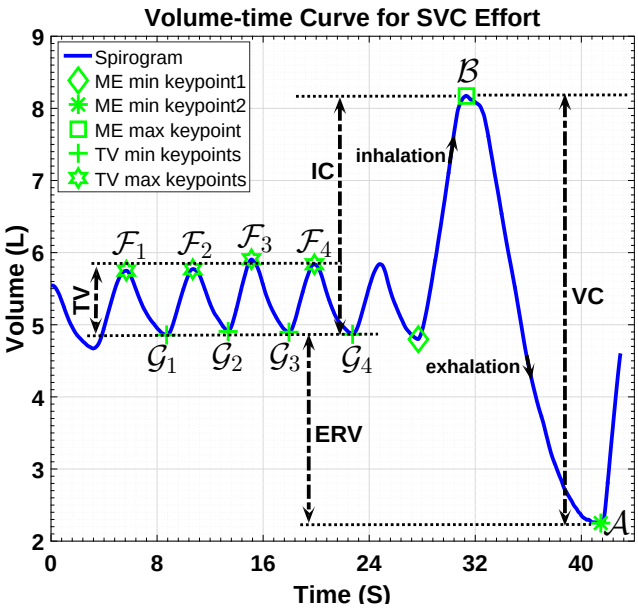


Figure 4.12: *Clinical SVC measures illustrated on a sample volume–time data extracted from our single-Kinect PFT dataset.*

(i) *VC* (Vital Capacity) as the volume change between full inspiration and complete expiration between keypoints $\mathcal{B}$ & $\mathcal{A}$,

$$VC = \big[V(t_{\mathcal{B}}) - V(t_{\mathcal{A}})\big], \tag{4.27}$$

(ii) *IC* (Inspiratory Capacity) as the volume change between taking a slow, full inspiration and the passive end-tidal expiration, i.e. difference of volume at keypoint $\mathcal{B}$ and the average volume at group keypoints $\mathcal{G}$ within the *tidal volume* section,

$$IC = V(t_{\mathcal{B}}) - \frac{1}{4}\sum_{i=1}^{4} V(t_{\mathcal{G}i}), \tag{4.28}$$

(iii) *TV* (Tidal Volume) as the volume of air inspired and expired at rest condition, i.e. the average volume difference between group keypoints $\mathcal{F}$ & $\mathcal{G}$,

$$TV = \frac{1}{4}\sum_{i=1}^{4} \big[V(t_{\mathcal{F}i}) - V(t_{\mathcal{G}i})\big], \tag{4.29}$$

(iv) *ERV* (Expiratory Reserve Volume) as the volume change between passive end-tidal expiration and complete expiration, i.e. difference of the average volume at group keypoints $\mathcal{G}$ within the *tidal volume* section and volume at keypoint $\mathcal{A}$,

$$ERV = \frac{1}{4}\sum_{i=1}^{4} V(t_{\mathcal{G}i}) - V(t_{\mathcal{A}}). \tag{4.30}$$

Figure 4.12 illustrates these measures on a sample SVC volume–time data from our single-Kinect PFT dataset. As noted previously in Chapter 3, each patient performed FVC and SVC tests several times (at least three) to ensure that the obtained PFT measures are in accordance with the spirometry ATS/ERS guidelines [128, 129].

4.5 Experimental Results

We evaluate our proposed method on the single-Kinect PFT dataset previously explained in Chapter 3. This dataset contains 529 sequences (260 FVC and 269 SVC) collected from 85 patients using both a Microsoft Kinect V2 and a spirometer.

4.5.1 Intra-Test PFT Evaluation Results

Tables 4.1 and 4.2 report the chest-wall 3-D modelling and the chest-wall averaging correlation coefficients (λ_v & λ_m) between the Kinect and the spirometer for all FVC and SVC test measures, along with the mean (μ_v & μ_m) and STD (σ_v & σ_m) of the L_2 error for all 85 subjects (529 sequences). For each measure, we also report the ratio of the mean of the L_2 error to the mean value of that measure (Ω_v & Ω_m).

The results show that the Kinect and the spirometer correlate well for the *FEV1* measure in the FVC tests and across all the SVC measures. The correlation amongst the other FVC measures is less strong due to the potential issues we describe later in Section 4.5.4. The results from both volume estimation methods are very close, with those from the chest-wall averaging method just edging ahead. Although the results from the chest-wall

Table 4.1: *Intra-test correlation coefficient ($\boldsymbol{\lambda_v}$ & $\boldsymbol{\lambda_m}$), mean ($\boldsymbol{\mu_v}$ & $\boldsymbol{\mu_m}$) and STD ($\boldsymbol{\sigma_v}$ & $\boldsymbol{\sigma_m}$) of L_2 error, and ratio of each measure's L_2 error to the mean value of that measure ($\boldsymbol{\Omega_v}$ & $\boldsymbol{\Omega_m}$) for FVC measures.*

		FVC (L)	***FEV1*** (L)	***PEF*** (L/S)	***FEF$_{25\%}$*** (L/S)	***FEF$_{50\%}$*** (L/S)	***FEF$_{75\%}$*** (L/S)	***FEF$_{25\text{-}75\%}$*** (L/S)
3-D Model	$\boldsymbol{\mu_v}$	0.006	0.285	1.685	1.696	0.931	0.576	0.757
	$\boldsymbol{\sigma_v}$	0.041	0.241	1.284	1.282	0.916	0.637	0.676
	$\boldsymbol{\Omega_v}$	0.002	0.137	0.490	0.597	0.375	0.559	0.414
	$\boldsymbol{\lambda_v}$	0.999	0.929	0.756	0.701	0.687	0.577	0.719
Averaging	$\boldsymbol{\mu_m}$	0.005	0.266	1.618	1.650	0.877	0.528	0.737
	$\boldsymbol{\sigma_m}$	0.039	0.217	1.259	1.246	0.830	0.576	0.665
	$\boldsymbol{\Omega_m}$	0.002	0.127	0.464	0.572	0.340	0.554	0.409
	$\boldsymbol{\lambda_m}$	0.999	0.940	0.774	0.719	0.729	0.595	0.728

Table 4.2: *Intra-test correlation coefficient ($\boldsymbol{\lambda_v}$ & $\boldsymbol{\lambda_m}$), mean ($\boldsymbol{\mu_v}$ & $\boldsymbol{\mu_m}$) and STD ($\boldsymbol{\sigma_v}$ & $\boldsymbol{\sigma_m}$) of L_2 error, and ratio of each measure's L_2 error to the mean value of that measure ($\boldsymbol{\Omega_v}$ & $\boldsymbol{\Omega_m}$) for SVC measures.*

		VC (L)	***IC*** (L)	***TV*** (L)	***ERV*** (L)
3-D Model	$\boldsymbol{\mu_v}$	0.011	0.045	0.066	0.049
	$\boldsymbol{\sigma_v}$	0.043	0.040	0.066	0.048
	$\boldsymbol{\Omega_v}$	0.004	0.019	0.072	0.105
	$\boldsymbol{\lambda_v}$	0.999	0.998	0.973	0.991
Averaging	$\boldsymbol{\mu_m}$	0.011	0.043	0.059	0.046
	$\boldsymbol{\sigma_m}$	0.045	0.040	0.065	0.046
	$\boldsymbol{\Omega_m}$	0.004	0.019	0.065	0.098
	$\boldsymbol{\lambda_m}$	0.999	0.998	0.976	0.992

averaging and 3-D modelling approaches should be identical in theory, we believe the 3-D modelling approach is slightly less accurate due to the chest-wall surface reconstruction errors. This confirms that the 3-D modelling approach, with its greater space requirements and time complexity, does not necessarily obtain better results than the simple and fast averaging method. The *FVC* and *VC* results (gray background columns) are highly correlated due to the rescaling of the y-axis in the volume–time data using their respective keypoints $\mathcal{A}$ & $\mathcal{B}$.

4.5.2 Intra-Subject PFT Evaluation Results

Tables 4.3 and 4.4 present the correlation coefficients (λ'_v & λ'_m), and the mean (μ'_v & μ'_m) and STD (σ'_v & σ'_m) of L_2 error for FVC and SVC computed measures for all 85

Table 4.3: *Intra-subject correlation coefficient ($\boldsymbol{\lambda'_v}$ & $\boldsymbol{\lambda'_m}$), mean ($\boldsymbol{\mu'_v}$ & $\boldsymbol{\mu'_m}$) and STD ($\boldsymbol{\sigma'_v}$ & $\boldsymbol{\sigma'_m}$) of L_2 error, and ratio of each measure's L_2 error to the mean value of that measure ($\boldsymbol{\Omega'_v}$ & $\boldsymbol{\Omega'_m}$) for FVC measures.*

		FVC (L)	***FEV1*** (L)	***PEF*** (L/S)	***FEF$_{25\%}$*** (L/S)	***FEF$_{50\%}$*** (L/S)	***FEF$_{75\%}$*** (L/S)	***FEF$_{25\text{-}75\%}$*** (L/S)
3-D Model	$\boldsymbol{\mu'_v}$	0.213	0.332	1.756	1.757	0.933	0.570	0.758
	$\boldsymbol{\sigma'_v}$	0.215	0.280	1.301	1.272	0.910	0.606	0.662
	$\boldsymbol{\Omega'_v}$	0.074	0.163	0.523	0.633	0.385	0.564	0.425
	$\boldsymbol{\lambda'_v}$	0.968	0.906	0.753	0.703	0.682	0.585	0.717
Averaging	$\boldsymbol{\mu'_m}$	0.200	0.299	1.717	1.735	0.882	0.509	0.727
	$\boldsymbol{\sigma'_m}$	0.186	0.243	1.286	1.254	0.822	0.540	0.670
	$\boldsymbol{\Omega'_m}$	0.071	0.146	0.508	0.621	0.354	0.553	0.417
	$\boldsymbol{\lambda'_m}$	0.975	0.927	0.769	0.715	0.715	0.603	0.721

Table 4.4: *Intra-subject correlation coefficient ($\boldsymbol{\lambda'_v}$ & $\boldsymbol{\lambda'_m}$), mean ($\boldsymbol{\mu'_v}$ & $\boldsymbol{\mu'_m}$) and STD ($\boldsymbol{\sigma'_v}$ & $\boldsymbol{\sigma'_m}$) of L_2 error, and ratio of each measure's L_2 error to the mean value of that measure ($\boldsymbol{\Omega'_v}$ & $\boldsymbol{\Omega'_m}$) for SVC measures.*

		VC (L)	***IC*** (L)	***TV*** (L)	***ERV*** (L)
3-D Model	$\boldsymbol{\mu'_v}$	0.237	0.269	0.118	0.297
	$\boldsymbol{\sigma'_v}$	0.239	0.269	0.137	0.310
	$\boldsymbol{\Omega'_v}$	0.084	0.116	0.129	0.592
	$\boldsymbol{\lambda'_v}$	0.956	0.915	0.888	0.737
Averaging	$\boldsymbol{\mu'_m}$	0.214	0.279	0.098	0.280
	$\boldsymbol{\sigma'_m}$	0.248	0.271	0.110	0.300
	$\boldsymbol{\Omega'_m}$	0.075	0.119	0.107	0.561
	$\boldsymbol{\lambda'_m}$	0.963	0.919	0.924	0.750

subjects. They also report the ratio of mean of the L_2 error to the mean value of that measure (Ω'_v & Ω'_m). Similar to the intra-test results, the chest-wall averaging method provides slightly better results.

The FVC test results for λ'_v & λ'_m in Table 4.3 indicate strong correlation of the *FVC* and *FEV1* measures against the spirometer, with the other five measures correlating reasonably well at a minimum of 0.603 for $FEF_{75\%}$ in the chest-wall averaging method. Furthermore, good correlation can be seen between the intra-subject and intra-test FVC measures (Tables 4.1 and 4.3).

The SVC test results for λ'_v & λ'_m in Table 4.4 also show strong correlation against the spirometer for *VC*, *IC*, and *TV* measures and good correlation for *ERV*. However, the differences between intra-subject mean (μ'_v & μ'_m) and STD (σ'_v & σ'_m) of errors (Table 4.4) and their intra-test counterparts (μ_v & μ_m and σ_v & σ_m from Table 4.2) are higher than these differences in FVC test. This is because SVC requires two scaling factors for the *tidal volume* and *main effort* parts of the curve, in addition to estimating the offset ψ'_{me}.

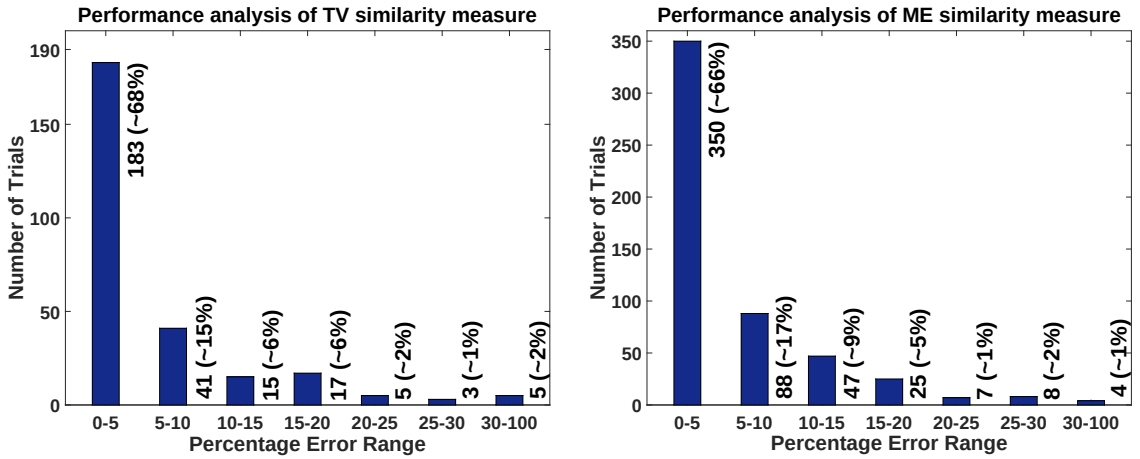
4.5.3 Statistical Analysis of Scaling Factors

The *tidal volume* and *main effort* test trials are calibrated using intra-subject scaling factors ξ'_{tv} & ξ'_{me}, which are chosen from the training sets $\{\xi^{\ell}_{tv}\}_{\ell=1}^{n_S}$ & $\{\xi^{\ell}_{me}\}_{\ell=1}^{n_{FS}}$, respectively, using (4.23), (4.24), and (4.25) based on the similarity measures in (4.21) and (4.22). The performance of the similarity measures, in terms of choosing the best intra-subject scaling factors from the training set, is evaluated by computing the normalised L_2 error:

$$\texttt{SME}^{tv} = \frac{\sqrt{(\xi'_{tv} - \xi^{c}_{tv})^2}}{\xi^{c}_{tv}} \qquad \text{and} \qquad \texttt{SME}^{me} = \frac{\sqrt{(\xi'_{me} - \xi^{c}_{me})^2}}{\xi^{c}_{me}}, \tag{4.31}$$

where ξ^{c}_{tv} & ξ^{c}_{me} are the closest scaling factors in the training set to the original scaling factors of the test trial ξ^{o}_{tv} & ξ^{o}_{me}. The original scaling factors are computed using the corresponding spirometer data as explained in Sections 4.2.2 and 4.2.3.

Figures 4.13a and 4.13b report the distribution of these errors for all *tidal volume* and *main effort* trials, respectively, in the range $0-30\%$ at 5% interval and then in the entire $30-100\%$ range. As can be seen, ∼83% of *tidal volume* scaling factors and ∼83%



(a) *269 Trials* **(b)** *529 Trials*

Figure 4.13: *(a) Performance analysis of intra-subject tidal volume similarity measure, i.e., F_{tv} (4.21). (b) Performance analysis of intra-subject main effort similarity measure, i.e., F_{me} (4.22).*

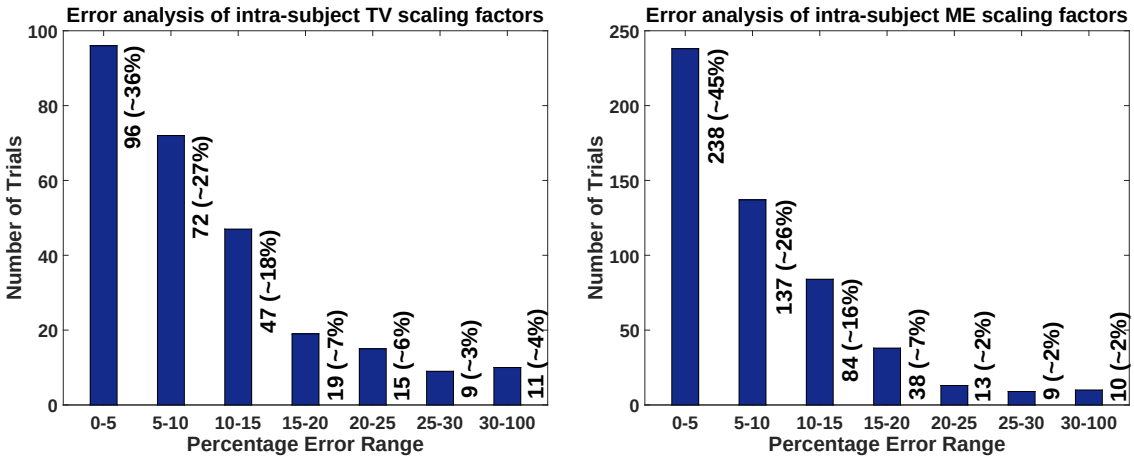
of *main effort* scaling factors are within an error of less than 10%. Only ∼2% of *tidal volume* scaling factors and ∼1% of *main effort* scaling factors have errors of greater than 30%.

Further, for each test trial, to compare the estimated intra-subject *tidal volume* and *main effort* scaling factors ξ'_{tv} & ξ'_{me} to the original scaling factors ξ^o_{tv} & ξ^o_{me}, their normalised L_2 error is computed as

$$\texttt{SCE}^{tv} = \frac{\sqrt{(\xi'_{tv} - \xi^o_{tv})^2}}{\xi^o_{tv}} \qquad \text{and} \qquad \texttt{SCE}^{me} = \frac{\sqrt{(\xi'_{me} - \xi^o_{me})^2}}{\xi^o_{me}}. \tag{4.32}$$

As seen in Figures 4.14a and 4.14b, which present the distribution of errors for all *tidal volume* and *main effort* trials, ∼81% of *tidal volume* scaling factors and ∼87% of *main effort* trials have an error of less than 15%. Only ∼4% of *tidal volume* scaling factors and ∼2% of *main effort* scaling factors have an error of greater than 30%.

Considering ATS/ERS acceptable repeatability criterion [128, 129], i.e., obtaining at least three *FVC* or *VC* measures with < 0.150 litres error across the repeated manoeuvres, the acceptable range of intra-subject scaling factor is based on the subject's lung volume and pathology, age, body physical specifications and other clinical aspects. Thus, if the estimated intra-subject scaling factor can calibrate the Kinect data with < 0.150 litres error, the estimated scaling factor provides a clinically approved measurement and the error would be in an acceptable range. For example, if the *main effort*

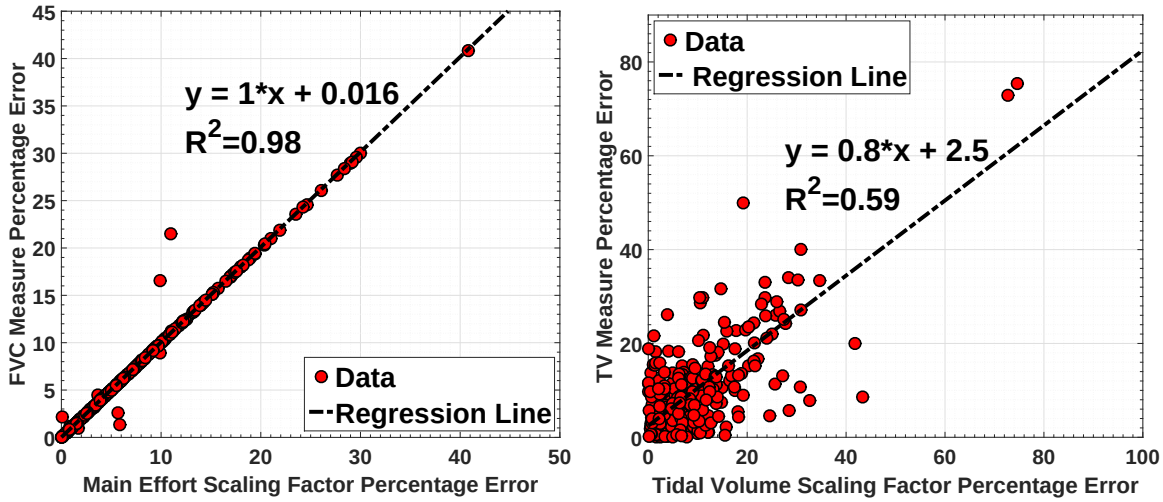


(a) *269 Trials* **(b)** *529 Trials*

Figure 4.14: *(a) Intra-subject tidal volume scaling factors error analysis. (b) Intra-subject main effort scaling factors error analysis.*

original scaling factor $\xi^{o}_{me} = 2.92$ calibrates the Kinect *VC* measure to 3.66 litres, then the estimated intra-subject scaling factor $\xi'_{me} = 2.81$ for which the error is ∼4%, calibrates the measure to 3.52 litres which is still in the clinically approved range.

We also analysed the correlation between the *tidal volume* and *main effort* scaling factor normalized L_2 errors $\texttt{SCE}^{tv}$ & $\texttt{SCE}^{me}$, and error of FVC and SVC computed measures. Figures 4.15a and 4.15b, present this correlation for *FVC* and *TV* measures. As can



(a) **(b)**

Figure 4.15: *(a) Correlation between the FVC measure error and the main effort intra-subject scaling factor error* $\texttt{SCE}^{me}$*. (b) Correlation between the TV measure error and the tidal volume intra-subject scaling factor error* $\texttt{SCE}^{tv}$*.*

be seen, there is a high correlation between the FVC measure error and the *main effort* scaling factor error across all trials. This correlation is less strong for the TV measure error and *tidal volume* scaling factor error. The reason for this is, *tidal volume* scaling factors are computed using all data points of *tidal volume* part of volume–time data, and TV measure itself is computed using group keypoints $\mathcal{F}$ & $\mathcal{G}$ (4.29). However, FVC measure and *main effort* scaling factors are both computed using the same keypoints $\mathcal{A}$ & $\mathcal{B}$. Thus, they are better correlated (Figure 4.15a) than the TV measure error and the *tidal volume* scaling factor error (Figure 4.15b).

4.5.4 Measurement Stability

It is important to note that even spirometer readings differ between multiple consecutive trials for the same subject, thus requiring at least three trials with similar readings before a clinician considers the results. This is illustrated in Figures 4.16a, 4.16b, 4.16c and 4.16d which present some example measures (FVC, $FEF_{25\%}$, PEF, and TV), provided by the spirometer and the proposed method for one subject from four consecutive trials.

To find out the correlation between spirometry reproducibility and the proposed method's error in computation of measures, we obtained the STD of each measure and its corresponding error in all repeated trials for each patient. Figures 4.16e, 4.16f, 4.16g and 4.16h show the computed correlation for FVC, $FEF_{25\%}$, PEF, and TV measures. These results indicate that when the measures provided by the spirometer are less consistent, the error between measures obtained by the proposed method and the spirometer increases.

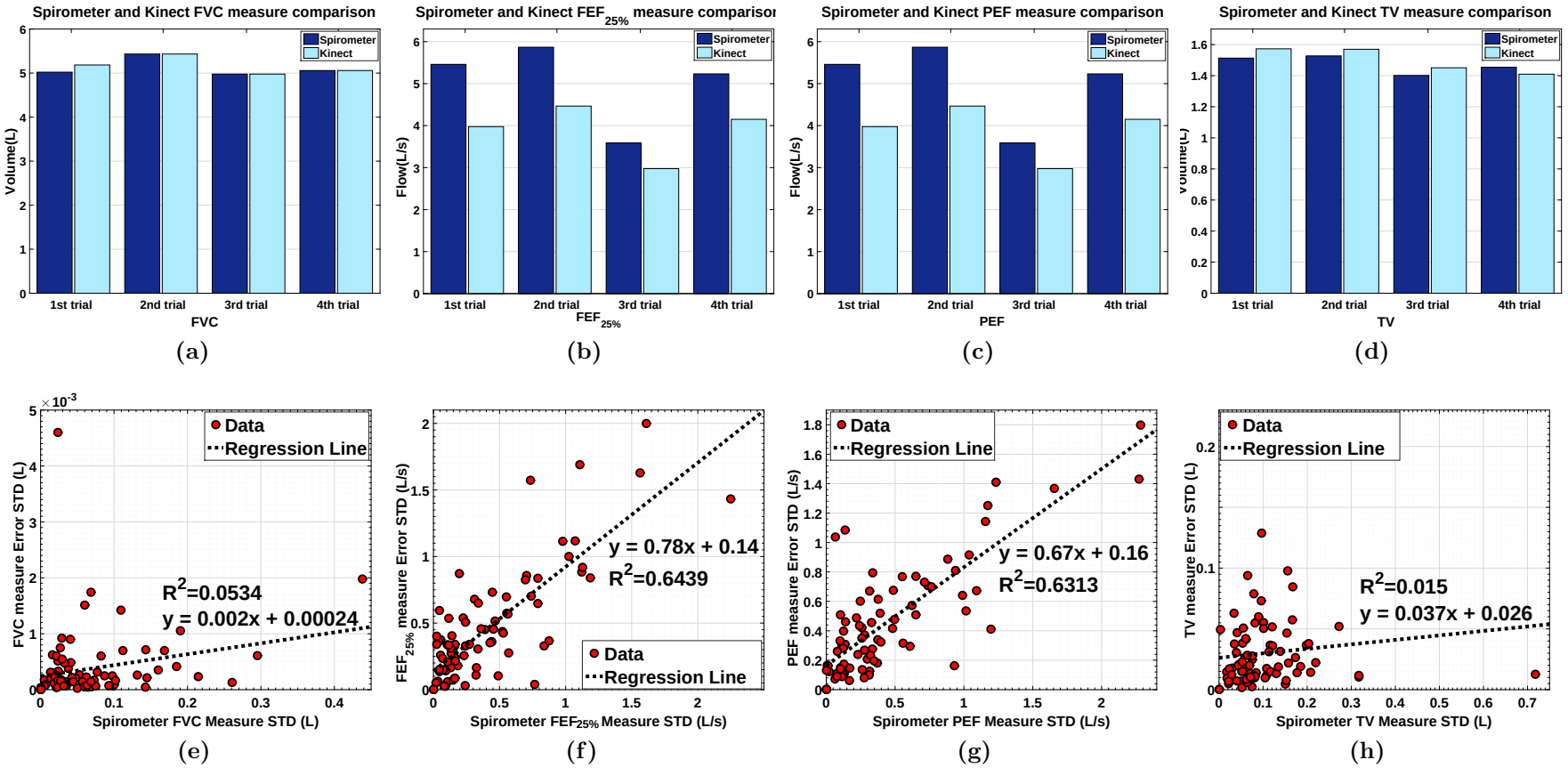


Figure 4.16: *(a)-(d) Comparing the Kinect PFT measures FVC, $FEF_{25\%}$, PEF and TV, to the spirometer measures for 4 trials of a sample patient. (e)-(h) Correlation analysis between the standard deviation of the Kinect computed PFT measures (FVC, $FEF_{25\%}$, PEF and TV) error, and the standard deviation of the PFT measures ground truth values from the spirometer for the whole trials of all 85 patients.*

The subject's trunk movement during a test is a primary reason for poor correlation and this is more evident in *main effort* measures. A specific example of how trunk movement (due to expiration pressure) can affect the $FEF_{25-75\%}$ measure is in Figure 4.17, where the estimation of $0.75FVC$ is sometimes compromised. In another observation, illustrated in Figure 4.18, we found that as the $FEF_{25\%}$ and PEF readings from the spirometer

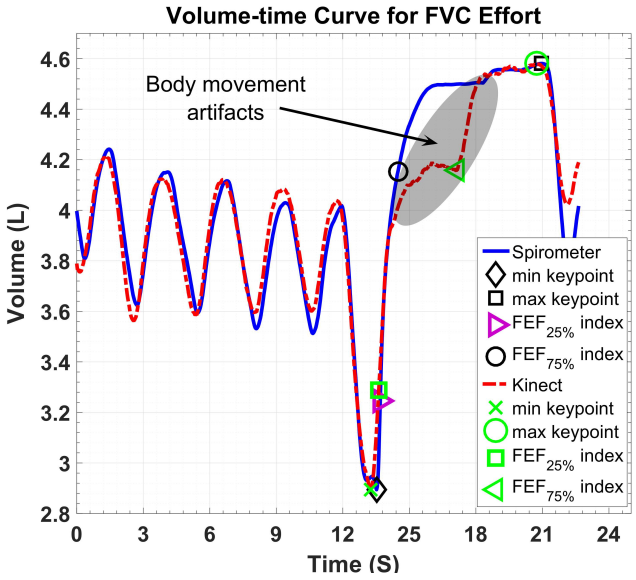
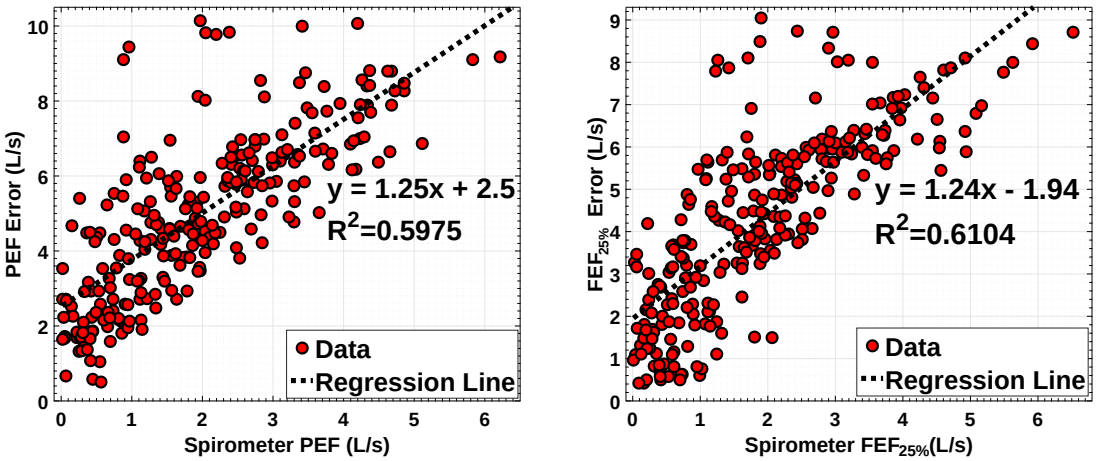


Figure 4.17: *Trunk motion artifacts effect on the computation of* $FEF_{25-75\%}$.



(a) (b)

Figure 4.18: *Analysis of the correlation between the Kinect PEF and* $FEF_{25\%}$ *measures error and their spirometer values over the whole FVC sequences. As seen, there is a linear relationship between the error of the Kinect PEF and* $FEF_{25\%}$ *measures and the ground truth values. The greater the measures is, the greater their error. Potentially, this is due to extensive trunk motion artifacts when patients blow harder and faster into the spirometer to achieve higher (better) PEF and* $FEF_{25\%}$ *measures.*

increases, our proposed method's error also increases. To the best of our knowledge, this happens as subjects try to attain better lung function measures by blowing faster into the spirometer which inevitably results in more trunk movement. *PEF* and $FEF_{25\%}$ are more affected by the patient's trunk translation because,

- they are calculated using flow data which is the first derivative of the volume over time and so is more sensitive to displacements, and
- *PEF* and $FEF_{25\%}$ are located at the beginning of the *main effort* section (Figure 4.19), which is more affected by the movement.

Even subtle movements caused by leaning forward, due to forcible expiration, affects keypoint positions of these measures. In Figure 4.19a, although the *main effort* parts of the curves match very well, their flow–volume curve is considerably different in Figure 4.19b between the start of exhalation and the location of the $FEF_{50\%}$ point.

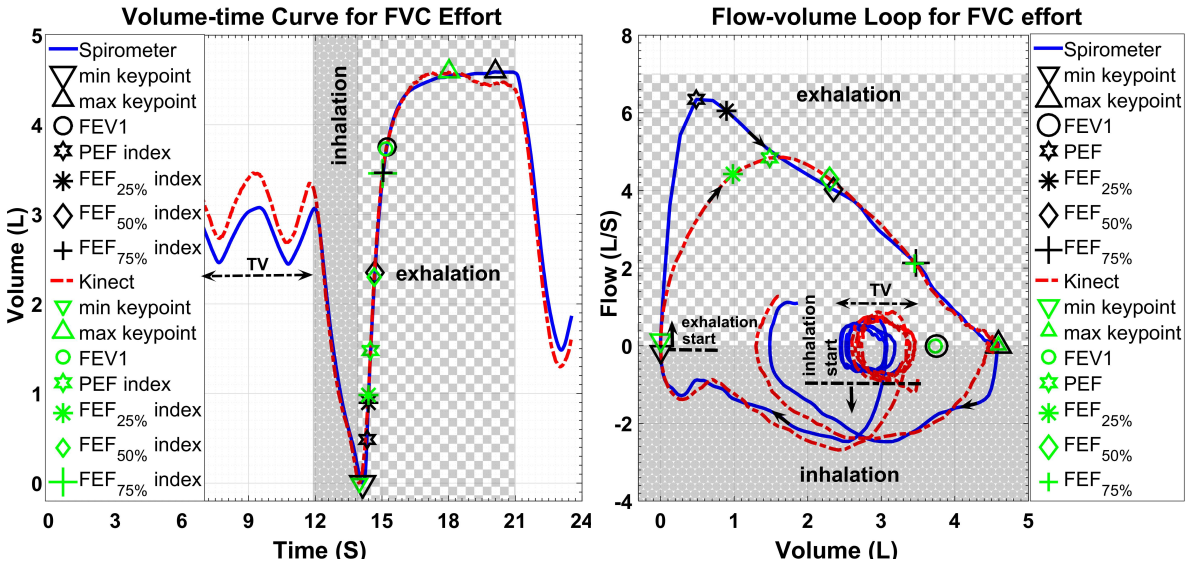


Figure 4.19: *Qualitative comparison of a sample Kinect/spirometer volume–time data and their PFT measure. Although the Kinect and spirometer volume–time data are very similar in (a) at the main effort part, the corresponding flow–volume data is different in (b) at the beginning of exhalation.*

4.6 Conclusion

We proposed a remote, non-invasive depth-based approach for Pulmonary Function Testing. The proposed system generates Kinect-based volume-time and flow-time curves, and by locating several keypoints automatically, we computed several FVC and SVC measures which we compared against a spirometer, and evaluated their reproducibility. We analysed the subject's trunk motion pattern to generalize scaling factors to be able to compute intra-subject PFT measures for each subject, without having to use a spirometer to calibrate against for each trial. We validated our system in a clinical environment with 85 actual patients and achieved high intra-test and intra-subject correlation against the spirometer.

In our future work, we plan to use two Kinects to decouple trunk motion and chest-surface motion to increase the accuracy of our PFT measures. Another potential development could be applying machine learning techniques to generalise the scaling factors by introducing parameters such as height, weight and age in the estimation and remove the need for subject-specific spirometry.

Chapter 5

3-D Data Acquisition and Registration by Two Opposing Depth Sensors

Methods for capturing the full extent of an object, or a complete scene, have been proposed using handheld depth sensors and temporal fusion [96, 135, 212]. Alternatively, static multi-sensor setups with varying overlapping requirements between the sensors have been proposed to reconstruct dynamic scenes on frame-level basis [28, 60, 106, 130, 155, 171]. These avoid the need, and challenge, for alignment and fusion between frames and can readily reconstruct dynamic scenes and deformable objects in real-time. For example, Kowalski et al. [106] recently presented a 3D data acquisition system, using up to four Kinect V2 sensors, in which they manually calibrated their system to register the point clouds from each sensor in a two-step procedure involving rough estimation and refinement. Their qualitative-only results showed good subjective performance for general static and dynamic object reconstruction. However, their calibration stage is cumbersome requiring self-designed markers, manual labelling of marker's locations, and sufficient overlap between the sensors.

While this work belongs to the category of static multi-sensor setup, we rely on a simplified approach of using two static *opposing* Kinect sensors with no overlap. Each sensor can perceive *nearly* half of the object, resulting in frame-level reconstruction of dynamic objects. This is valuable in applications where a less intrusive and more easily configured setup is necessary, which inherently means as few sensors as possible and as little incon-

venience to the subject as possible. One such application is within healthcare pertaining to respiratory measurements or pulmonary function testing, for which depth-based approaches have recently emerged [136, 137, 180, 182, 206], albeit for a single sensor.

The proposed approach is able to reconstruct rigid and dynamic objects to high accuracy, which we evaluate quantitatively on rigid objects and qualitatively on animated subjects. Ease of setup and high accuracy (range of average errors is $0.21-0.84$ cm across 3 objects and 3 placements) is achieved through,

- a fast and automatic calibration process using double-sided calibration chessboards placed at varying depths,
- synchronising intra-Kinect RGB and depth channels as well as two data acquisition PCs, and
- a highly accurate point cloud registration approach using only the infrared stream to specify real world coordinates, as opposed to using RGB and depth which is likely to increase registration error.

The main contributions of our work are twofold. First, the deployment of only two Kinect sensors for 3D data capture minimises the overall operation space, reduces the system setup and calibration effort, lowers system costs, and minimises the temporal frame alignment error. Second, unlike many other previous approaches, which require a considerable overlap between point clouds for registration [100, 106, 130, 155, 171], our proposed method is able to perform temporal and spatial alignment of two non-overlapping point clouds.

In this chapter, we present our open source[1] 3-D data acquisition and calibration pipeline using two opposing RGB-D sensors (Kinect V2). First, the relative pose of the two Kinects is estimated through a calibration stage. In this stage, after correcting the lens distortion of the two Kinect depth sensors, the crossing points of three *double-sided* chessboards are automatically detected from the illumination-normalised infrared images captured by the two facing Kinects. Automatically establishing the corresponding crossing points, the rigid transformation parameters are computed using a singular value decomposition approach (Section 5.1). Transferring the two Kinects' point clouds into a joint coordinate system, these parameters are then used to align and register point

[1] https://github.com/BristolVisualPFT/

clouds at frame level in the registration and reconstruction stage (Section 5.2). We validate the proposed system quantitatively by performing experiments on rigid objects in which dimension, surface information and volume of three boxes of known-size are automatically measured and compared against ground truth measurements. The quantitative validation results confirm the proposed method's accurate spatial alignment. To assess our method's ability in temporal synchronisation, we qualitatively report on non-rigid object reconstruction by way of human subjects in motion, e.g., waving hands or undergoing respiratory functional assessment (Section 5.3). We finish this chapter with a conclusion on our proposed method and giving directions for potential extension for future work (Section 5.4).

5.1 Synchronisation and Calibration

Figure 5.1 presents an overview of our proposed calibration approach.

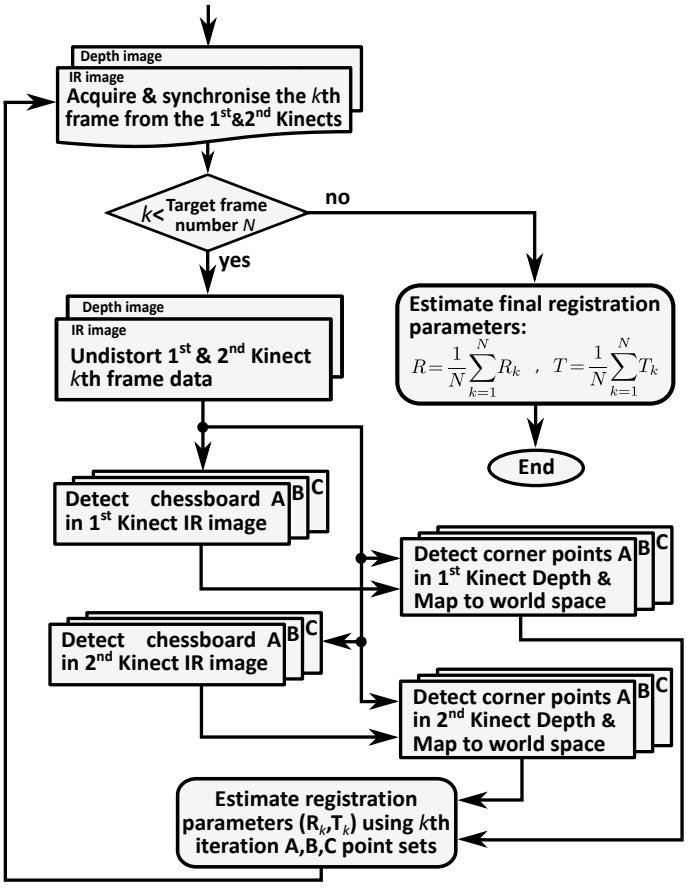


Figure 5.1: *Schematic of our proposed calibration approach.*

Figure 5.2: *Two facing Kinects are placed at a distance of ∼3m from each other which provides sufficient space for the data acquisition. The optimal location for minimal depth measurement noise is in the middle of the two Kinects according to our study in Chapter 3. Three double-sided calibration chessboards are placed at different depths to provide more distinct correspondences over the whole calibration space.*

5.1.1 System Configuration and Setup

We use two Kinects facing each other with ∼3m distance between them as seen in Figure 5.2, which allows objects to be captured at the optimal distance away from each sensor. For registering and aligning two sets of 3-D points, we need at least three corresponding and distinct 3-D points in each point set [35]. However, using more distinct points distributed over the calibration space improves the alignment accuracy and decreases the registration error.

The easiest solution for calibrating multiple depth sensors would be deploying calibration boxes [100] by which the calibration parameters can be estimated using the boxes specifications, e.g., corners for establishing the correspondence points. However, this is not an appropriate solution for us because,

- Kinects cannot acquire all of the corners of calibration boxes depth data due to the Kinects facing topology,
- depth measurement at edges and corners is noisy and inaccurate, and
- calibration cannot be performed automatically as corners must be selected man-

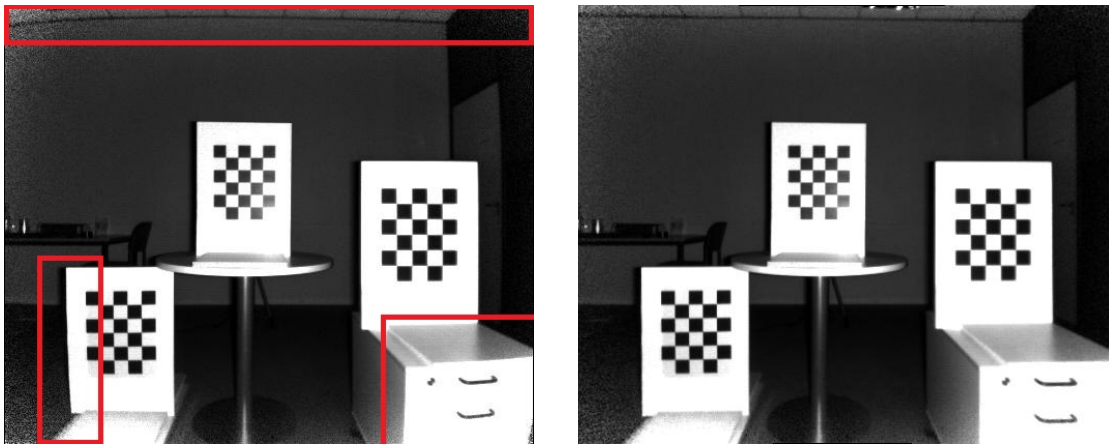
ually which not only affects the overall registration accuracy but also makes it a bothersome calibration method for users.

Thus, to help with the calibration, we use three double-sided chessboards which are placed at different depths from the Kinects (Figure 5.2). To make a double-sided chessboard, a 5×6 square pattern (with each square size of $55 \times 55mm^2$) is printed on two $A3$ sheets, which are then joined back to back and held by a frame such that the chessboards' crossing points are aligned as precisely as possible. This solution provides us with three groups of points (3×20 inner points in total) so that the points in any group have considerably different (x, y, z) coordinates from points in the other groups.

5.1.2 Data Acquisition and Synchronization

Unlike Kowalski et al. [106], our system is designed to capture all four of RGB, depth, infrared and body joints data in simultaneous processing threads at full device frame rate ($30\,fps$). Online visualisation is possible, although at the expense of lower frame rates. Our proposed system is able to generate RGB point clouds from pre-recorded and synchronised RGB and depth data.

Here, we wish to achieve 'synchronization' between corresponding frames of different data modalities in each Kinect separately (intra-Kinect) and also, between corresponding frames of the same type in different Kinects (inter-Kinect). Intra-Kinect synchronization is necessary to identify temporally corresponding RGB, depth, and skeleton data frames in each Kinect, which is simply performed by using the timestamps provided for each data frame in a Kinect. Inter-Kinect synchronization is achieved by synchronising the system time of two locally networked PCs (one for each Kinect) using Network Time Protocol (NTP) and recording the Kinects' system and threads timestamps for aligning each data frame. Since there are no means of triggering multiple Kinects simultaneously by software control commands, this can cause a maximum lag of 33 ms between our two Kinects, which would cause a synchronization error of at most one frame. We reduce this error by sending only one trigger command at the beginning of the capture, from one machine to another through the network, however the error is dependant on the network traffic and speed. Note that, the more Kinect sensors within a system, as in [13, 14, 15, 28, 29, 106, 123], the greater is this error.



(a) *Before distortion correction* **(b)** *After distortion correction*

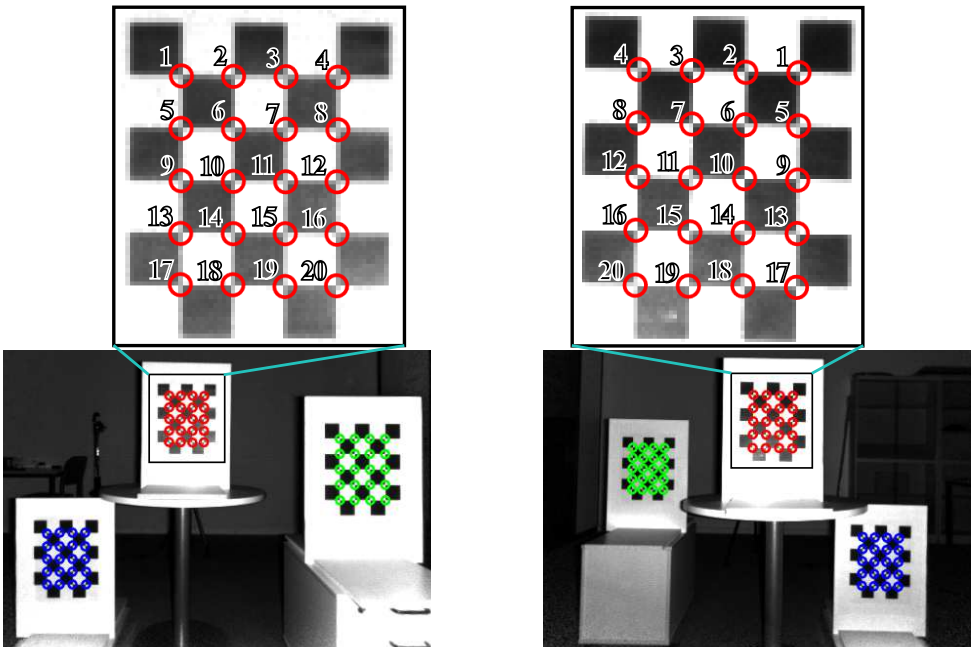
Figure 5.3: *A sample of Kinect infrared images before (a) and after (b) lens distortion correction.*

5.1.3 Lens Distortion Correction

Kinect V2 depth images are computed from the captured infrared images and therefore, both images have the same optical specifications. Similar to other lens-based imaging devices, the Kinect also suffers from lens distortion. Thus, both the infrared and depth image distortions are corrected by applying the Brown model [43] using the manufacturer's built-in radial distortion coefficients specifically extracted from each Kinect. Since the manufacturer's built-in tangential distortion coefficients are zero, they have not been taken into account in distortion correction. Figures 5.3a and 5.3b show an infrared captured image before and after lens distortion correction, respectively.

5.1.4 Establishing Crossing Points Correspondences

We use real world coordinates of the crossing points of three double-sided chessboards (3×20 points) to align point clouds and register them to a joint coordinate system. Previous approaches [13, 14, 15, 100, 106, 123] combined RGB and depth data to obtain coordinates of points required for calibration. However, we detect the coordinates of the crossing points [79] from the intensity and depth space obtained by illumination-normalised infrared images and depth images, respectively. Using only infrared sensor instead of using both RGB and depth, increases point cloud registration accuracy by eliminating the error caused by RGB to depth space mapping. Figures 5.4a and 5.4b illustrate the detected chessboards' crossing points where the corresponding crossing point sets of each chessboard in the two Kinects' infrared images are indicated in the



(a) (b)

Figure 5.4: *Establishing crossing points correspondences from the illumination-normalised infrared images captured by the two Kinects. The corresponding crossing point sets of each chessboard each view are indicated in the same color. The zoomed-in chessboard shows the established correspondences in more details.*

same color.

5.1.5 Kinects Pose Estimation

As there is an insufficient number of overlapping points in our point clouds, an iterative point matching algorithm, like ICP [35, 208], is unsuitable for aligning them. Thus, we considered one Kinect's coordinate system as reference, and then the other Kinect's relative pose was estimated using translation $\mathbf{T}$ and rotation $\mathbf{R}$ matrices. The transformation matrices $\mathbf{T}$ and $\mathbf{R}$ was computed by registering 3×20 corresponding crossing points of the reference and the second Kinect, i.e., $\mathbf{Q}$ and $\mathbf{Q}'$, as

$$\mathbf{Q}' = \mathbf{R} \times \mathbf{Q} + \mathbf{T}. \tag{5.1}$$

The rotation matrix $\mathbf{R}$ is computed by applying singular value decomposition on a cross-covariance matrix $\mathbf{M}$ created using $\mathbf{Q}$ and $\mathbf{Q}'$ point sets [49],

$$\mathbf{M} = \frac{1}{N}\sum_{j=1}^{N}\left[(\mathbf{Q}_j - \mathbf{Q}_\mu)(\mathbf{Q'}_j - \mathbf{Q'}_\mu)^T\right], \tag{5.2}$$

where $\mathbf{Q}_j$ and $\mathbf{Q'}_j$ denote the jth points in $\mathbf{Q}$ and $\mathbf{Q}'$ point sets, $\mathbf{Q}_\mu$ and $\mathbf{Q'}_\mu$ are the point sets' centroids, and $N = 60$ is the number of points in each set. Since $\mathbf{M}$ is a real square matrix with a positive determinant, it can be decomposed into orthogonal square matrices $\mathbf{U}$ and $\mathbf{V}$, and diagonal non-negative matrix $\mathbf{\Sigma}$, such that $\mathbf{M} = \mathbf{U\Sigma V}^T$ [82], where $\mathbf{\Sigma}$, and $\mathbf{U}$ and $\mathbf{V}^T$, are considered as scaling matrix and rotation matrices, respectively. Thus, $\mathbf{M}$ can be intuitively interpreted as a geometrical transformation composed of a rotation, a scaling, and another rotation. As the transformations need to be rigid, we omit $\mathbf{\Sigma}$ to preserve the objects' shape and size. Thus, the rotation matrix is computed as $\mathbf{R} = \mathbf{UV}^T$. Then, the translation matrix is determined as

$$\mathbf{T} = -\mathbf{R} \times \mathbf{Q}_\mu + \mathbf{Q'}_\mu. \tag{5.3}$$

5.2 Registration and Reconstruction

For a pair of temporal sequences of our dynamic object captured by our Kinects, e.g., of a human being breathing forcefully through a spirometer, we first found the corresponding frames using our intra-Kinect and inter-Kinect synchronisation approach. Then, the reference point cloud $\mathbf{P}_1^{ref}$ and the second point cloud $\mathbf{P}_2$, were generated. Using the computed translation and rotation matrices, we transformed $\mathbf{P}_2$ into the coordinate system of $\mathbf{P}_1^{ref}$, such that $\mathbf{P}'_2 = \mathbf{R} \times \mathbf{P}_2 + \mathbf{T}$. Finally, we created a merged point cloud,

$$\mathbf{P} = \mathbf{P}_1^{ref} \bigcup \mathbf{P}'_2. \tag{5.4}$$

The point cloud $\mathbf{P}$ is the final registered point cloud.

Figure 5.5 shows example scene point clouds, in purple and green, obtained by our two Kinects after aligning and registration. An enlarged side-view of the three calibration chessboards, held by 5mm-thick frames, is also shown, which presents the accuracy of our method in aligning the chessboards data points. Note that as long as the Kinects' position (fixed in calibration stage) remains unchanged, the computed rotation and translation matrices are valid for aligning the two point clouds.

Figure 5.5: *Aligned scene point clouds from the facing Kinects: the scene as viewed by each Kinect, after alignment. The zoomed-in side view shows the data points on the calibration chessboards.*

5.3 Experimental Results

5.3.1 Implementation Details

Two laptops with Intel® Core™ i7 quad core processors running at 2.8GHz and 16GB memory were used to acquire the data streams from our two Kinects, respectively. The proposed approach, comprising data acquisition, registration, reconstruction, and visualization were implemented in Microsoft Visual C++, using OpenCV [40] and Visualization Toolkit (VTK) [170] libraries, and Matlab 2015b. For concurrent processing, we used the Intel® Threading Building Blocks (TBB) library [147] to grab, buffer and record RGB, depth, infrared and body joint data in separate threads which enabled us to reconstruct a 3-D dynamic object at a consistent, full device frame rate of 30*fps*.

5.3.2 Registration Accuracy: Quantitative Evaluation

We measure the accuracy of our calibration method by computing the root mean square error (RMSE) of distances between corresponding chessboards' crossing points in $\mathbf{P}_1^{ref}$ and $\mathbf{P}_2'$. This error was computed as 4.6 mm for the lung function test setup ($\sim$3 m distance between Kinects at a height of $\sim$0.6 m).

We evaluate the calibration accuracy by estimating the dimensions of three boxes of known sizes placed at different depths and validating them against the ground truth values. Since more capturing space is required to be able to position the boxes in the scene, in this experiment the Kinects were placed at $\sim$4 m away from each other and at a height of $\sim$1.2 m. The RMSE of distances between corresponding crossing points in this setup was computed as 6.8 mm.

We evaluate the spatial registration accuracy of the proposed method by measuring dimensions and volumes, and performing surface analysis, of the 3 differently sized boxes. Figure 5.6 presents the 3-D models of these boxes reconstructed by our approach. Each box was captured three times, i.e., once at each of 3 different depths or locations, and all measurements made. Table 5.1 presents the 3 locations at which (the centroid of) each box was placed in the world coordinate system and the real dimensions of the 3 boxes. In each of the 9 captured sequences, the box was segmented from the registered point clouds by depth value thresholding. For each reconstructed box, sides planarity

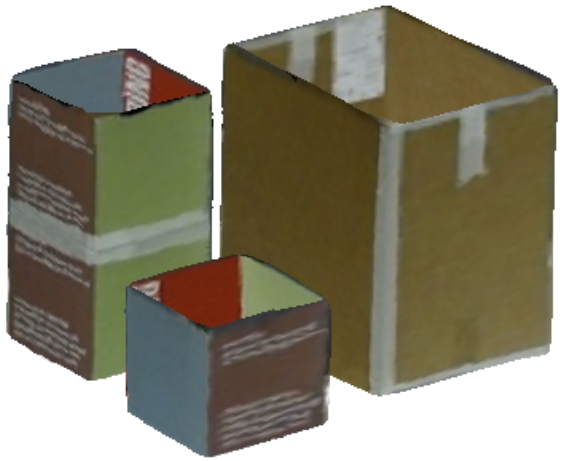


Figure 5.6: *3-D model of boxes used for quantitative evaluation of point clouds registration. The boxes 3-D model has been reconstructed using our proposed approach.*

Table 5.1: *Centroid location of the 3 boxes in the Kinects' joint coordinate system and their actual dimensions (all in cm).*

Box Location				Box Dimension			
Location	**X**	**Y**	**Z**	**Box**	**W**	**H**	**D**
Location 1	−17.7	18.7	241	***Box 1***	34.0	47.0	43.5
Location 2	−41.4	23.7	202	***Box 2***	23.2	45.0	23.2
Location 3	10.0	41.1	166	***Box 3***	23.2	22.5	23.2

and orthogonality, height, width, depth, and volume were automatically estimated by performing surface analysis, and then compared against ground truth measurements.

The boxes' four sides are automatically apportioned into separate point sets using the M-estimator SAmple Consensus (MSAC) approach [191]. Then, a plane is fitted on the point set of each side (see Figure 5.7) using a first degree polynomial, and R^2 and RMSE are computed for the fitted plane. The angles between the sides are estimated using the normal vectors of the fitted planes. Table 5.2 presents these estimations for *Box 1*, *Box 2* and *Box 3* in the 3 locations in which S1 to S4 refers to 4 sides of each box. The R^2 and RMSE values illustrate that the side-planarity is preserved very well in the reconstructed models. Furthermore, the estimated angles between the fitted planes show that our proposed system performs well in measuring orthogonality.

Table 5.2: *The surface analysis results of 3 boxes in 3 different locations computed using plane fitting. The RMSE and estimated angles are measured in cm and degree, respectively.*

	Box 1						***Box 2***						***Box 2***					
	Location 1		**Location 2**		**Location 3**		**Location 1**		**Location 2**		**Location 3**		**Location 1**		**Location 2**		**Location 3**	
Side	$\mathbf{R^2}$	**RMSE**	$\mathbf{R^2}$	**RMSE**	$\mathbf{R^2}$	**RMSE**	$\mathbf{R^2}$	**RMSE**	$\mathbf{R^2}$	**RMSE**	$\mathbf{R^2}$	**RMSE**	$\mathbf{R^2}$	**RMSE**	$\mathbf{R^2}$	**RMSE**	$\mathbf{R^2}$	**RMSE**
S1	0.997	0.003	0.997	0.004	0.997	0.003	0.994	0.004	0.997	0.002	0.993	0.004	0.997	0.003	0.998	0.002	0.996	0.003
S2	0.998	0.004	0.998	0.004	0.998	0.003	0.997	0.002	0.992	0.004	0.995	0.003	0.990	0.004	0.990	0.004	0.995	0.004
S3	0.996	0.004	0.997	0.003	0.997	0.003	0.995	0.003	0.993	0.005	0.993	0.004	0.995	0.003	0.996	0.003	0.997	0.002
S4	0.998	0.004	0.998	0.003	0.997	0.004	0.992	0.005	0.996	0.003	0.996	0.003	0.992	0.004	0.991	0.004	0.994	0.004
$\perp$	**Angle**		**Angle**		**Angle**		**Angle**		**Angle**		**Angle**		**Angle**		**Angle**		**Angle**	
S1–S2	88.89°		89.45°		89.17°		88.85°		89.96°		88.78°		88.65°		88.93°		89.19°	
S1–S4	89.41°		88.52°		88.42°		88.89°		89.02°		89.27°		88.45°		88.40°		89.22°	
S2–S3	89.19°		89.16°		89.94°		89.88°		89.26°		88.82°		88.99°		88.88°		88.64°	
S3–S4	89.10°		88.18°		90.83°		89.92°		89.81°		89.32°		91.20°		91.64°		88.67°	
S1–S3	0.59°		1.25°		1.08°		1.09°		0.78°		0.46°		0.51°		0.18°		0.71°	
S2–S4	2.35°		2.56°		1.90°		0.24°		1.37°		1.05°		0.37°		0.63°		0.54°	

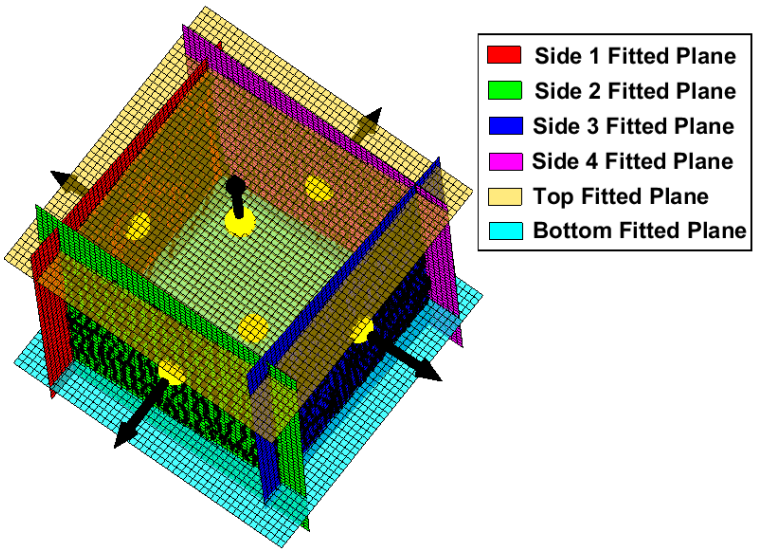


Figure 5.7: *Plane fitting of Box 3 sides.*

To help achieve the best estimation of height, width, and depth automatically, we use the planes fitted on the lateral sides of the boxes. Since the bottom of the boxes was not captured and the top was too sparse to rely on, the corresponding planes are computed by inference from the existing sides. First, the cross product of the normals of the lateral sides' fitted planes are computed to define two planes perpendicular to the side planes. These planes, which represent the bottom and top planes, are then placed respectively at the bottom-most and top-most of the box's point cloud, where there is a significant change in the number of points. Then, the eight corner points of the boxes are computed using the intersection of the fitted planes. Since the fitted planes are not exactly parallel, height, width, and depth are estimated by computing the average distance between the relevant four corner points of each side-plane and its facing side-plane.

Even though the box volume can be approximated using the estimated height, width, and depth ($V = W \times H \times D$), we estimate the volume by applying Gauss's Divergence Theorem as described in Section 4.1.3 of Chapter 4, since that would have to be used for geometrically non-uniform or non-rigid objects in any case. To be able to perform the surface integral over the box boundary, the box surface is reconstructed by applying a 2D Delaunay triangulation [57] on the registered point cloud. Note that dimensions and volume are presented in centimetres and litres, respectively.

Table 5.3 reports the estimated dimensions, volume and their L_2 error for *Box 1*, *Box 2* and *Box 3* against the ground truth at each of the 3 locations. We note that the extent of the error is a little different across height, width and depth, with an average of L_2

Table 5.3: *Automatically estimated width, height, depth (in cm) and volume (in litre) of 3 boxes using surface analysis.*

		Location 1		Location 2		Location 3	
		Estimated	**Error**	**Estimated**	**Error**	**Estimated**	**Error**
Box 1	**W**	34.3	0.3	34.2	0.2	34.3	0.3
	H	47.2	0.2	47.2	0.2	47.3	0.3
	D	44.4	0.9	44.4	0.9	44.2	0.7
	V	69.7 L	0.2 L	69.3 L	0.2 L	68.3 L	0.2 L
Box 2	**W**	24.0	0.8	23.7	0.5	23.8	0.6
	H	45.3	0.3	44.9	0.1	45.4	0.4
	D	24.0	0.8	23.9	0.7	24.1	0.9
	V	24.7 L	0.5 L	24.9 L	0.7 L	24.9 L	0.7 L
Box 3	**W**	23.9	0.7	23.8	0.6	23.9	0.7
	H	22.7	0.1	22.4	0.1	22.7	0.3
	D	24.2	1.0	24.1	0.9	23.9	0.7
	V	12.6 L	0.5 L	12.5 L	0.4 L	12.5 L	0.4 L

error for the 3 boxes in all locations across the height at 0.21, width at 0.54, and depth at 0.84 centimetre. Considering there is a $\sim 4m$ distance between the two Kinects, our results show very good accuracy for the estimated measurements, independent of the location of the boxes.

5.3.3 Dynamic Object Reconstruction: Qualitative Evaluation

We demonstrate the ability of the proposed method to achieve non-rigid 3-D object reconstruction with two different examples. The first is based on dynamic human trunk 3-D reconstruction for use in remote respiratory monitoring. The current remote respiratory sensing approaches, e.g., [58, 136, 137, 202, 203, 206], attempt to simulate traditional breathing tests, such as spirometry, however, none of these methods is able to decouple the subject's trunk motion from the subject's chest-wall surface motion as previously discussed in Chapter 2, which greatly affects the test results. Acquiring accurate and dynamic 3-D body shape data using our proposed method during the breathing test, can address this problem. In this test, the distance between each Kinect and the subject is $\sim 1.5m$ (optimal distance), at a height of $\sim 0.6m$, to be able to observe chest-wall motion as accurately as possible. Then, a 3-D surface of the subject's trunk performing real lung function assessment test, e.g., FVC, is reconstructed per frame. The analysis of such data demands precise point cloud alignment, accurate temporal frame synchronization,

body joints data acquisition (to estimate body pose), and consistent full frame rate data recording, all of which are provided by our system. Figure 5.8 shows sample 3-D reconstructed frames of a subject performing the FVC test. The reconstructions enable monitoring of the subject's trunk during the test. Although, the gap between the two aligned point clouds is not important in this application, it can be filled by interpolation.

The second example was performed to show accurate temporal and spatial point cloud alignment by way of the subject performing different actions, e.g., waving hands, dancing, and jumping. The two facing Kinects are placed $\sim 4m$ away from each other at a height of $\sim 1.2m$. Sample 3-D reconstruction of a subject waving hands in different frames are presented in Fig 5.9. As seen, the fingers have been well aligned and reconstructed. A sample video of the dynamic 3-D reconstruction is available at: https://www.youtube.com/watch?v=-JKHG3UJG9Q.

5.4 Conclusion

We proposed a 3-D RGB-D data acquisition pipeline which can provide accurate temporal and spatial 3-D reconstruction that can be used in applications such as remote respiratory monitoring and lung function assessment. The extrinsic parameters of the two facing Kinects are computed in a calibration stage, using three double-sided chessboards placed at varying depths. Then, these parameters are exploited to register point clouds and reconstruct 3-D, dynamic objects, for example performing lung function testing using a spirometer, and other actions such as waving. We evaluate the proposed system's accuracy by automatically measuring the dimensions, volume, and surface information of three different boxes and show that it is efficient in reconstructing the boxes and estimating their dimensions. Compared to the currently existing state-of-the-art dynamic 3-D data acquisition approaches, our proposed system only uses two sensors achieving frame-level reconstruction suitable for capturing fast and abrupt motions of dynamic objects. One shortcoming of our approach is that the current arrangement of our Kinects can result in missing information on parts of the object obscured from the Kinects' view (e.g., see the side of the person's trunk in Figure 5.8). However, The system has been designed such that it can be easily extended by more Kinects as long as each Kinect can see the three chessboards.

Figure 5.8: *Dynamic 3-D reconstruction of a subject performing lung function test while using a spirometer.*

Figure 5.9: *Dynamic 3-D reconstruction of a subject waving hands.*

Chapter 6

dPPG: Depth-based Whole-body Photoplethysmography in PFT

Our findings in Chapter 4 showed that when the respiratory motion is acquired using *a single* depth sensor, it can be easily affected by the subject's trunk motion during PFT. These motion artifacts not only provide unreliable measures, but also affect the volume–time data calibration due to the non-identical training and testing scaling factors. Filtering these motion artifacts under such circumstances is potentially impossible. Thus, in our single depth sensor approach (Chapter 4), we partially overcame this problem, by calibrating the *tidal volume* and *main effort* parts of the data separately using distinct *tidal volume* and *main effort* scaling factors.

Although subjects are asked to be completely still during the test, most of them inevitably move their trunk, especially during the *deep forced* inhalation–exhalation. Similar body motion artifacts have been also reported in [136, 137] where the main solution was to constrain the body movement, which is neither easy to achieve, nor particularly comfortable for patients. Further, this body movement is a natural reaction of the human respiratory system when required to maximally inhale and exhale, and restraining it can prevent subjects from performing their best breathing effort and would therefore affect their lung function measures.

In this chapter, we present our whole-body dPPG approach which allows subjects to perform the test as in routine spirometry without restraining their respiratory-related trunk reaction. Particularly, by decoupling trunk motion and chest-surface respiratory

motion, the proposed dPPG method reduces motion artifacts in volume–time data which consequently improves the accuracy of PFT measures.

Using our 3-D data acquisition and registration pipeline (Chapter 5), a dynamic 3-D model of the subject is reconstructed during the test. Then, a pair of depth variation time series automatically estimated from the subject's chest-wall and posterior-wall ROIs, are combined to retrieve the *depth-based* respiratory volume–time data, which is subsequently smoothed in a twofold filtering process (Section 6.1). Next, the training depth-based and spirometer volume–time data are analysed to extract keypoints and establish the intra-subject scaling factors (Section 6.2). Depth-based volume–time data of a test sequence is then calibrated using the learnt scaling factors from which 11 clinical FVC and SVC measures are computed (Section 6.3). We evaluate our dPPG method on the dual-Kinect PFT dataset containing 298 sequences from 35 subjects (Section 6.4). This method reduces the L_2 error mean and STD of $FEF_{50\%}$, $FEF_{75\%}$, $FEF_{25-75\%}$, IC, and ERV measures by half, compared to the single-Kinect approach. To investigate the accomplishment of the dPPG approach, we perform various comparative evaluations, i.e., Bland-Altman analysis, similarity measures performance analysis, intra-subject scaling factor error evaluation, and statistical analysis of within-subject *tidal volume* and *main effort* scaling factors, between the dPPG and the single-Kinect approaches, all of which show the superior accuracy of the dPPG. Finally, we conclude this chapter by summarising our approach and its achievements, and providing directions for future extensions (Section 6.5).

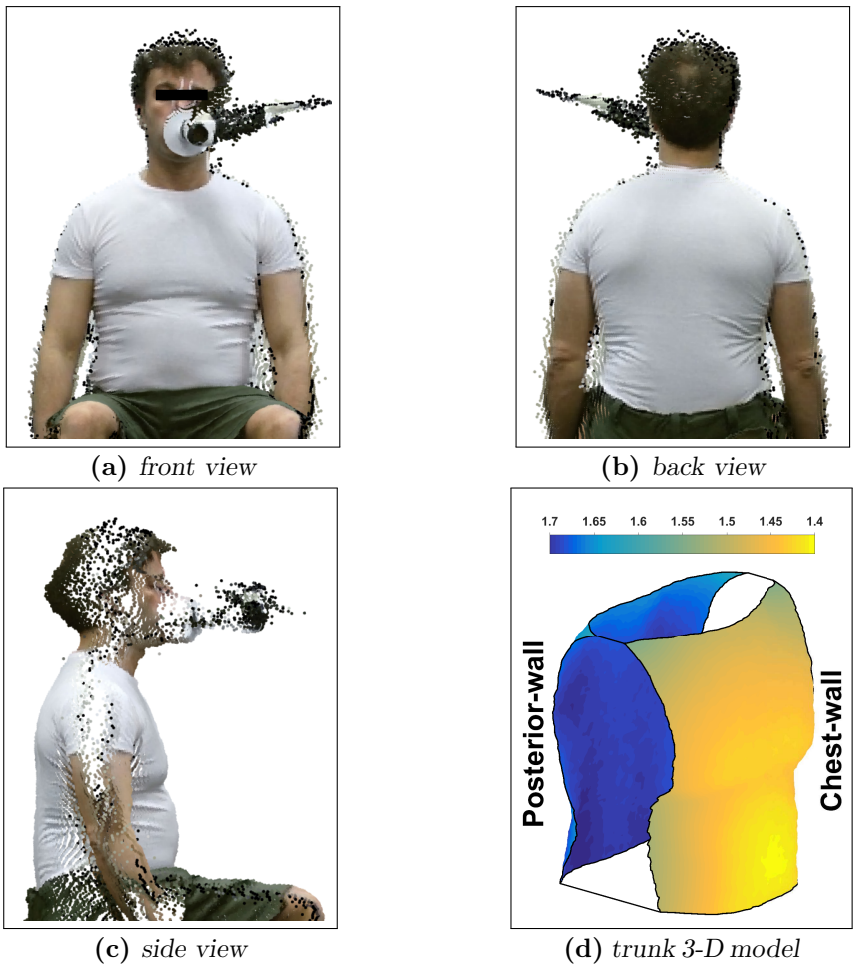
6.1 Depth-based Volume–time Data Retrieval

In order to compute PFT measures correctly, especially the timed measure *FEV1*, and the flow-based measures, *PEF* and $FEF_{25\%}$, $FEF_{50\%}$, $FEF_{75\%}$ and $FEF_{25-75\%}$, it is essential for the depth-based volume–time data to have a constant and high sampling rate. Since it is impossible to trigger multiple Kinects simultaneously, an exact frame level synchronisation between them cannot be achieved. Thus, the more Kinects used, the greater the temporal synchronization error would be.

As the only multiple-Kinect respiratory approach that this author is aware of, Harte et al. [88] estimated breathing time series from depth measurements acquired by four Kinects. They used off-the-shelf commercial 3-D software and plug-ins [6] for spatial calibration of the Kinects, trunk surface isolation, 3-D meshing and hole filling. Due

to their inability in precise inter-Kinect synchronisation, they reported temporal errors in the dynamic trunk reconstruction. According to the authors, these errors resulted in obtaining breathing time series with inconsistent sampling rates and sudden changes in the estimated volume.

Our dual-Kinect 3-D data acquisition and registration pipeline reconstructs an almost complete 3-D model of a subject performing the breathing test in a sitting position, at the Kinect's full frame rate (30 *fps*). Deploying only two sensors, (*a*) reduces the system setup and calibration effort, (*b*) keeps system costs low, (*c*) minimises the overall operation space, and most importantly, (*d*) minimises the temporal and spatial alignment errors. With this topology, there would be no overlapped views of the scene and the thoracoabdominal regions occluded by the arms are not considered pertinent to volume estimation accuracy. Figures 6.1a, 6.1b and 6.1c show an example 3-D reconstructed model of a subject's upper-body from different viewpoints during PFT performance.



(a) *front view* **(b)** *back view*

(c) *side view* **(d)** *trunk 3-D model*

Figure 6.1: *(a)–(c) 3-D reconstructed model of a subject performing PFT from different points of view. (d) 3-D reconstructed model of trunk for this subject.*

Inessential body parts are filtered using a chest-wall ROI mask which is automatically generated using skeleton joints similar to Section 4.1.2. Since the masking is performed on the whole reconstructed body in three dimensional space, the mask is transformed into the real world coordinate system using the camera parameters and rigid transformation matrices. Note that the skeleton joints data is acquired from the back Kinect as it is more stable than the data of the front Kinect as there is no occlusion of the face. Figure 6.1d shows the final 3-D reconstructed model of the subject's trunk after removing the head and limbs.

After registering the models of the chest and posterior walls into a joint real-world coordinate system for each frame of the sequence, their depth variations are computed as a pair of time series, i.e., $V_{ch}(t)$ and $V_{po}(t)$, using an averaging-based method. As an en bloc object, the subject's trunk movements are reflected on both the chest and the posterior walls, whereas the breathing motions mainly appear on the chest wall, with the posterior considerably less affected. Taking this into consideration, the trunk movements can be cancelled out by subtracting the motions of the chest and the posterior walls per frame, due to their similarity in direction and magnitude. However, this subtraction intensifies the breathing motions because expansion and contraction of the lungs move the chest and the posterior walls in nearly opposite directions. Thus, the final dual-Kinect dPPG volume–time data is computed as $V_{dk}(t) = [V_{po}(t) - V_{ch}(t)]$. For comparison, we define the single-Kinect volume–time data as $V_{sk}(t) = V_{ch}(t)$. Note that, $V_{po}(t)$ does not present any meaningful or useful information on its own. Figures 6.2a and 6.2b present

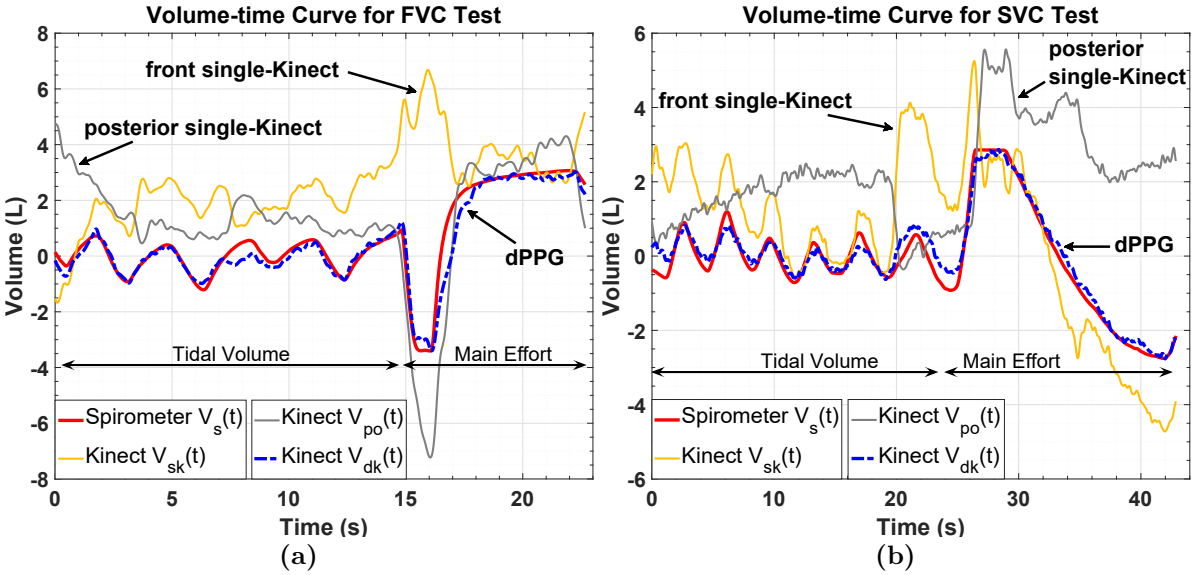


Figure 6.2: *Comparing volume–time data of the dPPG, i.e., $V_{dk}(t)$), front Kinect, i.e., $V_{sk}(t)$, and the back Kinect, i.e., $V_{po}(t)$, to the spirometer, i.e., $V_s(t)$.*

the retrieved $V_{dk}(t)$ and its corresponding $V_{sk}(t)$, and their comparison to the spirometer data $V_s(t)$, for FVC and SVC tests, respectively. As seen, the motion artifacts have been significantly reduced in $V_{dk}(t)$.

We improve the data filtering method in three ways compared to our previous work. Firstly, we choose not to apply a bilateral smoothing filter, as we noticed it eliminates subtle respiratory motions and affects the lung measures, especially the flow-based measures. Secondly, we realised that applying a moving-averaging filter over-smooths the *main effort* data and increases the error in the flow-based measures. Thus, here we use a 4th order Butterworth low-pass filter to smooth $V_{dk}(t)$ and $V_{sk}(t)$, similar to [137]. Thirdly, we perform twofold data filtering with two different cut-off frequencies, i.e., 1 Hz and 3 Hz. In the first data filtering stage, we choose the cut-off frequency as 1 Hz to

- identify the keypoints accurately,
- align the depth-based and spirometer volume–time data temporally, and
- segment them into *tidal volume* and *main effort* parts.

This frequency was chosen given the wide range of respiratory rates for adults and elderly at $12 - 36$ bpm ($0.2 - 0.6$ Hz) [55]. However, to avoid the respiratory data over-smoothing, especially at the *main effort* part where the curve slope is critical and needs to be preserved, we increase the cut-off frequency to 3 Hz and filter the original volume-time data for computing just PFT measures.

6.2 Volume–time Data Analysis

Since $V_{dk}(t)$ presents the subject's trunk volume variations rather than the real exchanged amount of air, it must be calibrated in order to obtain meaningful lung measures. This calibration is performed by linearly scaling the y-axis of the depth-based volume–time data using a scaling factor. Since scaling factors are subject-specific (intra-subject), they are learnt during a training phase for each subject by performing a linear regression analysis between the depth-based and spirometer training data. The main step towards this is to compute keypoints.

Keypoints computation – Multiple keypoints are automatically identified from the depth-based and spirometer volume–time data by performing an elaborate extrema anal-

ysis, as detailed in Chapter 4, using the same parameter and threshold values. We categorise these keypoints based on their application through as follows:

- Identifying *tidal volume* using $\{\mathcal{C}, \mathcal{D}\}$ and *main effort* using $\{\mathcal{E}, \mathcal{A}, \mathcal{B}\}$.
- Computing *main effort* scaling factors using $\{\mathcal{A}, \mathcal{B}\}$.
- Computing *tidal volume* and *main effort* similarity measures using $\{\mathcal{A}, \mathcal{B}\}$ and $\{\mathcal{F}_i, \mathcal{G}_i\}_{i=1}^{4}$.
- Computing PFT measures using $\{\mathcal{E}, \mathcal{A}, \mathcal{B}\}$, $\{\mathcal{F}_i, \mathcal{G}_i\}_{i=1}^{4}$, 'time zero' t_0 and 'Peak Flow' t_{PF}.

Figures 6.3a and 6.3b illustrate the computed keypoints for $V_s(t)$, $V_{dk}(t)$ and $V_{sk}(t)$ of FVC and SVC tests, respectively. As shown, all keypoints are computed correctly for $V_{dk}(t)$ and match their corresponding ones in the spirometer $V_s(t)$. However, for $V_{sk}(t)$, several keypoints, i.e., $\{\mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}\}$ and $\{\mathcal{F}_i, \mathcal{G}_i\}_{i=1}^{4}$, (labelled in red in Figures 6.3a and 6.3b), are computed incorrectly due to the effects of the subject's body movement on $V_{sk}(t)$. For example, in Figure 6.3b keypoint $\mathcal{B}$ is computed incorrectly resulted in an incorrect calibration of $V_{sk}(t)$, whilst $V_{dk}(t)$ is calibrated quite precisely for the same sequence.

Linear regression analysis – Linear regression is performed separately for *tidal volume* and *main effort* parts of depth-based and spirometer training volume–time data, and provides individual *tidal volume* and *main effort* scaling factors. In order to perform the linear regression, the depth-based and spirometer corresponding data samples are identified. For this, the spirometer data is sampled at the Kinect sampling rate of 30 Hz, and the depth-based and spirometer *tidal volume* are separated using $\{\mathcal{C}, \mathcal{D}\}$ keypoints. They are then detrended (see the trend in $V_{sk}(t)$ in Figures 6.3a and 6.3b) by applying EMD [93] to increase their similarity and attain better temporal alignment. Finally, the delay is computed using a windowed cross correlation and used to temporally align the whole depth-based and spirometer volume–time data. This process is carried out for $V_{dk}(t)$ and $V_{sk}(t)$ separately.

The *tidal volume* and *main effort* scaling factors are computed by establishing linear regression individually for *tidal volume* as

$$\widehat{V}_s^{tv} = \xi_{dk}^{tv} \cdot \widehat{V}_{dk}^{tv} + \psi_{dk}^{tv}, \quad \widehat{V}_{dk}^{tv} = \widehat{V}_{dk}(t)\Big|_{t_{\mathcal{C}}}^{t_{\mathcal{D}}}, \tag{6.1}$$

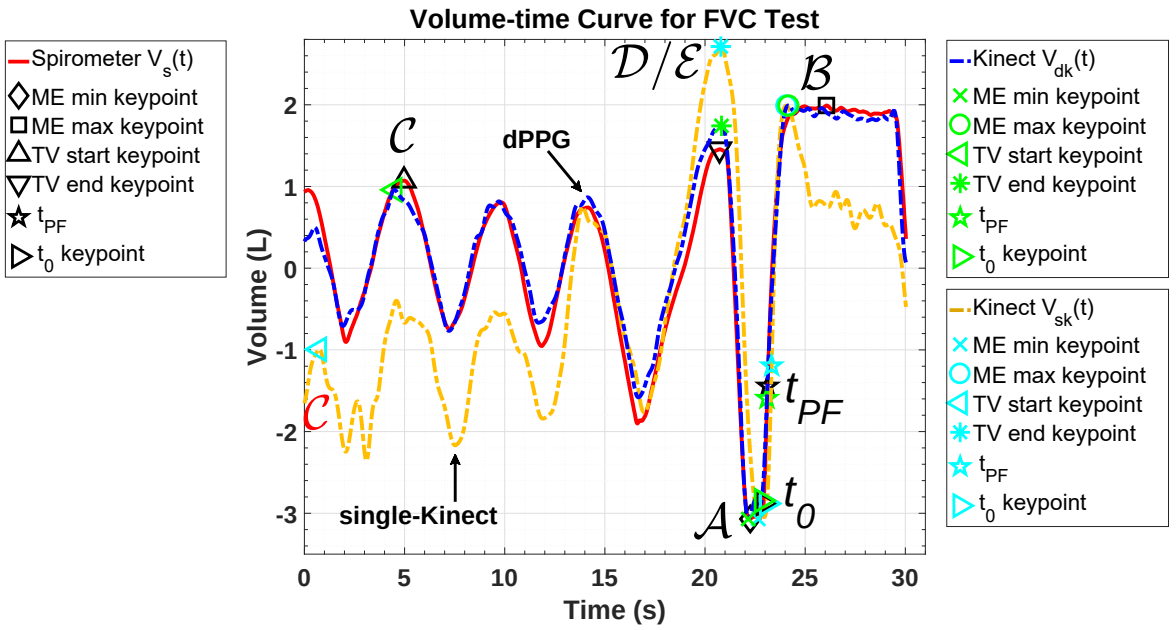


Figure 6.3: *Comparing the single-Kinect and dPPG volume–time data, i.e., $V_{sk}(t)$ and $V_{dk}(t)$, to the spirometer $V_s(t)$. Red labelled keypoints are incorrectly computed in $V_{sk}(t)$. $V_{sk}(t)$ in (b) is incorrectly calibrated due to incorrect computation of keypoints caused by the trunk motion.*

and for *main effort* as

$$\left\langle \widehat{V}_s(t_{\mathcal{A}}), \widehat{V}_s(t_{\mathcal{B}}) \right\rangle = \xi_{dk}^{me} \cdot \left\langle \widehat{V}_{dk}(t_{\mathcal{A}}), \widehat{V}_{dk}(t_{\mathcal{B}}) \right\rangle + \psi_{dk}^{me}, \tag{6.2}$$

where $\widehat{V}_{dk}(t)$ and $\widehat{V}_s(t)$ are detrended and zero-mean normalised data from the Kinect

and spirometer. Since the data is normalised to zero-mean of their *tidal volume*, then $\psi_{dk}^{tv} \approx 0$. Thus, the *tidal volume* and *main effort* scaling factors are defined as $\langle \xi_{dk}^{tv} \rangle$ and $\langle \xi_{dk}^{me}, \psi_{dk}^{me} \rangle$, respectively. Similarly, the *tidal volume* and *main effort* scaling factors for $V_{sk}(t)$, i.e., $\langle \xi_{sk}^{tv} \rangle$ and $\langle \xi_{sk}^{me}, \psi_{sk}^{me} \rangle$, are computed for comparative analysis.

6.3 Scaling Factors Generalisation

The aim of this study is to assess human lung function remotely and independently, without support from any other clinical device, e.g., a spirometer. The coefficients of the linear regression, i.e., the scaling factors, between the trunk volume and lungs air flow, are subject-specific and depend on physical body specifications, e.g., weight, height, BMI, gender and race. Thus, we train our system to learn the scaling factors per subject, which enables it to perform PFT independent of a spirometer at later trials.

In the training phase, intra-subject scaling factors are learnt using training trials, and computed as $\left\{ \langle \xi_{dk}^{tv} \rangle^{\ell} \right\}_{\ell=1}^{n_{tv}}$ & $\left\{ \langle \xi_{dk}^{me}, \psi_{dk}^{me} \rangle^{\ell} \right\}_{\ell=1}^{n_{me}}$ for $\left\{ \left\langle V_{dk}(t), V_s(t) \right\rangle^{\ell} \right\}_{\ell=1}^{n_T}$ as explained in Section 6.2, where n_{tv} and n_{me} are the number of *tidal volume* and *main effort* training trials, and $n_T = n_{tv} + n_{me}$.

In the testing phase, first, the volume–time data of a test trial, i.e., $V_{dk}^{test}(t)$, is retrieved using the method explained in Section 6.1. Then, *tidal volume* and *main effort* similarity measures are computed as F_{tv}^{test} & F_{me}^{test} and $\left\{ F_{tv}^{\ell} \right\}_{\ell=1}^{n_{tv}}$ & $\left\{ F_{me}^{\ell} \right\}_{\ell=1}^{n_{me}}$ using (4.21) and (4.22), for $V_{dk}^{test}(t)$ and $\left\{ \left\langle V_{dk}(t) \right\rangle^{\ell} \right\}_{\ell=1}^{n_T}$, respectively. These allow for optimisation of *tidal volume* and *main effort* training trials by matching training similarity measures with the similarity measures of $V_{dk}^{test}(t)$ as

$$\jmath_{tv} = \underset{j \in [1..n_{tv}]}{\arg\min} \left\{ \left| F_{tv}^{test} - F_{tv}^{j} \right| \right\}, \tag{6.3}$$

$$\jmath_{me} = \underset{j \in [1..n_{me}]}{\arg\min} \left\{ \left| F_{me}^{test} - F_{me}^{j} \right| \right\}. \tag{6.4}$$

The associated scaling factors of $\jmath_{tv}$ and $\jmath_{me}$ trials, declared as $\langle \xi_{dk}^{tv} \rangle^{\jmath_{tv}}$ and $\langle \xi_{dk}^{me}, \psi_{dk}^{me} \rangle^{\jmath_{me}}$ are then used to calibrate $V_{dk}^{test}(t)$ as

$$V_{dk}^{cal}(t) = \left[V_{dk}^{test}(t) \cdot \langle \xi_{dk}^{tv} \rangle^{\jmath_{tv}} \right]_{t=t_{\mathcal{C}}}^{t=t_{\mathcal{D}}} + \left[V_{dk}^{test}(t) \cdot \langle \xi_{dk}^{me} \rangle^{\jmath_{me}} + \langle \psi_{dk}^{me} \rangle^{\jmath_{me}} \right]_{t=t_{\mathcal{D}}}^{t=max(t_{\mathcal{A}}, t_{\mathcal{B}})}. \tag{6.5}$$

In order to compare our method to the single-Kinect approach, a similar process is carried out to obtain $\langle \xi_{sk}^{tv} \rangle^{\jmath'_{tv}}$ and $\langle \xi_{sk}^{me}, \psi_{sk}^{me} \rangle^{\jmath'_{me}}$, and calibrate $V_{sk}^{test}(t)$, where $\jmath'_{tv}$ and $\jmath'_{me}$ are the optimised *tidal volume* and *main effort* selected trials.

The intra-subject training and testing process is performed using leave-one-out cross-validation due to the limited number of trials for each subject.

6.4 Experimental Results

We evaluate the dPPG method on our dual-Kinect PFT dataset previously introduced in Chapter 3. This dataset contains 300 sequences (155 FVC and 145 SVC) collected from 35 subjects. From the 300 sequences, we drop only 2 sequences as our method fails to compute their volume–time data keypoints due to complex body motion patterns. Otherwise, the data of all the other 298 sequences are successfully retrieved and their measures are computed and considered in the experimental analysis.

We demonstrate the accomplishments of our dPPG approach by,

- achieving significant improvements in FVC and SVC measures compared to the single-Kinect approach,
- improving volume–time data calibration accuracy by computing more accurate similarity measures and reducing intra-subject scaling factor learning error,
- computing more consistent and stable *tidal volume* and *main effort* scaling factors, which increases the depth-based PFT measures' reproducibility, and
- achieving higher correlation between *tidal volume* and *main effort* scaling factors confirmed by performing a comparative statistical analysis across 35 subjects.

6.4.1 Evaluation of PFT Measures

Using the standard formulation explained in Chapter 4, seven FVC and four SVC measures are computed from $V_{dk}(t)$, and also $V_{sk}(t)$ for comparison. Tables 6.1 and 6.2 present the results of these measures for all 35 subjects, computed for 155 FVC and 143 SVC sequences respectively, from $V_{dk}(t)$ and $V_{sk}(t)$. These tables report, (i) mean

Table 6.1: *Intra-subject mean ($\boldsymbol{\mu_{dk}}$ & $\boldsymbol{\mu_{sk}}$) and STD ($\boldsymbol{\sigma_{dk}}$ & $\boldsymbol{\sigma_{sk}}$) of L_2 error, and ratio of each measure's L_2 error to the mean value of that measure ($\boldsymbol{\Omega_{dk}}$ & $\boldsymbol{\Omega_{sk}}$) as well as correlation coefficients ($\boldsymbol{\lambda_{dk}}$ & $\boldsymbol{\lambda_{sk}}$), for the dPPG and single-Kinect against the spirometer FVC measures.*

		FVC (L)	***FEV1*** (L)	***PEF*** (L/S)	***FEF$_{25\%}$*** (L/S)	***FEF$_{50\%}$*** (L/S)	***FEF$_{75\%}$*** (L/S)	***FEF$_{25\text{-}75\%}$*** (L/S)
dPPG	$\boldsymbol{\mu_{dk}}$	0.36	0.55	1.83	2.02	**1.55**	**0.85**	**1.28**
	$\boldsymbol{\sigma_{dk}}$	0.39	0.55	1.59	1.92	**1.28**	**0.86**	**1.04**
	$\boldsymbol{\Omega_{dk}}$	0.07	0.13	0.22	0.26	**0.32**	**0.39**	**0.31**
	$\boldsymbol{\lambda_{dk}}$	0.87	0.68	0.46	0.35	**0.34**	**0.47**	**0.43**
single-Kinect	$\boldsymbol{\mu_{sk}}$	0.45	0.68	2.32	2.48	3.29	2.60	2.67
	$\boldsymbol{\sigma_{sk}}$	0.40	0.49	2.30	2.45	3.17	2.97	2.48
	$\boldsymbol{\Omega_{sk}}$	0.09	0.17	0.28	0.32	0.68	1.22	0.64
	$\boldsymbol{\lambda_{sk}}$	0.86	0.67	0.41	0.22	0.09	0.15	0.19

Table 6.2: *Intra-subject mean ($\boldsymbol{\mu_{dk}}$ & $\boldsymbol{\mu_{sk}}$) and STD ($\boldsymbol{\sigma_{dk}}$ & $\boldsymbol{\sigma_{sk}}$) of L_2 error, and ratio of each measure's L_2 error to the mean value of that measure ($\boldsymbol{\Omega_{dk}}$ & $\boldsymbol{\Omega_{sk}}$) as well as correlation coefficients ($\boldsymbol{\lambda_{dk}}$ & $\boldsymbol{\lambda_{sk}}$), for the dPPG and single-Kinect against the spirometer SVC measures.*

		VC (L)	***IC*** (L)	***TV*** (L)	***ERV*** (L)
dPPG	$\boldsymbol{\mu_{dk}}$	0.421	**0.393**	0.148	**0.315**
	$\boldsymbol{\sigma_{dk}}$	0.474	**0.409**	0.150	**0.331**
	$\boldsymbol{\Omega_{dk}}$	0.095	**0.127**	0.160	**0.239**
	$\boldsymbol{\lambda_{dk}}$	0.869	**0.810**	0.839	**0.748**
single-Kinect	$\boldsymbol{\mu_{sk}}$	0.550	0.730	0.181	0.778
	$\boldsymbol{\sigma_{sk}}$	0.656	0.712	0.202	0.812
	$\boldsymbol{\Omega_{sk}}$	0.124	0.235	0.196	0.592
	$\boldsymbol{\lambda_{sk}}$	0.719	0.466	0.767	0.389

(μ_{dk} & μ_{sk}) and STD (σ_{dk} & σ_{sk}) of L_2 error for each measure, (*ii*) ratio of mean of L_2 error to the mean value of that measure (Ω_{dk} & Ω_{sk}), and (*iii*) correlation coefficients (λ_{dk} & λ_{sk}) for the dPPG & single-Kinect against the spirometer measures.

As can be seen in Table 6.1, ($\mu_{dk}, \sigma_{dk}, \Omega_{dk}$) are lower for the dPPG approach across all measures, compared to their single-Kinect counterparts ($\mu_{sk}, \sigma_{sk}, \Omega_{sk}$). In particular, these errors are reduced by half for $FEF_{50\%}$, $FEF_{75\%}$ and $FEF_{25-75\%}$ measures. This remarkable error reduction is due to the measures being computed using the top curvature of the *main effort* data, which is successfully recovered in $V_{dk}(t)$ by decoupling trunk movements from the respiratory motion (compare in Figure 6.3a against $V_{sk}(t)$).

For the other measures reported in Table 6.1, ($\mu_{dk}, \sigma_{dk}, \Omega_{dk}$) are not decreased signif-

icantly compared to their single-Kinect counterparts. For *FVC*, this is because the measure is computed using the same keypoints $\mathcal{A}$ and $\mathcal{B}$, which are also used in the *main effort* calibration. *FEV1*, *PEF* and $FEF_{25\%}$ measures are computed from the steepest part of the *main effort*, between keypoints t_0 and $t_{FEF25\%}$. Thus, we believe the trunk forward movement at the start of forceful exhalation increases the *main effort* curve slope and accidentally contributes in achieving better *FEV1*, *PEF* and $FEF_{25\%}$ measures.

These results confirm the superiority of the dPPG method to the single-Kinect approach, with λ_{dk} also showing better correlation of measures than λ_{sk}. However, λ_{dk} does not express strong correlation between the dPPG FVC measures and the spirometer, except for *FVC* and *FEV1*. This is expected as we exploit all of the acquired data and do not remove the trials that impose greater error. In particular, these trials appear as outliers and influence the correlation coefficients. To further clarify this issue, we perform a Bland-Altman analysis of measures (Section 6.4.2) and present more qualitative and quantitative comparison between the dPPG and single-Kinect methods against the spirometer measures.

Ostadabbas et al. [136] reported a 0.88 average correlation with a spirometer for *FEV1* (and no other measures). However, this cannot be directly compared to the *FEV1* correlation coefficient computed here which is on a different dataset, acquired by different protocols, under different criteria.

Table 6.2 reports the evaluation results for SVC measures, in which $(\mu_{dk}, \sigma_{dk}, \Omega_{dk})$ are also reduced by half for *IC* and *ERV* measures, compared to $(\mu_{sk}, \sigma_{sk}, \Omega_{sk})$. Moreover, λ_{dk} shows much better correlation for these two measures, compared to λ_{sk}. The improved results are due to the trunk motion corrections, which remove the offset between the *tidal volume* and the *main effort*. The *VC* measure is computed using the keypoints $\mathcal{A}$ and $\mathcal{B}$, which are also exploited for calibrating SVC volume–time data, thus the dPPG method achieves only a slight improvement in this measure. *TV* is also slightly improved as subjects' movements in the 'rest' condition is insignificant.

PFT measures' correlation coefficient and error, reported in Chapter 4, are relatively better than the results reported here because they are evaluated on the single-Kinect dataset in which the subject's trunk motion were strictly restrained during the test. Tables 6.1 and 6.2 report the results of applying the same single-Kinect method of Chapter 4 on the dual-Kinect PFT dataset in which subjects performed PFT as in routine spirometry and their body's normal reaction to deep and forced inhalation–exhalation was not restricted. Comparing the evaluation results of the dPPG and single Kinect approaches

on the dual-Kinect dataset (Tables 6.1 and 6.2), confirms that eliminating trunk motion, achieved by the dPPG approach, highly improves the measures' correlation and reduces the error, even when both approaches use the same volume–time data analysis and intra-subject scaling factor learning methods.

6.4.2 Bland-Altman Analysis of PFT Measures

Tables 6.3 and 6.4 report Bland-Altaman [37] range of agreement between the dPPG and the spirometer measures, i.e., $\mathcal{L}_{dk}-\mathcal{U}_{dk}$, and also between the single-Kinect and the spirometer measures, i.e., $\mathcal{L}_{sk}-\mathcal{U}_{sk}$, where $\mathcal{L}_{dk}\,\&\,\mathcal{U}_{dk}$ and $\mathcal{L}_{sk}\,\&\,\mathcal{U}_{sk}$ indicate the *lower & upper limits of agreement* for the dPPG and the single-Kinect measures, respectively. Results confirms that the dPPG better agree with the spirometer across all the measures, particularly for $FEF_{50\%}$, $FEF_{75\%}$, $FEF_{25-75\%}$, *IC* and *ERV*.

Further, in order to better compare the error between the dPPG and the single-Kinect measures, $\mathcal{M}_{dk}$ is computed as the percentage of trials where the difference between the dPPG and the spirometer measure lies in $\mathcal{L}_{dk}-\mathcal{U}_{dk}$. Similarly, $\mathcal{M}_{sk}$ specifies the percentage of trials in the same range of agreement between the single-Kinect and the spirometer measure (see Tables 6.3 and 6.4). Although $\mathcal{M}_{dk}$ is greater than $\mathcal{M}_{sk}$ across

Table 6.3: *Bland-Altman analysis of the dPPG and the single-Kinect FVC measures.*

Metric	FVC Measures						
	FVC (L)	***FEV1*** (L)	***PEF*** (L/S)	$\boldsymbol{FEF_{25\%}}$ (L/S)	$\boldsymbol{FEF_{50\%}}$ (L/S)	$\boldsymbol{FEF_{75\%}}$ (L/S)	$\boldsymbol{FEF_{25\text{-}75\%}}$ (L/S)
$\mathcal{L}_{dk}-\mathcal{U}_{dk}$	-0.9–1.1	-1.0–1.7	-3.5–5.3	-3.9–6.1	**-3.4–4.2**	**-2.1–2.5**	**-2.2–3.6**
$\mathcal{L}_{sk}-\mathcal{U}_{sk}$	-1.1–1.2	-1.6–1.7	-7.0–5.5	-6.0–7.4	**-10–6.0**	**-8.6–5.5**	**-7.9–5.7**
$\mathcal{M}_{dk}$ (%)	98.0%	95.5%	94.2%	93.5%	**95.5%**	**94.2%**	**92.9%**
$\mathcal{M}_{sk}$ (%)	92.2%	92.2%	83.2%	85.1%	**70.3%**	**67.7%**	**63.8%**

Table 6.4: *Bland-Altman analysis of the dPPG and the single-Kinect SVC measures.*

Metric	SVC Measures			
	VC (L)	***IC*** (L)	***TV*** (L)	***ERV*** (L)
$\mathcal{L}_{dk}-\mathcal{U}_{dk}$	-1.4–0.8	**-1.2–0.7**	-0.4–0.4	**-0.9–0.8**
$\mathcal{L}_{sk}-\mathcal{U}_{sk}$	-1.8–1.5	**-2.1–1.9**	-0.4–0.6	**-2.2–2.1**
$\mathcal{M}_{dk}$ (%)	95.1%	**95.8%**	93.0%	**95.1%**
$\mathcal{M}_{sk}$ (%)	90.2%	**76.2%**	88.1%	**64.3%**

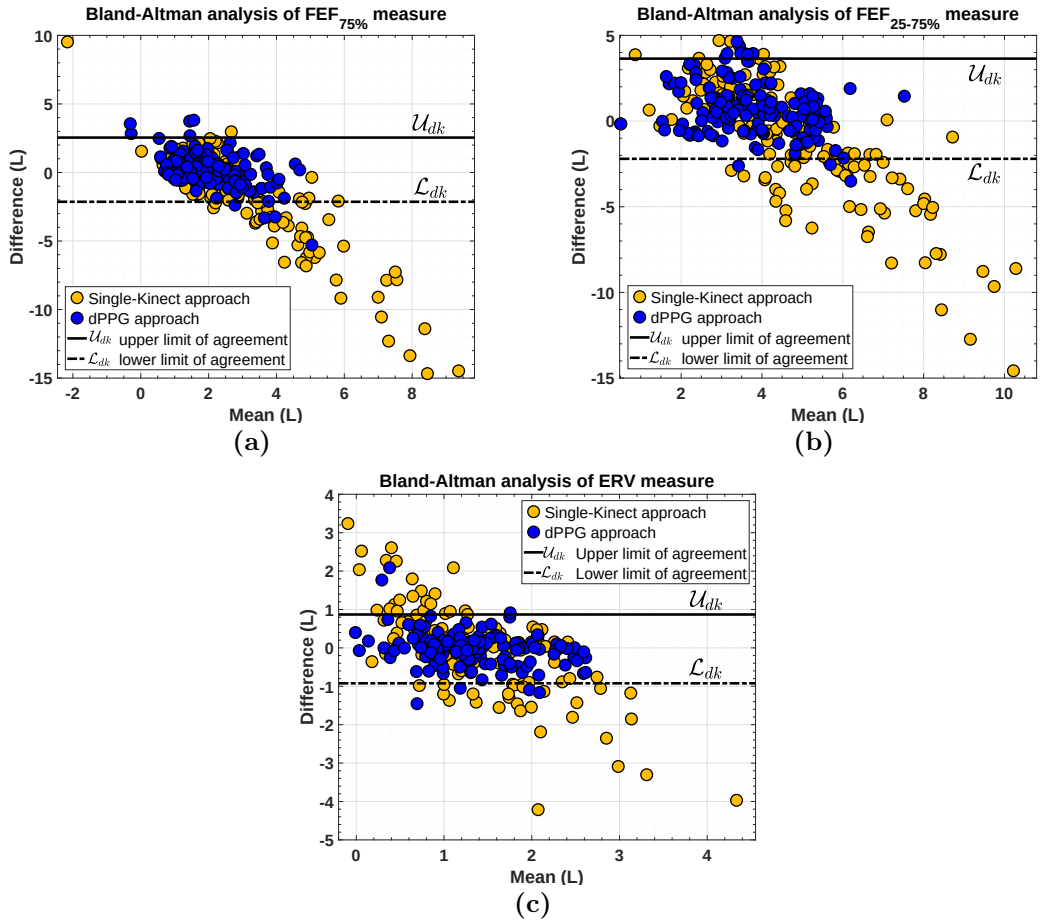


Figure 6.4: *Bland-Altman plots for (a) $FEF_{75\%}$, (b) $FEF_{25-75\%}$ and (c) ERV.*

all measures, the difference between $\mathcal{M}_{dk}$ and $\mathcal{M}_{sk}$ is more distinguishable for $FEF_{50\%}$, $FEF_{75\%}$, $FEF_{25-75\%}$, IC and ERV. Figure 6.4 shows Bland-Altman plots of $FEF_{75\%}$, $FEF_{25-75\%}$ and ERV measures for the dPPG and the single-Kinect approaches. Many of the single-Kinect measures lie outside of the dPPG lower and upper limits of agreement, i.e., $\mathcal{L}_{dk}$ and $\mathcal{U}_{dk}$.

6.4.3 Performance Evaluation of Similarity Measures

We evaluate the performance of the *tidal volume* and *main effort* similarity measures, in terms of their ability to choose the intra-subject scaling factors $\langle \xi_{dk}^{tv} \rangle^{\jmath_{tv}}$ & $\langle \xi_{dk}^{me} \rangle^{\jmath_{me}}$, which are supposed to calibrate the test volume–time data with the minimum error, among the training scaling factors $\left\{ \langle \xi_{dk}^{tv} \rangle^{\ell} \right\}_{\ell=1}^{n_{tv}}$ & $\left\{ \langle \xi_{dk}^{me} \rangle^{\ell} \right\}_{\ell=1}^{n_{me}}$. Thus, we use normalised L_2 error $\texttt{SME}_{dk}^{tv}$ & $\texttt{SME}_{dk}^{me}$, computed as the ratio of L_2 error between $\langle \xi_{dk}^{tv} \rangle^{\jmath_{tv}}$ & $\langle \xi_{dk}^{me} \rangle^{\jmath_{me}}$ and $\langle \xi_{dk}^{tv} \rangle^{c}$ & $\langle \xi_{dk}^{me} \rangle^{c}$, to $\langle \xi_{dk}^{tv} \rangle^{c}$ & $\langle \xi_{dk}^{me} \rangle^{c}$. $\langle \xi_{dk}^{tv} \rangle^{c}$ & $\langle \xi_{dk}^{me} \rangle^{c}$ are the numerically closest scaling

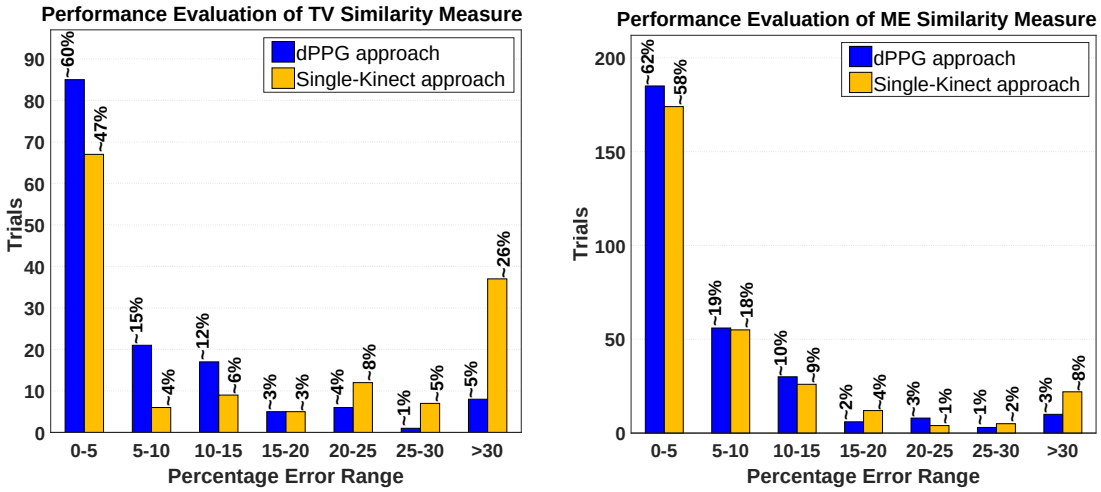


Figure 6.5: *Performance evaluation of similarity measures by distributing (a) 143 tidal volume trials, and (b) 298 main effort trials over* $\texttt{SME}^{tv}_{dk}$ & $\texttt{SME}^{tv}_{sk}$ *and* $\texttt{SME}^{me}_{dk}$ & $\texttt{SME}^{me}_{sk}$ *at various intervals for the dPPG (blue) and the single-Kinect (orange) approaches.*

factors to the spirometer scaling factors of the test trial, i.e., $\langle \xi^{tv}_{dk} \rangle^{o}$ & $\langle \xi^{me}_{dk} \rangle^{o}$. Note that, $\langle \xi^{tv}_{dk} \rangle^{o}$ & $\langle \xi^{me}_{dk} \rangle^{o}$ are computed using the spirometer volume–time data of the test trial and are only used for evaluation and comparison. Similarly, $\texttt{SME}^{tv}_{sk}$ and $\texttt{SME}^{me}_{sk}$ are also computed for the single-Kinect approach. Figures 6.5a and 6.5b show the distribution of *tidal volume* and *main effort* trials over the computed error $\texttt{SME}^{tv}_{dk}$ & $\texttt{SME}^{tv}_{sk}$ and $\texttt{SME}^{me}_{dk}$ & $\texttt{SME}^{me}_{sk}$ for the dPPG (blue) and the single-Kinect (orange) approaches, respectively, in the range $0 - 30\%$ at 5% intervals and then for more than 30%. As can be seen, ∼75% of *tidal volume* trials and ∼81% of *main effort* trials in the dPPG approach have $< 10\%$ error. This reduces to ∼51% and ∼76% in the single-Kinect approach. Also, many fewer trials with $> 30\%$ error occur in the dPPG approach, i.e., ∼5%, as opposed to ∼26% in the single-Kinect method.

6.4.4 Error Analysis of Intra-subject Scaling Factors

We obtain the spirometer scaling factors $\langle \xi^{tv}_{dk} \rangle^{o}$ & $\langle \xi^{me}_{dk} \rangle^{o}$ to assist us in evaluating our intra-subject scaling factors by computing the normalised L_2 error, i.e., $\texttt{SCE}^{tv}_{dk}$ and $\texttt{SCE}^{me}_{dk}$, between $\langle \xi^{tv}_{dk} \rangle^{\jmath_{tv}}$ & $\langle \xi^{me}_{dk} \rangle^{\jmath_{me}}$ and $\langle \xi^{tv}_{dk} \rangle^{o}$ & $\langle \xi^{me}_{dk} \rangle^{o}$. We also compare against the single-Kinect approach by computing $\texttt{SCE}^{tv}_{sk}$ and $\texttt{SCE}^{me}_{sk}$. Figures 6.6a and 6.6b present the distribution of *tidal volume* and *main effort* trials over the intra-subject scaling factor errors $\texttt{SCE}^{tv}_{dk}$ & $\texttt{SCE}^{tv}_{sk}$ and $\texttt{SCE}^{me}_{dk}$ & $\texttt{SCE}^{me}_{sk}$ for the dPPG (blue) and the single-Kinect (orange)

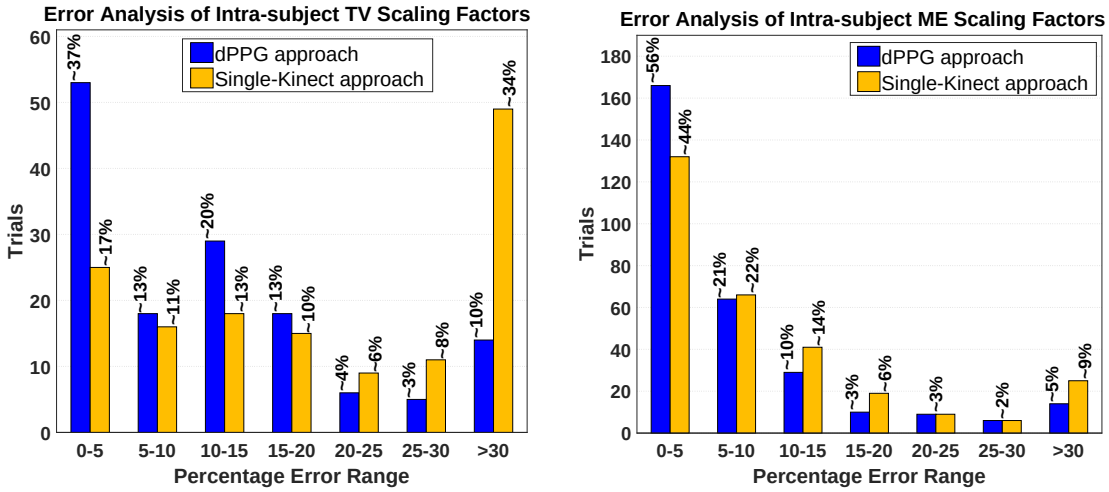


Figure 6.6: *Error analysis of intra-subject scaling factors by distributing (a) 143 tidal volume trials, and (b) 298 main effort trials over* $\texttt{SCE}^{tv}_{dk}$ & $\texttt{SCE}^{tv}_{sk}$ *and* $\texttt{SCE}^{me}_{dk}$ & $\texttt{SCE}^{me}_{sk}$ *at various intervals for the dPPG (blue) and the single-Kinect (orange) approaches.*

approaches, respectively. For example, in Figure 6.6a, ∼50% of *tidal volume* trials have $< 10\%$ error in the dPPG approach against ∼28% in the single-Kinect approach. Also, only ∼10% of the *tidal volume* trials have $> 30\%$ error for the dPPG against ∼34% in the single-Kinect method. In the *main effort* trials, the dPPG approach similarly performs better (see Figure 6.6b).

6.4.5 Statistical Analysis of Within-subject Scaling Factors

Table 6.5 reports the mean and STD of within-subject *tidal volume* and *main effort* scaling factors, for the dPPG and single-Kinect approaches for all 35 participants, denoted as $\texttt{M}^{tv}_{dk}, \texttt{M}^{me}_{dk}$ & $\Sigma^{tv}_{dk}, \Sigma^{me}_{dk}$ and $\texttt{M}^{tv}_{sk}, \texttt{M}^{me}_{sk}$ & $\Sigma^{tv}_{sk}, \Sigma^{me}_{sk}$, respectively. Minimum to maximum range of scaling factors and their distribution between the 1st and 3rd quartiles along with the outliers and median are presented in Figure 6.7.

The comparison between the scaling factors' STD, i.e., $\Sigma^{tv}_{dk}, \Sigma^{me}_{dk}$ versus $\Sigma^{tv}_{sk}, \Sigma^{me}_{sk}$ in Table 6.5, shows that the dPPG within-subject scaling factors are more consistent than the single-Kinect method, especially for the *tidal volume*. This can be better realised by comparing the minimum to maximum range of the scaling factors, and also their interquartile ranges in Figure 6.7, for the dPPG (blue boxes) and the single-Kinect (orange boxes) approaches.

Table 6.5: *Mean and STD of within-subject tidal volume and main effort scaling factors for the dPPG approach, i.e.,* $\mathtt{M}^{tv}_{dk}, \mathtt{M}^{me}_{dk}$ *&* $\Sigma^{tv}_{dk}, \Sigma^{me}_{dk}$*, and the single-Kinect approach, i.e.,* $\mathtt{M}^{tv}_{sk}, \mathtt{M}^{me}_{sk}$ *&* $\Sigma^{tv}_{sk}, \Sigma^{me}_{sk}$*. Blue numbers highlight some samples of the best agreement between* $\mathtt{M}^{tv}_{dk}$ *and* $\mathtt{M}^{me}_{dk}$ *whereas their corresponding* ***values*** *in the single-Kinect approach, i.e.,* $\mathtt{M}^{tv}_{sk}$ *and* $\mathtt{M}^{me}_{sk}$*, disagree with each other.* ***Bold red*** *numbers are related to* Σ^{tv}_{sk} *or* Σ^{me}_{sk}*, where their values are considerably higher than the corresponding* Σ^{tv}_{dk} *and* Σ^{me}_{dk}*. Red numbers highlight the* Σ^{tv}_{dk} *or* Σ^{me}_{dk} *where their value are higher than value of* Σ^{tv}_{sk} *and* Σ^{me}_{sk}*.*

Subject Specifications					dPPG				Single-Kinect			
No	**GEN**	**AGE**	**BMI**	**SMK**	$\mathtt{M}^{tv}_{dk}$	$\mathtt{M}^{me}_{dk}$	Σ^{tv}_{dk}	Σ^{me}_{dk}	$\mathtt{M}^{tv}_{sk}$	$\mathtt{M}^{me}_{sk}$	Σ^{tv}_{sk}	Σ^{me}_{sk}
#01	M	33	22.8	*Quit*	4.82	4.72	0.57	0.41	4.48	4.67	1.11	0.46
#02	M	26	23	Never	6.81	5.73	1.03	0.29	9.05	5.79	**2.34**	0.54
#03	F	23	22	Never	3.19	3.40	0.49	0.56	2.71	3.43	0.90	0.61
#04	F	35	23.2	Never	3.29	3.30	0.48	0.43	3.32	3.48	0.66	0.69
#05	M	32	25.1	*Smokes*	9.05	7.18	0.55	0.47	9.30	7.11	1.01	0.40
#06	M	27	23.8	Never	5.95	5.65	0.36	0.48	6.88	5.59	0.39	0.51
#07	M	22	23.4	Never	6.33	5.75	0.30	0.73	5.93	5.84	0.40	0.74
#08	M	29	26.3	Never	**6.47**	**6.56**	1.01	0.62	**2.98**	**12.5**	1.05	**4.31**
#09	F	37	20.4	Never	3.08	2.97	0.55	0.23	3.09	3.00	**1.46**	0.27
#10	M	27	20.4	Never	4.51	4.30	0.32	0.40	4.30	4.34	0.65	0.57
#11	M	36	22.2	Never	7.74	4.61	0.32	0.27	6.71	4.86	**2.68**	0.68
#12	F	26	22.2	Never	4.06	3.77	0.19	0.94	4.16	3.64	**0.77**	0.67
#13	M	27	26	Never	5.69	4.93	0.28	0.19	5.54	4.91	0.53	0.16
#14	M	32	24.6	Never	5.53	5.74	0.85	1.01	5.25	6.08	**1.82**	**2.38**
#15	F	31	23	Never	4.78	4.18	0.69	0.17	4.69	4.25	0.52	0.29
#16	M	**45**	26.8	*Quit*	4.23	5.32	1.51	0.61	5.95	6.34	**3.47**	**2.03**
#17	M	25	24.5	Never	5.14	4.17	0.54	0.52	5.58	5.32	**1.73**	**1.94**
#18	M	27	20.2	Never	5.25	3.62	0.42	0.30	5.39	3.56	0.45	0.24
#19	M	33	27.2	Never	6.92	6.10	0.64	0.27	5.53	6.11	**2.55**	0.21
#20	M	38	22.8	Never	5.47	5.01	0.58	0.50	5.59	5.59	0.92	**1.07**
#21	M	29	27.8	*Smokes*	10.2	5.17	2.61	1.04	7.03	5.14	2.76	1.30
#22	M	26	22.5	*Smokes*	5.63	4.42	0.76	0.25	5.88	4.45	0.99	0.30
#23	M	32	19.6	Never	7.99	5.29	0.31	0.48	8.81	5.29	**1.08**	0.46
#24	M	23	24	Never	**5.71**	**5.45**	1.30	1.09	**4.96**	**8.82**	1.48	**6.81**
#25	M	26	**33.4**	Never	**5.22**	**5.16**	0.35	0.22	**7.14**	**11.9**	**1.32**	**6.16**
#26	M	27	24.1	Never	5.90	4.89	0.54	0.43	7.67	5.18	1.11	0.70
#27	M	25	26.5	Never	5.59	5.33	0.78	0.62	5.72	5.31	0.73	0.60
#28	M	31	20.9	*Quit*	4.52	5.00	1.84	0.46	4.95	5.21	**2.27**	0.84
#29	M	33	26.1	Never	5.95	4.97	0.24	0.19	8.41	4.92	**1.43**	0.21
#30	F	31	19.4	Never	4.36	3.91	0.33	0.45	3.95	3.92	0.75	0.47
#31	F	34	19.6	*Smokes*	6.34	3.92	1.31	0.10	7.78	4.04	**2.70**	0.36
#32	F	37	27.7	*Smokes*	3.27	3.04	0.16	0.17	3.43	3.42	0.35	**0.76**
#33	M	35	29.1	*Quit*	7.14	5.42	0.72	1.03	9.28	5.88	**1.69**	1.56
#34	M	24	21.5	*Quit*	6.07	4.31	0.54	0.61	6.24	4.30	1.05	0.62
#35	M	38	23.7	*Quit*	5.28	7.91	3.48	1.75	3.76	7.86	1.87	1.62

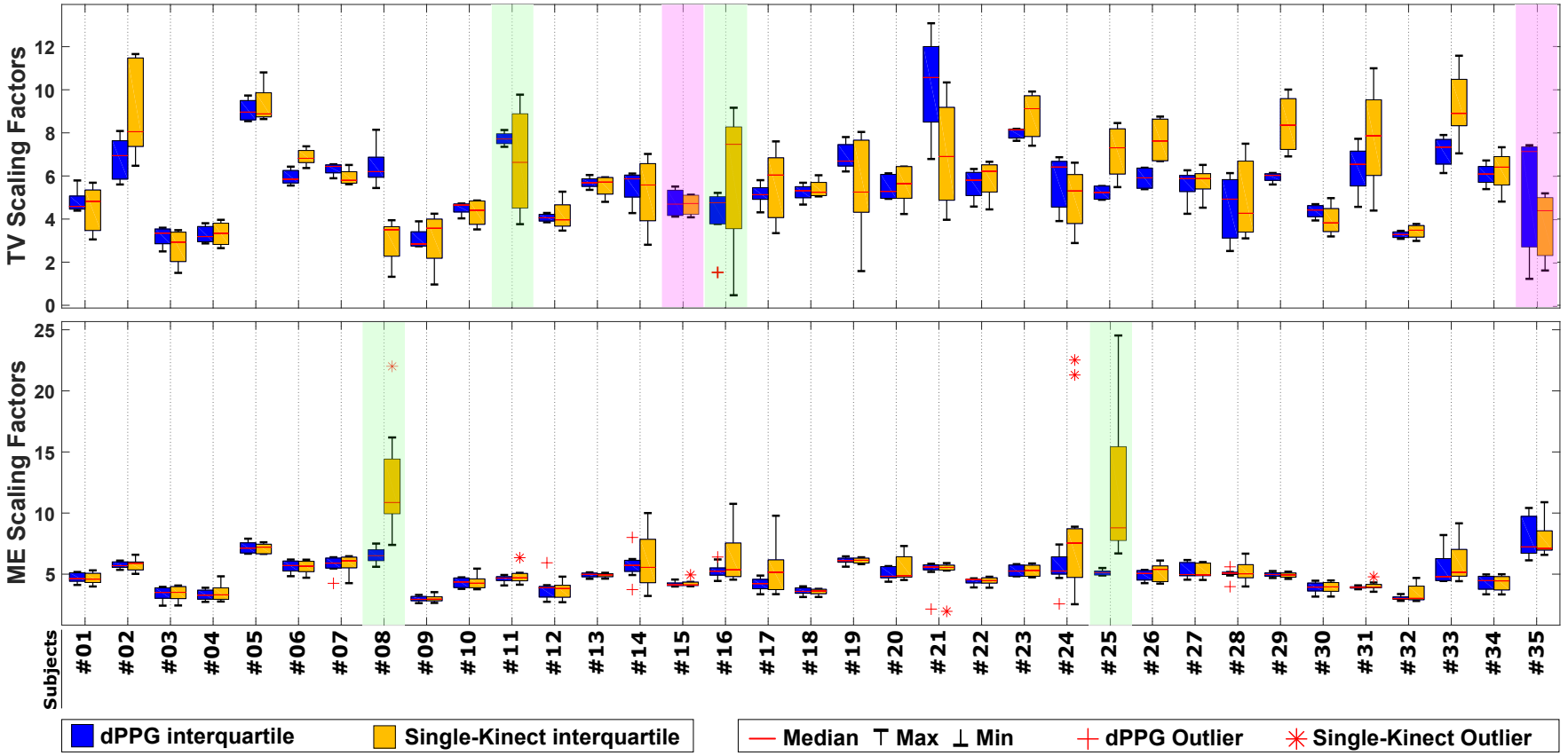


Figure 6.7: *Boxplot of statistics of within-subject tidal volume (top) and main effort (bottom) scaling factors of 35 subjects' trials, in which interquartile range, median, max, min and outliers of tidal volume and main effort scaling factors are illustrated for the dPPG (blue) and the single-Kinect (orange) approaches. The interquartile range of the single-Kinect tidal volume scaling factors are wider across all subjects except for a few, e.g., subjects #15 and #35 (pink highlighted). In particular,* Σ_{sk}^{tv} *is 8.4 and 2.3 times higher than* Σ_{dk}^{tv} *for the green highlighted subjects #11 and #16. Similarly, for the main effort scaling factors,* Σ_{sk}^{me} *is 6.9 and 28 times higher than* Σ_{dk}^{me} *for the green highlighted subjects #8 and #25.*

Among all subjects, only subject #35 has a considerably greater Σ^{tv}_{dk} (bold red) than Σ^{tv}_{sk}, whereas $\Sigma^{tv}_{sk}, \Sigma^{me}_{sk}$ are higher for numerous subjects (bold red). For example, Σ^{tv}_{sk} is 8.4 and 2.3 times higher than Σ^{tv}_{dk} for subjects #11 and #16, and Σ^{me}_{sk} is 6.9 and 28 times higher than Σ^{me}_{dk} for subjects #8 and #25 (highlighted in green in Figure 6.7). The greater the scaling factors' STD, the higher the depth-based PFT measures' error would be. For example, the average error of TV and FVC measures decreases from 0.23 and 0.84 in the single-Kinect method to 0.07 and 0.19 in the dPPG approach for subjects #16 and #25, respectively.

Closely looking into our statistical analysis in Table 6.5 and investigating any potential association of subject specifications, i.e., gender, age, BMI and smoking condition (SMK), to the scaling factors, no meaningful relationship could be established. For example, subject #25 with the highest BMI = 33.4 Kg/cm^2 has very closed dPPG *tidal volume* and *main effort* scaling factors with also very low STDs. Similar scaling factor statistics has been reported for subject #09 with a very low BMI =20.4 Kg/cm^2. Thus, the *tidal volume* and *main effort* scaling factors coherency and consistency does not seem to be relevant to the BMI. Similar examples also exist for smoking condition. For example, while the smoker subject #32 has very close *tidal volume* and *main effort* scaling factors with small STDs, subject #21 with similar smoking condition has completely different *tidal volume* and *main effort* scaling factors with rather high STDs. Regarding the age, subjects #20 and #35 who are both 38 years old male subjects with similar BMIs, have different scaling factors coherency and consistency. Also, no association between gender and scaling factor statistics was found. However, noting that our dataset is rather small, establishing such associations requires larger datasets with more variation in the subjects lung pathology, body specifications and background.

The scaling factor outliers, shown by (+) and (∗) in Figure 6.7 for the dPPG and single-Kinect approaches, specify the scaling factors which are highly different with the other scaling factors obtained for the same subject. Comparing the *tidal volume* to the *main effort* within-subject scaling factors, almost all of the outliers belong to the *main effort* data calibration with only one *tidal volume* outlier for the subject #16. This implies that the trunk motion artifacts caused by deep forced *inhalation–exhalation* could be the main reason for the *main effort* scaling factor outliers.

Finally, Table 6.6 shows the mean ($\mu_{\mathtt{M}'}$) and STD ($\sigma_{\mathtt{M}'}$) of the absolute difference between 'the average of within-subject *tidal volume* scaling factors' and 'the average of within-subject *main effort* scaling factors', i.e., $\mathtt{M}' = |\mathtt{M}^{tv}_x - \mathtt{M}^{me}_x|$, where $x = dk$ for the dPPG approach and $x = sk$ for the single-Kinect method. It also shows the normalised mean

Table 6.6: *Statistics of* $\mathtt{M}' = |\mathtt{M}_x^{tv} - \mathtt{M}_x^{me}|$ *and* $\Sigma' = |\Sigma_x^{tv} - \Sigma_x^{me}|$, *x=sk or dk across 35 subjects in the dPPG and the single-Kinect methods.*

Method	$\mu_{\mathtt{M}'}$	$\sigma_{\mathtt{M}'}$	$\Omega_{\mathtt{M}'}$	$\mu_{\Sigma'}$	$\sigma_{\Sigma'}$	$\Omega_{\Sigma'}$
dPPG	**1.02**	**1.10**	**0.19**	**0.37**	**0.44**	**0.57**
single-Kinect	1.72	1.98	0.81	1.03	1.28	0.85

of $\mathtt{M}'$ as $\Omega_{\mathtt{M}'} = \mu_{\mathtt{M}'}/\alpha_{\mathtt{M}'}$, in which the normalisation factor $\alpha_{\mathtt{M}'}$ is defined as the average of $\{\mathtt{M}_x^{tv}, \mathtt{M}_x^{me}\}$ across all subjects. Table 6.6 also presents similar statistics $(\mu_{\Sigma'}, \sigma_{\Sigma'}, \Omega_{\Sigma'})$ for the absolute difference between 'the STD of within-subject *tidal volume* scaling factors' and 'the STD of within-subject *main effort* scaling factors', i.e., $\Sigma' = |\Sigma_x^{tv} - \Sigma_x^{me}|$. As seen, mean, STD and the normalised mean of Σ', are notably smaller for the dPPG, where it shows better agreement between *tidal volume* and *main effort* scaling factors. For example, $\mathtt{M}_{dk}^{tv}$ and $\mathtt{M}_{dk}^{me}$ are almost equal for subjects #8, #24 and #25 (in blue in Table 6.5), whereas $\mathtt{M}_{sk}^{tv}$ and $\mathtt{M}_{sk}^{me}$ show considerable disagreement for these subjects (in orange in Table 6.5).

6.5 Conclusion

We introduced depth-based whole-body photoplethysmography to increase remote PFT measure accuracy by decoupling a subject's trunk movements from their chest-surface respiratory motions using two opposing depth sensors. First, two Kinects are calibrated and synchronised to construct a dynamic 3-D model of a subject during a breathing test. Using a 3-D mask, thoracoabdominal volume is automatically segmented and used to retrieve volume–time data. This data is then calibrated using the intra-subject scaling factors, learnt in a training phase, and 11 clinical FVC and SVC measures are computed. We validate the dPPG computed measures by comparing them to the measures obtained from a spirometer. The evaluation results show very good improvement compared to the single-Kinect approach.

The proposed dPPG method does not perform in real-time as the body data acquisition, trunk reconstruction and PFT computation stages operate separately. While the data acquisition and the PFT computation stages perform in nearly real-time, the trunk reconstruction for each breathing test performance is accomplished in less than a minute. However, we feel confident to project that by applying GPU-based 3-D reconstruction techniques, and incorporating these stages using further development, dPPG can operate

in real-time.

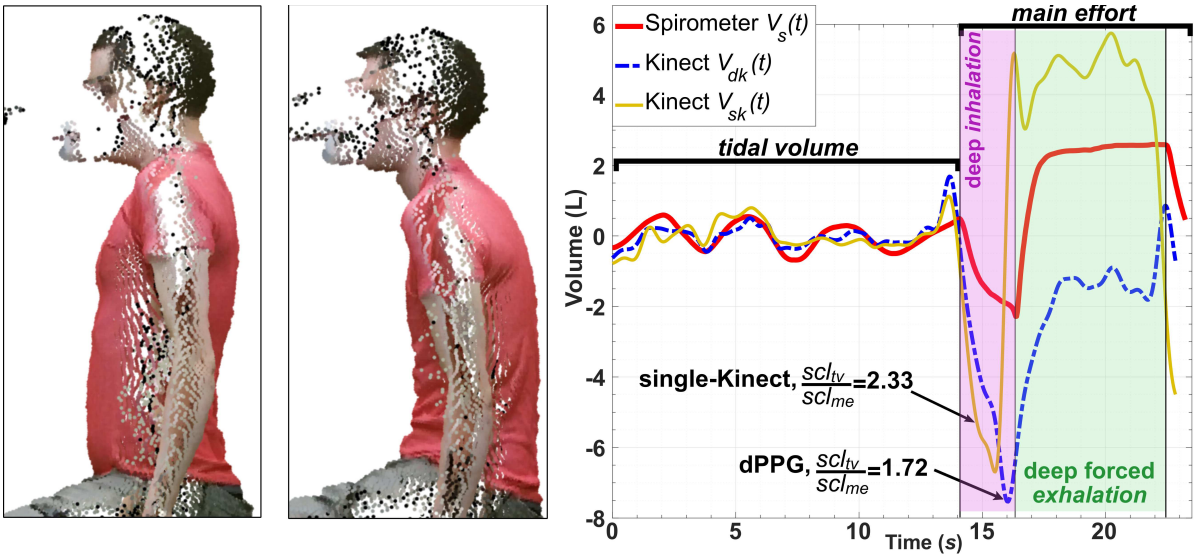
The proposed method for decoupling body movements from respiratory motions results in *tidal volume* and *main effort* scaling factors that are more consistent and better agree with each other than the single-Kinect method. However, they are not identical enough to be a unique intra-subject scaling factor that could be used to calibrate the whole volume–time data. As a potential reason this could be because of estimating the respiratory data using a simple averaging based method which is not very effective in filtering complex trunk motions specifically during the *main effort* breathing. Thus, future work can include the investigation of other methods of volume estimation for more effective trunk motion filtering.

Chapter 7

Trunk Shape Modelling for Body Motion Artifacts Correction

In Chapter 6, we presented the dPPG approach which filtered the natural reaction of the human respiratory system by constructing a 3-D model of a subject's trunk and then subtracting the average depth of the chest-wall from the average depth of the posterior-wall per frame. This was intended to reduce the trunk motion artifacts during the breathing test, and therefore, allowed subjects to perform PFT as in routine spirometry.

While the dPPG method is able to reduce the trunk motion artifacts to some extent and show advancement in the accuracy of the computed measures and also in the consistency of the scaling factors, there is still room for further improvements. First, the dPPG method cannot filter complex trunk motion patterns, particularly at the *main effort* inhalation–exhalation stage, due to using only the average depth of the chest and posterior walls. Second, different patterns of trunk motion in *tidal volume* and *main effort* breathing, requires individual calibration of their volume–time data rather than calibrating the whole data using a single scaling factor. For example, Figures 7.1a and 7.1b show the body's natural reaction to the *main effort* deep inhalation and deep forced exhalation during a routine spirometry FVC test, respectively. Correspondingly, Figure 7.1c presents the volume–time data of the single-Kinect $V_{sk}(t)$ and dPPG $V_{dk}(t)$ methods for this test which are entirely calibrated using only the *tidal volume* scaling factor. As seen, both the single-Kinect and dPPG methods fail to provide an accurate enough volume–time data, though dPPG performed better. Also, the ratios of the *tidal volume* to *main effort* scaling factors, i.e., scl_{tv}/scl_{me}, for the single-Kinect (2.33) and

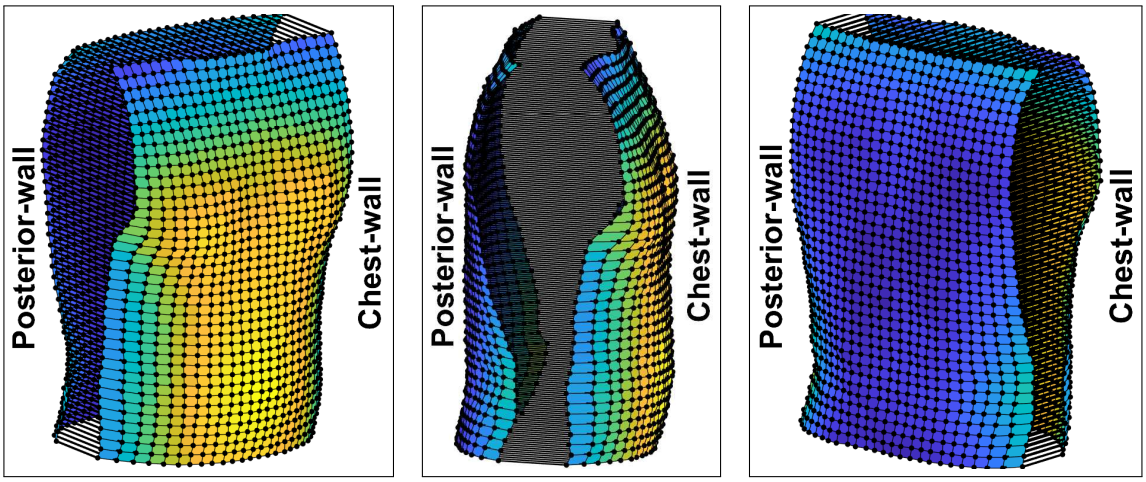


(a) *deep inhalation* **(b)** *deep forced exhalation* **(c)** *volume-time data*

Figure 7.1: *Trunk's natural reaction to (a) deep inhalation, and (b) deep forced exhalation, in main effort breathing. (c) The single-Kinect and dPPG volume–time data and their tidal volume to main effort scaling factors ratio.*

dPPG (1.72), denote that both methods obtained considerably different *tidal volume* and *main effort* scaling factors. This can be perceived by comparing the scale of depth-based *main effort* to the spirometer's *main effort* in Figure 7.1c.

In this chapter, we present our active shape modelling approach to extract the chest-surface respiratory pattern by temporal modelling and tracking of the trunk shape, generated using the measurements acquired by two opposing depth sensors (Section 7.1). Instead of filtering the trunk motion by subtracting the average depth of the chest from the average depth of the posterior in the dPPG method, here we extract the respiratory pattern by performing PCA on temporal 3-D geometrical features extracted from the chest and posterior shape models in $\mathbb{R}^3$ space. To present the real volume of exchanged air, the respiratory data is then calibrated using scaling factors learnt in an efficient training phase (Section 7.2). We validate our method on the dual-Kinect PFT dataset, by computing the normalised L_2 error, dynamic time warping and Fréchet [23] distances, as well as correlation of determination between the depth-based and spirometer volume–time data. Comparing our results to the single-Kinect and dPPG methods verifies that our active trunk shape modelling approach outperforms previous approaches across all evaluation metrics. We also demonstrate that this method calibrates the whole volume–time data using only *tidal volume* scaling factors with a significantly lower error than the single-Kinect and dPPG approaches (Section 7.3). Finally, we conclude this chapter by briefly reviewing our method and its key achievements (Section 7.4).



(a) *Front view* **(b)** *Side view* **(c)** *Back view*

Figure 7.2: *(a)-(c) A sample of a subject's reconstructed 3-D model of the trunk in different points of view. The black dots show the uniformly generated data points over the chest and posterior walls. The partial volumes are demonstrated by a direct line between the chest and posterior data points.*

7.1 Trunk Shape Modelling

Using depth measurement acquired by our 3-D data acquisition and registration pipeline (Chapter 5), we first reconstruct the dynamic 3-D model of the subject's body during the breathing test. Then, chest-wall ROI and its corresponding region on the posterior-wall are automatically isolated by a 3-D mask generated from skeletal joint data. Figures 7.2a, 7.2b and 7.2c present a sample of the subject's final reconstructed 3-D model of the trunk from different points of view.

Noting that the trunk motion is reflected on both chest and posterior walls with the same direction and magnitude, while the respiratory motion mainly appears on the chest-wall, dPPG filters the trunk motion by merely subtracting the average distance of the chest-wall data points $\mathbf{P}_{ch}$ from the average distance of the posterior-wall data points $\mathbf{P}_{po}$ in each frame. Although this simple subtraction reduces the trunk motion artifacts to a notable extent, it fails to correctly filter more complex motion patterns due to,

(i) using the depth as the only feature of the chest and posterior walls, i.e., z-coordinate of $\mathbf{P}_{ch}$ and $\mathbf{P}_{po}$ in $\mathbb{R}^3$ space, and

(ii) eliminating useful potential features by filtering $\mathbf{P}_{ch}$ and $\mathbf{P}_{po}$ to single average depth values.

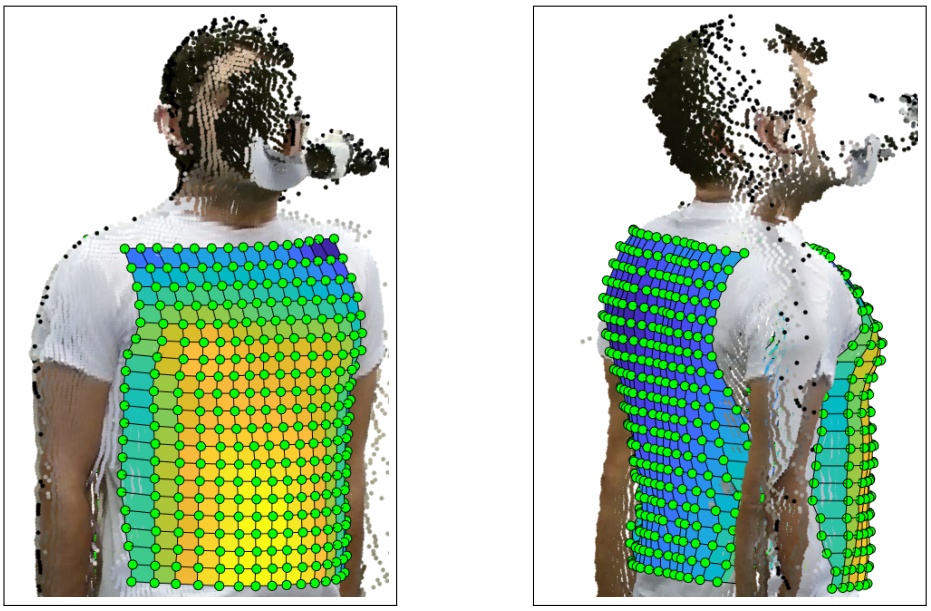
We address these issues by extracting a set of temporal geometrical features over the entire chest and posterior walls within a sequence. Using these features, partial volumes are uniformly computed from the full trunk, and tracked per frame. However, as a markerless trunk reconstruction approach, we are faced with two main challenges. First, the number and location of data points in either of the chest and the posterior regions may vary from one frame to another. Second, since each Kinect captures either the chest or the posterior, these two regions are different in the number of the data points and their locations. Thus, it is quite unlikely that there would be a corresponding co-located posterior data point for a chest data point in the xy-plane.

We address these issues by,

(i) defining a fixed size 2-D region $\boldsymbol{\mathcal{R}}_{xy}(t)$ which covers the chest in the xy-plane at time t through the sequence, and generating a fixed number of points uniformly distributed on it, and

(ii) defining temporal interpolant functions $\boldsymbol{\mathcal{I}}^t_{ch}$ and $\boldsymbol{\mathcal{I}}^t_{po}$ to compute corresponding chest and posterior data points in $\mathbb{R}^3$ space which are located at the same position in the xy-plane.

Figure 7.3 shows the uniformly distributed data points on the chest and posterior shapes



(a) *Front view* **(b)** *Side view*

Figure 7.3: *A subject's chest and posterior shape models with their data points presented as small green spheres.*

of a subject's trunk as small green spheres. The interpolant functions are generated by applying a Delaunay triangulation on the original data points [20].

7.2 Volume–time Data Retrieval and Calibration

Using $\boldsymbol{\mathcal{I}}^t_{ch}$ and $\boldsymbol{\mathcal{I}}^t_{po}$, partial volumes $\mathbf{v}_{ij}(t)$ are computed over the trunk as

$$\mathbf{v}_{ij}(t) = \left[\boldsymbol{\mathcal{I}}^t_{po}\Big(\mathbf{p}_{ij}(t)\Big) - \boldsymbol{\mathcal{I}}^t_{ch}\Big(\mathbf{p}_{ij}(t)\Big)\right], \forall \mathbf{p}_{ij}(t) \in \boldsymbol{\mathcal{R}}_{xy}(t). \tag{7.1}$$

Figure 7.2 demonstrates the partial volumes within the whole trunk as the direct lines between the corresponding data points of the chest-wall and posterior-wall ROIs. Since the extent of body motion along the trunk varies during the breathing test, i.e., from the minimum at the hips to its maximum at the shoulders, we accumulate $\mathbf{v}_{ij}(t)$ along the longitudinal direction to further reduce their sensitivity to the trunk motion. Thus, the final feature matrix for the whole sequence is created by extracting these partial volumes and their location in the xy-plane for each frame of the sequence as

$$\mathbf{F} = \left[\left\langle \boldsymbol{\mathcal{R}}_{xy}(t), \textstyle\sum_j \mathbf{v}_{ij}(t) \right\rangle\right]_{t=0}^{t=\tau}, \tag{7.2}$$

where τ indicates the PFT sequence duration.

The feature matrix $\mathbf{F}$ is normalised to zero mean for each frame of the sequence (variable) and its covariance matrix is computed as

$$\mathbf{C}_{\mathbf{F}} = \frac{1}{N_\tau - 1} \sum_{i=1}^{N_\tau} (\mathbf{F}_i - \mu_{\mathbf{F}_i})(\mathbf{F}_i - \mu_{\mathbf{F}_i})^T, \tag{7.3}$$

where N_τ and $\mu_{\mathbf{F}_i}$ indicate the number of the variables (frames) and the mean value of each frame's data samples, respectively. Finally, PCA is computed by factorisation of $\mathbf{C}_{\mathbf{F}}$ to orthogonal and diagonal matrices. Since $\mathbf{C}_{\mathbf{F}}$ is a square symmetric matrix with rank Υ, it can be diagonalised as

$$\mathbf{C}_{\mathbf{F}} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T = \sum_{i=1}^{\Upsilon} \lambda_i v_i v_i^T, \tag{7.4}$$

where v_i is the ith principal component (PC) and λ_i is the ith eigenvalue of $\mathbf{C}_{\mathbf{F}}$ which

Table 7.1: *Statistical analysis of the percentage of the variance presented by the first 10 principal components across the whole dataset.*

Metric	1st PC	2nd PC	3rd PC	4th PC	5th PC	6th PC	7th PC	8th PC	9th PC	10th PC
Mean	**97.76%**	1.64%	0.28%	0.11%	0.05%	0.03%	0.02%	0.01%	0.01%	0.01%
STD	**2.66%**	2.16%	0.36%	0.16%	0.06%	0.03%	0.02%	0.01%	0.01%	0.01%
Median	**98.52%**	0.94%	0.17%	0.06%	0.03%	0.02%	0.01%	0.01%	0.01%	0.01%
Min	**78.99%**	0.11%	0.03%	0.01%	0.005%	0.004%	0.003%	0.003%	0.003%	0.002%
Max	**99.78%**	15.59%	4.01%	2.16%	0.66%	0.29%	0.16%	0.11%	0.08%	0.06%

indicates the variance of the data along the ith PC. Statistical analysis of the first 10 PCs across the whole dataset, reported in Table 7.1, shows that the first component presents an average 97.76% variation of the data. Thus, the trunk shape modelling volume–time data $V_r(t)$ is provided by choosing only the first PC and the other PCs are not used as they only present $< 2.24\%$ of the data variability. Figure 7.4 presents the plot of the first three PCs for a sample sequence. As seen, the first PC presents the main variability of the data while the other two PCs hardly present any variability. It is worth noting that deep convolutional neural networks can also be investigated for extracting more effective and distinguishable features per subject instead of extracting hand-crafted geometrical features.

To eliminate high frequency noise, $V_r(t)$ is then smoothed using a 4th order Butterworth low-pass filter. Given a wide range of respiratory rates for adults and elderly at $12 -$

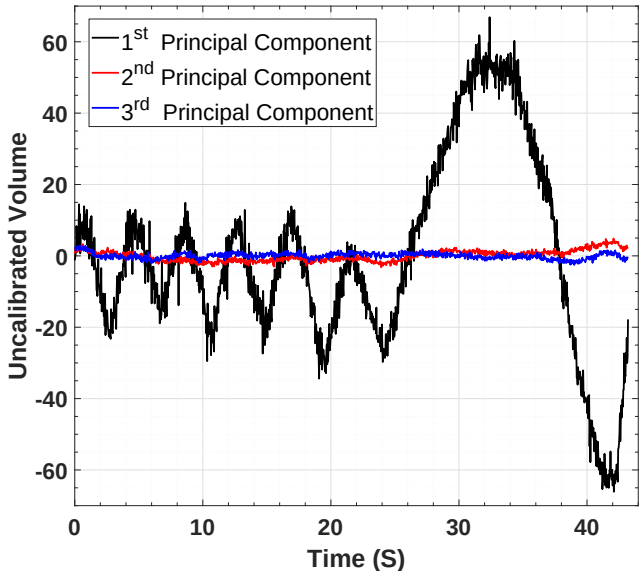
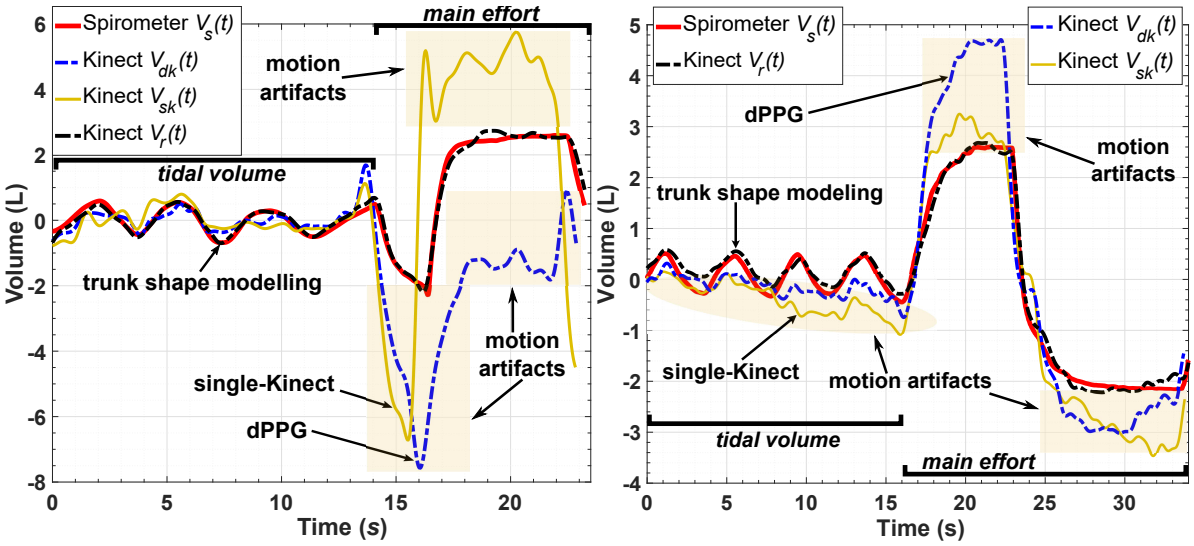


Figure 7.4: *The first three PCs for a sample breathing sequence. The first PC presents the main variability of the data compared to the second and third PCs.*



(a) *FVC volume–time data* **(b)** *SVC volume–time data*

Figure 7.5: *Comparing the depth-based volume–time data $V_r(t)$, $V_{sk}(t)$ and $V_{dk}(t)$ obtained by the trunk shape modelling approach, single-Kinect and dPPG methods, respectively, for (a) FVC and (b) SVC tests. Our proposed approach greatly corrects the motion artifacts that the single-Kinect and the dPPG methods fail to improve.*

36 bpm ($0.2 - 0.6$ Hz) [55], the cut-off frequency was chosen as 1.2 Hz, which is twice the maximum respiratory frequency and ensures preserving the respiratory information while filtering the high frequency noise.

Figures 7.5a and 7.5b present $V_r(t)$ (dashed black curve) and its comparison to the volume–time data obtained by the spirometer $V_s(t)$ (solid red curve), single-Kinect $V_{sk}(t)$ (solid orange curve), and dPPG $V_{dk}(t)$ (dashed blue curve) methods for FVC and SVC tests, respectively. As seen, while $V_{sk}(t)$ and $V_{dk}(t)$ are significantly affected by the trunk motion (yellow highlighted regions), especially at the deep forced inhalation–exhalation stage, $V_r(t)$ is much more accurate.

To present the real volume of exchanged air (in litres), $V_r(t)$ is calibrated by linearly scaling the y-axis using scaling factors learnt per subject in a *training* phase from spirometer and depth-based training volume–time data. In the *training* phase, corresponding data samples of $V_s(t)$ and $V_r(t)$ are identified by sampling $V_s(t)$ at 30 Hz, and detrending $V_s(t)$ and $V_r(t)$ (using EMD), and temporally aligning them. Next, the *tidal volume* and *main effort* scaling factors are separately computed by solving linear least square equations for $V_s(t)$ and $V_r(t)$ using an overdetermined system. This process is repeated for every

pair of training spirometer and depth-based volume–time data which provides several pairs of training scaling factors.

For a test sequence, $V_r(t)$ is first computed using the trunk shape modelling approach, and then calibrated by an *average* of all the training scaling factors, instead of applying similarity metrics exploited in the single-Kinect and dPPG methods. This is because our trunk shape modelling approach filters the trunk motion to a great extent such that calibrating a test volume–time data using the average of the training scaling factors would be more accurate than applying similarity metrics.

7.3 Experimental Results

7.3.1 Volume–time Data Evaluation

Since the PFT measures are computed from specific parts of volume–time data [129], evaluation based on only these measures does not provide an accurate enough validation of the computed depth-based respiratory data. Thus, unlike our single-Kinect and dPPG works in which the evaluation is based on the measures, here we validate our proposed method at the signal level over the whole calibrated respiratory data. In particular, we validate our method on the dual-Kinect PFT dataset by computing the mean $\pm$ STD of

(i) normalised L_2 error (NL_2),

(ii) Fréchet distance (FRD) [23],

(iii) dynamic time warping distance (DTW), and

(iv) correlation of determination (R^2),

between the computed and spirometer volume–time data for 155 FVC and 145 SVC sequences of all 35 subjects. Note that our trunk shape modelling approach is able to process and retrieve the volume–time data of the 2 sequences which could not be processed by the dPPG due to their complex trunk motion.

The L_2 error is normalised to the sequence duration and the difference between the

maximum and minimum volume of exchanged air, computed as

$$\mathrm{NL}_2 = \frac{\sum_{t=0}^{t=\tau} \Big(V_s(t) - V_d(t)\Big)^2}{\tau * \Big(V_s(t_{max}) - V_s(t_{min})\Big)}, \tag{7.5}$$

where $V_s(t)$ and $V_d(t)$ are the spirometer and depth-based volume–time data, t_{max} and t_{min} are the timestamps of the global maximum and minimum of $V_s(t)$, and τ is the sequence data duration. Other metrics, i.e., FRD [23], DTW and R^2 are standard signal comparison measures.

Tables 7.2 and 7.3 present the evaluation results of the single-Kinect, dPPG, and the trunk shape modelling approaches for the *tidal volume* and *main effort* parts of volume–

Table 7.2: *Depth-based volume–time data (tidal volume + main effort) evaluation results for 155 FVC sequences, calibrated using tidal volume and main effort scaling factors*

	Metric	Single-Kinect	dPPG	Trunk Shape Modelling
tidal volume	**NL_2**	0.033 ± 0.041	0.013 ± 0.018	$\mathbf{0.004 \pm 0.006}$
	FRD	7.87 ± 4.98	4.86 ± 3.16	$\mathbf{3.02 \pm 1.77}$
	DTW	91.54 ± 76.06	43.68 ± 38.51	$\mathbf{22.49 \pm 19.08}$
	R^2	0.42 ± 0.31	0.66 ± 0.24	$\mathbf{0.85 \pm 0.11}$
main effort	**NL_2**	0.908 ± 2.796	0.068 ± 0.122	$\mathbf{0.014 \pm 0.015}$
	FRD	22.34 ± 32.61	7.80 ± 6.45	$\mathbf{3.89 \pm 1.92}$
	DTW	211.52 ± 499.81	38.06 ± 48.05	$\mathbf{18.38 \pm 12.48}$
	R^2	0.74 ± 0.29	0.91 ± 0.15	$\mathbf{0.98 \pm 0.02}$

Table 7.3: *Depth-based volume–time data (tidal volume + main effort) evaluation results for 145 SVC sequences, calibrated using tidal volume and main effort scaling factors*

	Metric	Single-Kinect	dPPG	Trunk Shape Modelling
tidal volume	**NL_2**	0.038 ± 0.066	0.012 ± 0.011	$\mathbf{0.005 \pm 0.006}$
	FRD	10.00 ± 7.76	6.07 ± 3.12	$\mathbf{4.03 \pm 2.00}$
	DTW	147.59 ± 155.73	71.59 ± 51.78	$\mathbf{39.13 \pm 26.34}$
	R^2	0.41 ± 0.29	0.63 ± 0.26	$\mathbf{0.80 \pm 0.15}$
main effort	**NL_2**	0.951 ± 6.681	0.025 ± 0.026	$\mathbf{0.013 \pm 0.013}$
	FRD	20.44 ± 43.88	6.60 ± 3.30	$\mathbf{4.97 \pm 2.65}$
	DTW	239.88 ± 779.48	35.66 ± 34.84	$\mathbf{30.28 \pm 23.89}$
	R^2	0.91 ± 0.13	$0.97 \pm 0.03^{\dagger}$	$\mathbf{0.98 \pm 0.02}$

time data which are separately calibrated using corresponding scaling factors. The evaluation results across all the metrics confirm that while the dPPG method shows a notable improvement to the single-Kinect approach, the trunk shape modelling method achieves even more significant accuracy than the dPPG. In particular, while the dPPG method reduces the *tidal volume* and *main effort* average NL_2 errors of 300 sequences from 0.0357 and 0.9288 (obtained by the single-Kinect method) to 0.0124 and 0.0473, respectively, the trunk shape modelling approach further reduces NL_2 errors to 0.0049 and 0.0138, which are 2.5 and 3.4 times lower than the errors achieved by the dPPG. Note that although dPPG achieves a close (†) correlation of determination (R^2) to the trunk shape modelling method for the SVC *main effort* tests, NL_2, FRD and DTW, still achieve considerably lower error for the trunk shape modelling method. This implies that the correlation analysis might not be as reliable as the other evaluation metrics for this application.

Table 7.4 reports the mean ± STD of the same metrics for 155 FVC and 145 SVC sequences in which the volume–time data is entirely calibrated using only *tidal volume* scaling factor. As the results confirm, the trunk shape modelling method outperforms the single-Kinect and dPPG methods across all of the evaluation metrics, except for R^2 of the SVC tests (⋆) where our method achieves almost the same R^2 as dPPG. In particular, the average NL_2 error across 300 sequences is reduced from 0.454 and 0.136 by the single-Kinect and dPPG methods to 0.050 by our method, which shows 9.1 and 2.7 times lower error.

Comparing Table 7.4 to Tables 7.2 and 7.3, verifies that calibrating the entire volume–time data using only the *tidal volume* scaling factor achieves lower accuracy than the individual calibration of *tidal volume* and *main effort*. However, as a significant achievement,

Table 7.4: *Depth-based volume–time data evaluation results for 155 FVC and 145 SVC sequences, calibrated using only tidal volume scaling factor*

	Metric	Single-Kinect	dPPG	Trunk Shape Modelling
FVC tests	**NL_2**	0.679 ± 1.243	0.181 ± 0.309	$\mathbf{0.051 \pm 0.079}$
	FRD	39.31 ± 35.51	21.08 ± 17.89	$\mathbf{11.95 \pm 8.08}$
	DTW	541.4 ± 566.1	248.3 ± 246.7	$\mathbf{136.2 \pm 125.8}$
	R^2	0.59 ± 0.32	0.78 ± 0.28	$\mathbf{0.93 \pm 0.06}$
SVC tests	**NL_2**	0.229 ± 0.327	0.092 ± 0.137	$\mathbf{0.049 \pm 0.047}$
	FRD	31.15 ± 20.64	19.72 ± 13.35	$\mathbf{15.49 \pm 7.34}$
	DTW	510.2 ± 470.9	279.6 ± 250.6	$\mathbf{204.1 \pm 132.6}$
	R^2	0.73 ± 0.23	$0.91 \pm 0.06^{\star}$	0.90 ± 0.07

the difference in this accuracy is notably lower in the trunk shape modelling approach than the single-Kinect and dPPG methods. While the average NL_2 error across all FVC sequences increases by 0.042 in the trunk shape modelling approach, it increases by 0.208 and 0.141 in the single-Kinect and dPPG methods. Thus, the trunk shape modelling approach is 5.0 and 3.4 times more accurate than the single-Kinect and dPPG methods in calibrating the whole volume–time data using only *tidal volume* scaling factor. Such calibration eliminates the threshold-based analysis required for separate calibration of *tidal volume* and *main effort*, which not only allows our approach to perform online, but also removes the calibration error due to incorrect keypoints computation.

Altogether, since the trunk shape modelling approach achieves a higher accuracy in the signal level, it is logical to induce that it would also provide more accurate clinical PFT measures.

7.3.2 Calibration Assessment and Scaling Factors Evaluation

We compare the performance of our simplified volume–time data calibration method, i.e., using the average of training scaling factors, to the similarity based measures exploited in the single-Kinect and dPPG approaches. For this, we calibrate the trunk shape modelling volume–time data using, (*i*) the average of training scaling factors, and (*ii*) similarity based measures, which provides two sets of calibrated volume–time data. Both of these data sets are then evaluated against the spirometer by computing the mean of NL_2, FRD, DTW and R^2 metrics for *tidal volume* and *main effort* parts of all 300 FVC and SVC sequences. As Table 7.5 shows, our simplified method attain higher or equal performance across all metrics due to achieving more consistent intra-subject scaling factors within different trials.

Table 7.5: *Comparing the performance of the volume–time data calibration using the average of training scaling factor to the similarity based measures exploited in the single-Kinect and dPPG methods across* 300 *PFT sequences*

		Evaluation Metrics			
	Calibration Method	**NL_2**	**FRD**	**DTW**	**R^2**
tidal volume	**Averaging**	**0.005**	**3.51**	**30.53**	**0.83**
	Similarity Measure	0.005	3.54	31.15	0.83
main effort	**Averaging**	**0.014**	**4.41**	**24.13**	**0.98**
	Similarity Measure	0.015	4.58	25.81	0.98

Table 7.6: *Comparing the ratio of tidal volume and main effort scaling factors, i.e., scl_{tv}/scl_{me} computed across 155 FVC and 145 SVC tests*

Metric	Single-Kinect	dPPG	Trunk Shape Modelling
FVC scl_{tv}/scl_{me}	1.11 ± 0.61	1.27 ± 0.41	$\mathbf{1.07 \pm 0.17}$
SVC scl_{tv}/scl_{me}	1.11 ± 0.56	1.30 ± 0.41	$\mathbf{1.10 \pm 0.14}$

To further evaluate our method effectiveness in filtering trunk motion which has different patterns in *tidal volume* and *main effort* data, we computed the mean $\pm$ STD of *tidal volume* to *main effort* scaling factors ratio, i.e., scl_{tv}/scl_{me}, for all FVC and SVC sequences. As reported in Table 7.6, the trunk shape modelling approach obtains the closest ratio to 1 with the lowest STD, which indicates achieving very close *tidal volume* and *main effort* scaling factors for both FVC and SVC tests. This implies that we can calibrate the entire volume–time data using only the *tidal volume* scaling factor which is more reliable and less affected by trunk motion.

7.4 Conclusion

We proposed a vision-based trunk-motion tolerant approach for estimating the respiratory data within forced and slow vital capacity tests. Our approach corrects the body motion artifacts to a remarkable extent which not only improves the accuracy of the depth-based volume–time data, but also increases the intra-subject calibration process accuracy. After constructing a dynamic 3-D model of a subject's trunk during a breathing test, temporal geometrical features are extracted from the entire chest-wall and posterior-wall shapes. Then, a feature matrix is created for the whole sequence from which the respiratory volume–time data is computed by performing PCA. The retrieved respiratory data is then calibrated by the average of training scaling factors.

Evaluation on the dual-Kinect PFT dataset verifies that the trunk shape modelling approach achieves more accurate volume–time data than the single-Kinect and dPPG methods when calibrating by only the *tidal volume* scaling factor, or *tidal volume* and *main effort* scaling factors separately. In particular, the trunk shape modelling method is 3.4 times more accurate than dPPG with only *tidal volume* scaling factor calibration, and 2.7 times more accurate with the *tidal volume* and *main effort* scaling factor calibration. Our new averaged based calibration method is more accurate and efficient than using similarity based metrics, and it also attains closer *tidal volume* and *main effort* scaling

factors, compared to the single-Kinect and dPPG methods.

The trunk shape modelling complexity mainly lies in the trunk reconstruction stage as explained in Section 6.5 with negligible computational costs for the feature extraction and PCA computation stages. As long as the breathing test is performed based on ATS/ERS guidelines in which subjects must perform the test in a sitting posture without any significant non-breathing-related body movements prevented by the breathing test protocols, e.g., standing up or walking, our shape modelling approach is able to filter the trunk and shoulder breathing-related motions to a great extent. However, our model is prone to any kind of chest-wall and posterior region occlusion. This can be occlusion caused by external objects or self-occlusion due to the subject over-bending or trunk sideways movement during the deep forced exhalation. While deep learning methods could be investigated for constructing a more robust trunk shape model, they demand a considerable amount of body shape data from various stages of breathing per subject.

Chapter 8

Conclusions and Future Work

In this chapter, we present a summary of the thesis and outline our main contributions to the remote respiratory assessment field. This is then followed by providing some directions and ideas for future works.

8.1 Thesis Summary

We have proposed remote depth-based methods for clinical PFT using low-cost and publicly available depth sensors, such as the Microsoft Kinect V2.

In Chapter 1, we outlined the concepts of clinical PFT using spirometry [129] and introduced its two main protocols, i.e., FVC and SVC, as the key techniques in diagnosis of restrictive and obstructive lung diseases. Motivated to address the spirometry drawbacks as a contact-based PFT technique, we set out our vision-based approaches for estimating clinical lung measures remotely. In Chapter 2, we presented a review of scene depth acquisition and registration works and considered previous related works in remote respiratory sensing.

In Chapter 3, we described the specifications of our single-Kinect and dual-Kinect PFT datasets. In summary, the single-Kinect dataset comprises 529 sequences collected from 85 patients attending Southmead Hospital Bristol for respiratory test. In the dual-Kinect dataset, we collected 300 PFT sequences from 35 subjects based on ATS/ERS [128, 129] guidelines, similar to the single-Kinect dataset, but using two opposing depth sensors.

We presented our single-Kinect approach for PFT in Chapter 4. Using a depth sensor, the respiratory volume–time data was estimated by measuring the chest-wall volume variation in a PFT sequence. Calibrating this data to present the real amount of exchanged air, we computed 11 clinical FVC and SVC measures. We evaluated this approach by comparing our results to the ground truth measures obtained directly from the spirometer software. The evaluation results showed a high correlation between our depth-based and the spirometer measures when the subject's trunk motion artifacts are minimised during the test.

In Chapter 5, we explained our 3-D data acquisition and registration pipeline. Two opposing depth sensors were calibrated and synchronised for reconstructing an almost complete 3-D model of dynamic objects with precise temporal and spatial alignment accuracy. Quantitative and qualitative validations on known-size rigid boxes and non-rigid objects showed a high accuracy in spatial registration and temporal synchronisation, respectively. While this is a general data acquisition pipeline, we specifically used it for collecting our dual-Kinect PFT dataset.

In Chapter 6, our dPPG approach for PFT was presented. Exploring our single-Kinect results and analysis, we noticed that many subjects had involuntary respiratory-related trunk motion reaction to deep forced inhalation–exhalation. These motion artifacts affected the acquired respiratory data and consequently decreased the accuracy of computed measures in our single-Kinect method. Motived to solve this, we reconstructed the 3-D model of the subject's trunk during the test and decoupled the trunk and the respiratory motions by subtracting the average depth of the chest-wall from the posterior-wall in each frame of the sequence. Calibrating the retrieved data, PFT measures were then computed. We evaluated this approach on the dual-Kinect dataset which showed a great improvement in the dPPG measures compared to the single-Kinect approach.

As the last part of our work, we presented a trunk shape modelling approach for correcting body motion artifacts in Chapter 7. Although dPPG gained a notable improvement in the computed PFT measures compared to the single-Kinect method, it was not very effective in correcting the complex trunk motion, particularly at *main effort* inhalation-exhalation. To address this issue, we proposed the trunk shape modelling approach which computes the respiratory data by applying PCA on the temporal 3-D geometrical features extracted from the chest and posterior shape models. Evaluating on the dual-Kinect PFT dataset, we validated this approach in the signal level by computing several curve similarity measures between the computed and spirometer volume–time data. Comparing to the single-Kinect and dPPG methods, the trunk shape modelling

approach shows a significant improvement in the estimated volume–time data for both of *tidal volume*, and *tidal volume* & *main effort* individual calibration.

8.2 Contributions

The main contributions of this thesis are:

- Presenting a novel remote PFT approach using a single Kinect for estimating 11 clinically approved PFT measures.
- Developing an accurate and efficient 3-D data acquisition and registration pipeline using only two opposing RGB-D sensors.
- Improving PFT measures accuracy by decoupling trunk motion and chest-surface respiratory motion.
- Correcting complex trunk motion artifacts by proposing an active trunk shape modelling approach.
- Releasing the dual-Kinect PFT dataset, which is the first publicly available dataset of such kind (to the best of author knowledge) for other researchers to use.

8.3 Clinicians Point of View

The work presented in this thesis is a huge step forward in the development of remote monitoring of patients with respiratory diseases. As elaborated in [175, 176], our 'real world' clinical data, collected from a large group of patients with a wide range of lung function, is unique, and given the correlation and Bland-Altman analysis results, we are able to accurately obtain respiratory measures remotely, which has potential clinical applications for monitoring of patients in the hospital inpatients, gating (timing) of thoracic imaging and synchronisation with ventilatory support. Also, this work has the potential to be implemented in residential settings with no specialist requirement. With only a limited dependency to the spirometer, i.e. to build our model per subject using only a few training manoeuvres, our work is specifically useful for monitoring of pulmonary function without the continuous use of the spirometer which is of significant hindrance for frail elderly, children, and physically and cognitively-impaired patients. In

summary, in this work, we have taken a vital step towards applying remote sensing as an independent surrogate for spirometry.

8.4 Directions for Future Work

Our current remote vision-based PFT work and dataset opens up new opportunities for other researchers to extend the current work and also investigate other remote respiratory assessment methods, or develop the related frameworks such as remote respiratory monitoring in the wild. Thus, in this section several technical directions and novel ideas are provided for further extensions and future developments.

For a remote PFT approach, it would be preferable to straightly perform on any subject without prior intra-subject training requirement or support from other devices, e.g., spirometer. It means the need for subject-specific scaling factor for calibration shall be addressed. The main step toward this would be applying machine learning techniques to train on depth and ground truth respiratory data of male and female subjects with a wide range of ages, weights, heights, upper-body sizes and shapes, ethnicities, respiratory physiologies and lung pathologies. With the recent advances in deep learning methods, this is possible, however, it requires a considerable amount of training data.

Instead of filtering trunk motion by reconstructing the subject's body model using depth data from two sensors, applying single-view spatio-temporal deformable shape models [53, 177, 205] may be investigated. For this, several key trunk templates can be built for a subject's trunk in various sizes, i.e., different inhalation and exhalation stages, using a single-Kinect fusion method [96, 134]. Then, these models can be temporally fit to the subject's trunk shape during the test to help with decoupling the trunk and respiratory motions.

Although external respiratory motion is reflected on both thoracic and abdominal regions, it can be dominant in one of these regions depending on the subject's respiratory system physiology. Using the breathing data of the region with the dominant respiratory reflection, can potentially obtain more accurate depth-based volume–time data. Moreover, we note that in different subjects, thoracic and abdominal regions contribute differently in the *tidal volume* breathing, and the *main effort* inhalation–exhalation. Hence, a multi-patch linear regression model may be investigated to help with both of these considerations for further improving the depth-based volume–time data and

calibration scaling factors.

Finally, deep learning techniques may be investigated for retrieving more reliable depth-based respiratory data and measures. The Long Short Term Memory (LSTM) recurrent network has proven to be applicable in time series regression and prediction [76, 151, 153]. Taking the advantage of LSTM network in learning short-term patterns in long period of time, and also its regression output, we can train it on a subject's trunk surface 3-D temporal data along with the ground truth air flow information from a spirometer. Next, this trained network can be used for on-line generation of respiratory data and lung measures directly from the depth measurements. This method not only may provide more accurate respiratory data, but also would eliminate the volume–time data analysis for keypoints extraction and scaling factors computation for calibration.

References

[1] Asus Xtion PRO LIVE. [Online]: https://goo.gl/vcbJQA. Accessed: 03-11-2017. 14

[2] Spirometry. Measuring Ventilation with a Spirometer. [Online]: https://goo.gl/BvrMuE. Accessed: 13-10-2017. vii, 3

[3] Microsoft Kinect for Windows. [Online]: https://goo.gl/bJMhk6. Accessed: 06-09-2017. 7, 14

[4] 3d data acquisition and registration using two opposing kinects. [Online]: https://goo.gl/ndGt31. Accessed: 06-05-2018. 44

[5] The economic burden of lung disease. [Online]: https://goo.gl/VG8t4y. Accessed: 18-12-2017. 1

[6] Geomagic studio 2012. [Online]: https://goo.gl/CucNQ4. Accessed: 31-03-2018. 24, 96

[7] Global strategy for the diagnosis, management and prevention of COPD, global initiative for Chronic Obstructive Lung Disease (GOLD) 2017. [Online]: http://goldcopd.org. Accessed: 28-12-2017. 1

[8] British Lung Foundation. Chronic obstructive pulmonary disease (COPD) statistics. [Online]: https://goo.gl/Lo1UFe. Accessed: 13-10-2017. 1

[9] Single kinect data acquisition. [Online]: https://goo.gl/1aT2fo. Accessed: 06-05-2018. 42

[10] Real-time tidal volume estimation using iso-surface reconstruction. [Online]: https://goo.gl/a8Esp8. Accessed: 03-05-2018. 25

[11] Department of Health. Facts about COPD 2010. [Online]: https://goo.gl/i8KPJk. Accessed: 13-10-2017. 1

[12] Healthcare Commission. Clearing the air: A national study of chronic obstructive pulmonary disease. 2006. 1

[13] H. Afzal, D. Aouada, D. Font, B. Mirbach, and B. Ottersten. Rgb-d multi-view system calibration for full 3d scene reconstruction. In *International Conference on Pattern Recognition (ICPR)*, pages 2459–2464, 2014. 83, 84

[14] D. Alexiadis, D. Zarpalas, and P. Daras. Fast and smooth 3d reconstruction using multiple rgb-depth sensors. In *IEEE Conference on Visual Communications and Image Processing Conference*, pages 173–176, 2014. 83, 84

[15] D. S. Alexiadis, D. Zarpalas, and P. Daras. Real-time, full 3-d reconstruction of moving foreground objects from multiple consumer depth cameras. *IEEE Transactions on Multimedia*, 15(2):339–358, 2013. 83, 84

[16] D. G. Aliaga and Y. Xu. A self-calibrating method for photogeometric acquisition of 3d objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4):747–754, 2010. 13, 15

[17] A. Aliverti and A. Pedotti. Opto-electronic plethysmography. *Monaldi archives for chest disease*, 59:12–16, 2003. 29

[18] M. R. Alnowam, E. Lewis, K. Wells, and M. Guy. Respiratory motion modelling and prediction using probability density estimation. In *Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE*, pages 2465–2469, 2010. 6, 7, 29, 37

[19] M. Alnowami, E. Lewis, K. Wells, and M. Guy. Inter-and intra-subject variation of abdominal vs. thoracic respiratory motion using kernel density estimation. In *Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE*, pages 2921–2924, 2010. 6, 7, 29, 30, 37

[20] I. Amidror. Scattered data interpolation methods for electronic imaging systems: a survey. *Journal of electronic imaging*, 11(2):157–176, 2002. 119

[21] H. Aoki, Y. Takemura, K. Mimura, H. Aoki, and M. Nakajima. A non-contact and non-restricting respiration monitoring method for a sleeping person with a fiber-grating optical sensor. *Sleep and Biological Rhythms*, 1(3):249–250, 2003. 6, 7, 29, 30, 37

[22] H. Aoki, M. Miyazaki, H. Nakamura, R. Furukawa, R. Sagawa, and H. Kawasaki. Non-contact Respiration Measurement Using Structured Light 3-D Sensor. In *Proceedings of SICE Annual Conference (SICE)*, pages 614–618, 2012. 6, 7, 20, 23, 26, 28, 37

[23] B. Aronov, S. Har-Peled, C. Knauer, Y. Wang, and C. Wenk. Fréchet distance for curves, revisited. In *ESA*, volume 4168, pages 52–63. Springer, 2006. 116, 122, 123

[24] K. Ashutosh, R. Gilbert, J. Auchincloss, J. Erlebacher, and D. Peppi. Impedance pneumotachograph and magnetometer methods for monitoring. 37:964–966, 01 1975. 1

[25] R. Avetisyan, M. Willert, S. Ohl, and O. Staadt. Calibration of depth camera arrays. In *Proceedings of SIGRAD, Visual Computing*, pages 41–48, 2014. 17

[26] M. Bartula, T. Tigges, and J. Muehlsteff. Camera-based system for contactless monitoring of respiration. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2672–2675, 2013. 6, 7, 20, 21, 36

[27] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008. 16

[28] S. Beck and B. Froehlich. Volumetric calibration and registration of multiple rgbd-sensors into a joint coordinate system. In *IEEE Symposium on 3D User Interfaces*, pages 89–96, 2015. 17, 18, 79, 83

[29] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):616–625, 2013. 83

[30] F. Benetazzo, A. Freddi, A. Monteriu, and S. Longhi. Respiratory rate detection algorithm based on RGB-D camera: theoretical background and experimental results. *Healthcare Technology Letters*, 1(3):81–86, 2014. 6, 7, 20, 23, 28, 37, 38, 53

[31] K. Berger, K. Ruhl, Y. Schroeder, C. Bruemmer, A. Scholz, and M. A. Magnor. Markerless motion capture using multiple color-depth sensors. In *VMV*, pages 317–324, 2011. 14

[32] N. Bernacchia, L. Scalise, L. Casacanditella, I. Ercoli, P. Marchionni, and E. P. Tomasini. Noncontact measurement of heart and respiration rates based on kinect. In *International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–5, 2014. 6, 7, 20, 23, 28, 37, 53

[33] E. A. Bernal, L. K. Mestha, and E. Shilla. Noncontact monitoring of respiratory function via depth sensing. In *International Conference on Biomedical and Health Informatics (BHI)*, pages 101–104, 2014. 6, 7, 20, 23, 26, 37

[34] P. Bert. *La pression barométrique*. G. Masson, 1878. 2

[35] P. J. Besl and H. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992. 16, 82, 85

[36] P. Bishop. A bibliography of john hutchinson. *Medical history*, 21(4):384, 1977. 3

[37] J. M. Bland and D. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476):307–310, 1986. 106

[38] J. Bogaard, S. Overbeek, A. Verbraak, C. Vons, H. Folgering, T. W. van der Mark, C. Roos, and P. Sterk. Pressure-volume analysis of the lung with an exponential and linear-exponential model in asthma and copd. dutch cnsld study group. *European Respiratory Journal*, 8(9):1525–1531, 1995. 32

[39] M. Bracci, E. Strafella, N. Croce, S. Staffolani, A. Carducci, M. Verani, M. Valentino, and L. Santarelli. Risk of bacterial cross infection associated with inspiration through flow-based spirometers. *American journal of infection control*, 39(1):50–55, 2011. 6

[40] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 88

[41] T. Breuer, C. Bodensteiner, and M. Arens. Low-cost commodity depth sensor comparison and accuracy analysis. In *Electro-Optical Remote Sensing, Photonic Technologies, and Applications VIII; and Military Applications in Hyperspectral Imaging and High Spatial Resolution Sensing II*, volume 9250, page 92500G, 2014. 40

[42] J. D. Bronzino. *Biomedical engineering handbook*, volume 2. CRC press, 1999. 1

[43] D. C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37 (8):855–866, 1971. 84

[44] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *European conference on computer vision*, pages 25–36, 2004. 23

[45] N. Burba, M. Bolas, D. M. Krum, and E. a. Suma. Unobtrusive measurement of subtle nonverbal behaviors with the microsoft kinect. In *IEEE Virtual Reality Short Papers and Posters*, pages 1–4, 2012. 53

[46] A. Butler, S. Izadi, O. Hilliges, D. Molyneaux, S. Hodges, and D. Kim. Shakensense: Reducing structured light interference when multiple depth cameras overlap. *Proc. Human Factors in Computing Systems (ACM CHI). New York, NY, USA: ACM*, 2012. 14

[47] S. J. Cala, C. M. Kenyon, G. Ferrigno, P. Carnevali, A. Aliverti, A. Pedotti, P. T. Macklem, and D. F. Rochester. Chest wall and lung volume estimation by optical reflectance motion analysis. *Journal of Applied Physiology*, 81(6):2680–2689, 1996. 6, 7, 29, 37, 38

[48] M. Camplani, T. Mantecón, and L. Salgado. Depth-color fusion strategy for 3-d scene modeling with kinect. *IEEE Transactions on Cybernetics*, 43(6):1560–1571, 2013. 49

[49] J. H. Challis. A procedure for determining rigid body transformation parameters. *Journal of biomechanics*, 28(6):733–737, 1995. 86

[50] A. Chatterjee, A. P. Prathosh, and P. Praveena. Real-time respiration rate measurement from thoracoabdominal movement with a consumer grade camera. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2708–2711, 2016. vii, 6, 7, 20, 21, 22, 36

[51] A. Chatterjee, A. P. Prathosh, P. Praveena, and V. Upadhya. Real-time visual respiration rate estimation with dynamic scene adaptation. In *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 154–160, 2016. 23, 36

[52] A. Chatterjee, A. P. Prathosh, P. Praveena, and V. Upadhya. A vision based method for real-time respiration rate estimation using a recursive fourier analysis. In *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 143–149, 2016. 23, 36

[53] Y. Chen, T.-K. Kim, and R. Cipolla. Inferring 3d shapes and deformations from single views. In *European Conference on Computer Vision*, pages 300–313. Springer, 2010. 132

[54] C. F. Clarenbach, O. Senn, T. Brack, M. Kohler, and K. E. Bloch. Monitoring of ventilation during exercise by a portable respiratory inductive plethysmograph. *CHEST Journal*, 128(3):1282–1290, 2005. 1

[55] M. A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris. Respiratory rate: the neglected vital sign. *Medical Journal of Australia*, 188(11): 657–659, 2008. 54, 99, 121

[56] C. Criée, S. Sorichter, H. Smith, P. Kardos, R. Merget, D. Heise, D. Berdel, D. Köhler, H. Magnussen, W. Marek, H. Mitfessel, K. Rasche, M. Rolke, H. Worth, and R. Jörres. Body plethysmography–its principles and clinical use. *Respiratory medicine*, 105(7):959–971, 2006. 1, 2

[57] M. de Berg, O. Cheong, M. van Kreveld, and M. Overmars. *Computational Geometry: Algorithms and Applications*. Springer-Verlag, 3rd edition, 2008. 51, 91

[58] W. de Boer, J. Lasenby, J. Cameron, R. Wareham, S. Ahmad, C. Roach, W. Hills, and R. Iles. Slp: A zero-contact non-invasive method for pulmonary function testing. In *Proceedings of the British Machine Vision Conference*, pages 85.1–85.12, 2010. 6, 7, 20, 23, 26, 28, 37, 92

[59] R. L. Dellacá, A. Aliverti, P. Pelosi, E. Carlesso, D. Chiumello, A. Pedotti, and L. Gattinoni. Estimation of end-expiratory lung volume variations by optoelectronic plethysmography. *Critical care medicine*, 29(9):1807–1811, 2001. 6, 7, 29, 37, 38

[60] T. Deng, J. C. Bazin, T. Martin, C. Kuster, J. Cai, T. Popa, and M. Gross. Registration of multiple rgbd cameras via local rigid transformations. In *International Conference on Multimedia and Expo*, pages 1–6, 2014. 17, 18, 79

[61] D. Desjardins and P. Payeur. Dense stereo range sensing with marching pseudorandom patterns. In *Computer and Robot Vision, 2007. CRV'07. Fourth Canadian Conference on*, pages 216–226. IEEE, 2007. 14

[62] A. B. DuBois, S. Y. Botelho, G. N. Bedell, R. Marshall, and J. H. Comroe Jr. A rapid plethysmographic method for measuring thoracic gas volume: a comparison with a nitrogen washout method for measuring functional residual capacity in normal subjects. *Journal of Clinical Investigation*, 35(3):322, 1956. 2

[63] A. B. DuBois, S. Y. Botelho, and J. H. Comroe Jr. A new method for measuring airway resistance in man using a body plethysmograph: values in normal subjects and in patients with respiratory disease. *Journal of Clinical Investigation*, 35(3): 327, 1956. 2

[64] N. G. Durdle, J. Thayyoor, and V. Raso. An improved structured light technique for surface reconstruction of the human trunk. In *Electrical and Computer Engineering, 1998. IEEE Canadian Conference on*, volume 2, pages 874–877. IEEE, 1998. 14

[65] C. H. Esteban, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):548–554, 2008. 15

[66] G. D. Evangelidis and E. Z. Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1858–1865, 2008. 14

[67] O. Faugeras, Q.-T. Luong, and T. Papadopoulo. *The geometry of multiple images: the laws that govern the formation of multiple images of a scene and some of their applications*. MIT press, 2004. 14

[68] J. Fei and I. Pavlidis. Analysis of breathing air flow patterns in thermal imaging. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 946–952, 2006. 53

[69] G. Ferrigno, P. Carnevali, A. Aliverti, F. Molteni, G. Beulcke, and A. Pedotti. Three-dimensional optical analysis of chest wall motion. *Journal of Applied Physiology*, 77(3):1224–1231, 1994. 6, 7, 29, 37, 38

[70] J. A. Fiz, M. Haro, J. Aguilar, J. Alvarez, J. Abad, E. Monso, and J. Morera. Spirometry and maximal respiratory pressures in patients with facial paralysis. *Chest*, 103(1):170–173, 1993. 6

[71] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002. 13, 49

[72] S. Fowler, R. Roush, J. Wise, and D. Stronck. *Concepts of Biology*. Openstax College, Rice University, 2013. vii, 5

[73] M. Frigola, J. Amat, and J. Pagès. Vision based respiratory monitoring system. In *Mediterranean Conference on Control and Automation*, pages 1–5, 2002. 6, 7, 20, 21, 22, 36

[74] Y. Furukawa and C. Hernndez. Multi-view stereo: A tutorial. *Foundations and Trends in Computer Graphics and Vision*, 9(1-2):1–148, 2015. 15

[75] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000. 14

[76] J. C. B. Gamboa. Deep learning for time-series analysis. *arXiv preprint arXiv:1701.01887*, 2017. 133

[77] W. F. Ganong and K. E. Barrett. *Review of medical physiology*. Appleton & Lange Norwalk, CT, 24 edition, 1995. 54

[78] GBD. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015. *The Lancet Respiratory Medicine*, 5(9):691–706, 2017. 1

[79] A. Geiger, F. Moosmann, . Car, and B. Schuster. Automatic camera and range sensor calibration using a single shot. In *International Conference on Robotics and Automation*, pages 3936–3943, 2012. 84

[80] J. Geng. Structured-light 3d surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2):128–160, 2011. 14

[81] R. B. George. *Chest medicine: essentials of pulmonary and critical care medicine*. Lippincott Williams & Wilkins, 2005. 31

[82] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 1996. 86

[83] P. M. Griffin, L. S. Narasimhan, and S. R. Yee. Generation of uniquely encoded light patterns for range data acquisition. *Pattern recognition*, 25(6):609–616, 1992. 14

[84] C. Gu and C. Li. Assessment of human respiration patterns via noncontact sensing using doppler multi-radar system. *Sensors*, 15(3):6383–6398, 2015. 6, 7, 29, 31, 37

[85] J. Han, L. Shao, D. Xu, and J. Shotton. Enhanced computer vision with microsoft kinect sensor: A review. *IEEE transactions on cybernetics*, 43(5):1318–1334, 2013. 14

[86] C. Hanning, H. Smith, and I. M. Ledingham. Clinical application of inductive plethysmography. *Scottish medical journal*, 23(4):310–311, 1978. 1

[87] M. Hansard, S. Lee, O. Choi, and R. P. Horaud. *Time-of-flight cameras: principles, methods and applications*. Springer Science & Business Media, 2012. 13, 14

[88] J. M. Harte, C. K. Golby, J. Acosta, E. F. Nash, E. Kiraci, M. A. Williams, T. N. Arvanitis, and B. Naidu. Chest wall motion analysis in healthy volunteers and adults with cystic fibrosis using a novel kinect-based motion tracking system. *Medical & biological engineering & computing*, 54(11):1631–1640, 2016. vii, 6, 20, 23, 24, 37, 38, 96

[89] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. 14

[90] R. Horaud, M. Hansard, G. Evangelidis, and C. Ménier. An overview of depth cameras and range scanners based on time-of-flight technologies. *Machine Vision and Applications*, 27(7):1005–1020, 2016. 13, 14, 15

[91] B. K. Horn and B. G. Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981. 22

[92] Y. C. Hsieh, D. M. McKeown Jr, and F. P. Perlant. Performance evaluation of scene registration and stereo matching for artographic feature extraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):214–238, 1992. 14

[93] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 454 (1971):903–995, 1998. 60, 100

[94] M. Innmann, M. Zollhöfer, M. Nießner, C. Theobalt, and M. Stamminger. Volumedeform: Real-time volumetric non-rigid reconstruction. In *European Conference on Computer Vision (ECCV)*, pages 362–379, 2016. 16, 17

[95] I. Ishii, K. Yamamoto, K. Doi, and T. Tsuji. High-speed 3d image acquisition using coded structured light projection. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 925–930. IEEE, 2007. 14

[96] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. In *ACM Symposium on User Interface Software and Technology*, pages 559–568, 2011. 16, 18, 79, 132

[97] R. Janssen, W. Wang, A. Moo, and G. de Haan. Video-based respiration monitoring with automatic region of interest detection. *Physiological Measurement*, 37(1), 2016. 6, 7, 20, 23, 36

[98] E. Jovanov, D. Raskovic, and R. Hormigo. Thermistor-based breathing sensor for circadian rhythm evaluation. *Biomedical sciences instrumentation*, 37:493–498, 2001. 1

[99] R. M. Kacmarek, J. K. Stoller, and A. Heuer. *Egan's Fundamentals of Respiratory Care-E-Book*. Elsevier Health Sciences, 2016. vii, 5

[100] B. Kainz, S. Hauswiesner, G. Reitmayr, M. Steinberger, R. Grasset, L. Gruber, E. Veas, D. Kalkofen, H. Seichter, and D. Schmalstieg. Omnikinect: Real-time dense volumetric data acquisition and applications. In *ACM symposium on Virtual reality software and technology*, pages 25–32, 2012. 9, 17, 18, 19, 80, 82, 84

[101] A. Kendrick, D. Johns, and J. Leeming. Infection control of lung function equipment: a practical approach. *Respiratory medicine*, 97(11):1163–1179, 2003. 6

[102] K. Khoshelham and S. O. Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454, 2012. 14

[103] B. Koeppen and B. Stanton. *Berne & Levy Physiology*. Elsevier, 6 edition, 2010. vii, 2

[104] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-flight cameras in computer graphics. In *Computer Graphics Forum*, volume 29, pages 141–159. Wiley Online Library, 2010. 40

[105] T. Kondo, T. Uhlig, P. Pemberton, and P. Sly. Laser monitoring of chest wall displacement. *European Respiratory Journal*, 10(8):1865–1869, 1997. 30

[106] M. Kowalski, J. Naruniec, and M. Daniluk. Livescan3d: A fast and inexpensive 3d data acquisition system for multiple kinect v2 sensors. In *International Conference on 3D Vision*, pages 318–325, 2015. vii, 9, 15, 17, 18, 19, 79, 80, 83, 84

[107] W. Krattenthaler, K. J. Mayer, and H. P. Duwe. 3d-surface measurement with coded light approach. In *Proceedings of the 17th Meeting of the Austrian Association for Pattern Recognition on Image Analysis and Synthesis*, pages 103–114, 1993. ISBN 3-85403-068-1. 14

[108] E. Lachat, H. Macher, T. Landes, and P. Grussenmeyer. Assessment and calibration of a rgb-d camera (kinect v2 sensor) towards a potential use for close-range 3d modeling. *Remote Sensing*, 7(10):13070–13097, 2015. 15

[109] I. Levine. *Physical chemistry*. McGraw-Hill, 6 edition, 2008. 2

[110] M. H. Li, A. Yadollahi, and B. Taati. A non-contact vision-based system for respiratory rate estimation. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2119–2122, 2014. 20, 22, 23, 36

[111] M. H. Li, A. Yadollahi, and B. Taati. Noncontact vision-based cardiopulmonary monitoring in different sleeping positions. *IEEE Journal of Biomedical and Health Informatics*, 21(5):1367–1375, 2017. 6, 7, 20, 22, 23, 36

[112] W. Li, B. Tan, and R. J. Piechocki. Non-contact breathing detection using passive radar. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–6, 2016. 6, 7, 29, 31, 37

[113] W. Li, B. Tan, and R. Piechocki. Passive radar for opportunistic monitoring in e-health applications. *IEEE Journal of Translational Engineering in Health and Medicine*, 6:1–10, 2018. 6, 7, 29, 31, 37

[114] J. Lim. Optimized projection pattern supplementing stereo systems. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 2823–2829. IEEE, 2009. 14

[115] K. Y. Lin, D. Y. Chen, and W. J. Tsai. Image-based motion-tolerant remote respiratory rate evaluation. *IEEE Sensors Journal*, 16(9):3263–3271, 2016. 6, 7, 20, 22, 36, 38

[116] M. Lindner, I. Schiller, A. Kolb, and R. Koch. Time-of-flight sensor calibration for accurate range sensing. *Computer Vision and Image Understanding*, 114(12): 1318–1328, 2010. 40

[117] C. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, pages 125–131. IEEE, 1999. 14

[118] J. L. López-Campos eta l. Global burden of copd. *Respirology*, 21(1):14–23, 2016. 1

[119] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. 16

[120] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'81, pages 674–679, 1981. 22

[121] T. Luk, J. Pik, and L. Chrenko. Contactless recognition of respiration phases using web camera. In *2014 24th International Conference Radioelektronika*, pages 1–4, 2014. 23, 36

[122] D. J. MacKay. A practical bayesian framework for backpropagation networks. *Neural computation*, 4(3):448–472, 1992. 24

[123] A. Maimone and H. Fuchs. Encumbrance-free telepresence system with real-time 3d capture and display using commodity depth cameras. In *IEEE International Symposium on Mixed and Augmented Reality*, pages 137–146, 2011. 83, 84

[124] T. Mallick, P. P. Das, and A. K. Majumdar. Characterizations of noise in kinect depth images: A review. *IEEE Sensors journal*, 14(6):1731–1740, 2014. 14

[125] R. Marani, G. Gelao, and A. G. Perri. A new system for continuous monitoring of breathing and kinetic activity. *Journal of Sensors*, 2010, 2010. 53

[126] M. Martinez and R. Stiefelhagen. Breath rate monitoring during sleep using near-ir imagery and pca. In *International Conference on Pattern Recognition (ICPR)*, pages 3472–3475, 2012. 6, 7, 20, 23, 29, 37

[127] J. McFadden, R. Price, H. Eastwood, and R. Briggs. Raised respiratory rate in elderly patients: a valuable physical sign. *British Medical Journal (Clinical Research Ed.)*, 284(6316):626–627, 1982. 54

[128] M. R. Miller, R. Crapo, J. Hankinson, V. Brusasco, F. Burgos, R. Casaburi, A. Coates, P. Enright, C. M. van der Grinten, P. Gustafsson, et al. General considerations for lung function testing. *European Respiratory Journal*, 26(1):153–161, 2005. 1, 6, 24, 38, 39, 41, 42, 45, 68, 72, 129

[129] M. R. Miller, J. Hankinson, V. Brusasco, F. Burgos, R. Casaburi, A. Coates, R. Crapo, P. Enright, C. P. M. van der Grinten, P. Gustafsson, R. Jensen, D. C. Johnson, N. MacIntyre, R. McKay, D. Navajas, O. F. Pedersen, R. Pellegrino, G. Viegi, and J. Wanger. Standardisation of spirometry. *The European Respiratory Journal*, 26(2):319–338, 2005. 1, 3, 4, 6, 38, 39, 41, 42, 45, 57, 68, 72, 122, 129

[130] S. Miller, A. Teichman, and S. Thrun. Unsupervised extrinsic calibration of depth sensors in dynamic scenes. In *International Conference on Intelligent Robots and Systems*, pages 2695–2702, 2013. 9, 17, 19, 79, 80

[131] S. D. Min, D. J. Yoon, S. W. Yoon, Y. H. Yun, and M. Lee. A study on a non-contacting respiration signal monitoring system using doppler ultrasound. *Medical & biological engineering & computing*, 45(11):1113–1119, 2007. 6, 7, 29, 31, 37

[132] S. D. Min, J. K. Kim, H. S. Shin, Y. H. Yun, C. K. Lee, and M. Lee. Noncontact respiration rate measurement system using an ultrasonic proximity sensor. *IEEE Sensors Journal*, 10(11):1732–1739, 2010. 6, 7, 29, 31, 37

[133] S. K. Mitra and Y. Kuo. *Digital signal processing: a computer-based approach*, volume 2. McGraw-Hill New York, 4 edition, 2006. 54

[134] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *International Symposium on Mixed and Augmented Reality*, pages 127–136, 2011. 16, 132

[135] R. A. Newcombe, D. Fox, and S. M. Seitz. Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. In *Computer Vision and Pattern Recognition (CVPR)*, pages 343–352, 2015. 16, 17, 79

[136] S. Ostadabbas, C. Bulach, D. N. Ku, L. J. Anderson, and M. Ghovanloo. A passive quantitative measurement of airway resistance using depth data. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5743–5747, 2014. vii, xi, 7, 20, 31, 32, 33, 34, 36, 37, 51, 53, 80, 92, 95, 105

[137] S. Ostadabbas, N. Sebkhi, M. Zhang, S. Rahim, L. J. Anderson, F. E. H. Lee, and M. Ghovanloo. A vision-based respiration monitoring system for passive airway resistance estimation. *IEEE Transactions on Biomedical Engineering*, 63(9):1904–1913, 2016. vii, 7, 20, 31, 34, 35, 36, 37, 53, 63, 80, 92, 95, 99

[138] J. Pages, J. Salvi, and C. Matabosch. Robust segmentation and decoding of a˜ grid pattern for structured light. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 689–696. Springer, 2003. 14

[139] D. Pagliari and L. Pinto. Calibration of kinect for xbox one and comparison between the two generations of microsoft sensors. *Sensors*, 15(11):27569–27589, 2015. xi, 15

[140] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand. Bilateral filtering: Theory and applications. *Foundations and Trends in Computer Graphics and Vision*, 4 (1):1–73, 2008. 49

[141] C. Park and B. Lee. Real-time estimation of respiratory rate from a photoplethysmogram using an adaptive lattice notch filter. *Biomedical engineering online*, 13 (1), 2014. 53

[142] P. Payeur and D. Desjardins. Structured light stereoscopic imaging with dynamic pseudo-random patterns. In *International Conference Image Analysis and Recognition*, pages 687–696. Springer, 2009. 14

[143] J. Penne, C. Schaller, J. Hornegger, and T. Kuwert. Robust real-time 3d respiratory motion detection using time-of-flight cameras. *Computer Assisted Radiology and Surgery*, 3(5):427–431, 2008. 6, 7, 23, 27, 37

[144] E. Peper, G. H. Groshans, J. Johnston, R. Harvey, and F. Shaffer. Calibrating respiratory strain gauges: What the numbers mean for monitoring respiration. *Biofeedback*, 44(2):101–105, 2016. 1

[145] E. Petriu, Z. Sakr, H. Spoelder, and A. Moica. Object recognition using pseudo-random color encoded structured light. In *Instrumentation and Measurement Technology Conference, 2000. IMTC 2000. Proceedings of the 17th IEEE*, volume 3, pages 1237–1241. IEEE, 2000. 14

[146] E. Pflüger. Das pneumonometer. *Archiv für die gesamte Physiologie des Menschen und der Tiere*, 29(1):244–246, 1882. 2

[147] C. Pheatt. Intel threading building blocks. *Journal of Computing Sciences in Colleges*, 23(4):298–298, 2008. 88

[148] R. Pierce. Spirometry: an essential clinical measurement. *Australian family physician*, 34(7):535–539, 2005. 4

[149] M. Z. Poh, D. J. McDuff, and R. W. Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7–11, 2011. 20, 23, 36

[150] F. Pomerleau, F. Colas, R. Siegwart, and S. Magnenat. Comparing icp variants on real-world data sets. *Autonomous Robots*, 34(3):133–148, 2013. 16

[151] S. C. Prasad and P. Prasad. Deep recurrent neural networks for time series prediction. *arXiv preprint arXiv:1407.5949*, 2014. 133

[152] A. P. Prathosh, P. Praveena, L. K. Mestha, and S. Bharadwaj. Estimation of respiratory pattern from video using selective ensemble aggregation. *IEEE Transactions on Signal Processing*, 65(11):2902–2916, 2017. 23, 36

[153] Y. Qin, D. Song, H. Cheng, W. Cheng, G. Jiang, and G. Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971*, 2017. 133

[154] K. F. Rabe, S. Hurd, A. Anzueto, P. J. Barnes, S. A. Buist, P. Calverley, Y. Fukuchi, C. Jenkins, R. Rodriguez-Roisin, C. Van Weel, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: Gold executive summary. *American Journal of Respiratory and Critical Care Medicine*, 176(6):532–555, 2007. xi, 4, 5

[155] A. Rafighi, S. Seifi, and O. Meruvia-Pastor. Automatic and adaptable registration of live rgbd video streams. In *ACM SIGGRAPH Conference on Motion in Games*, pages 243–250, 2015. 9, 17, 19, 79, 80

[156] B. Reyes, N. Reljin, Y. Kong, Y. Nam, and K. Chon. Tidal volume and instantaneous respiration rate estimation using a smartphone camera. *IEEE Journal of Biomedical and Health Informatics*, 21(3):764–777, 2017. 6, 7, 20, 21, 36

[157] S. Rihana, E. Younes, D. Visvikis, and H. Fayad. Kinect2 – respiratory movement detection study. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3875–3878, 2016. 6, 7, 20, 23, 28, 37

[158] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf. In *International Conference on Computer Vision (ICCV)*, pages 2564–2571, 2011. 19

[159] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-time 3d model acquisition. *ACM Transactions on Graphics (TOG)*, 21(3):438–446, 2002. 15

[160] R. B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (fpfh) for 3d registration. In *International Conference on Robotics and Automation*, pages 3212–3217, 2009. 16

[161] D. R. Rutala, W. A. Rutala, D. J. Weber, and C. A. Thomann. Infection risks associated with spirometry. *Infection Control & Hospital Epidemiology*, 12(2):89–92, 1991. 6

[162] F. Sadlo, T. Weyrich, R. Peikert, and M. Gross. A practical structured light acquisition system for point-based geometry and texture. In *IEEE VGTC Symposium Point-Based Graphics*, pages 89–145, 2005. 13, 15

[163] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern recognition*, 37(4):827–849, 2004. 14

[164] H. Sarbolandi, D. Lefloch, and A. Kolb. Kinect range sensing: Structured-light versus time-of-flight kinect. *Computer vision and image understanding*, 139:1–20, 2015. xi, 14, 15

[165] I. Sato and M. Nakajima. Non-contact breath motion monitoring system in full automation. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3448–3451, 2005. 6, 7, 29, 30, 37

[166] L. Scalise, M. De Melis, U. Morbiducci, P. Segers, and E. Tomasini. From cardiac to respiratory rate, from cardiac sounds to pulse velocity: a noncontact unified approach for the monitoring of vital signs by means of optical vibrocardiography. In *Eighth International Conference on Vibration Measurements by Laser Techniques: Advances and Applications*, volume 7098, page 70980S, 2008. 6, 7, 29, 30, 31, 37

[167] L. Scalise, P. Marchionni, and I. Ercoli. Optical method for measurement of respiration rate. In *International Workshop on Medical Measurements and Applications*, pages 19–22, 2010. 6, 7, 29, 30, 31, 37

[168] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47 (1-3):7–42, 2002. 14

[169] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2003. 14

[170] W. J. Schroeder, K. Martin, and B. Lorensen. *The visualization toolkit : an object-oriented approach to 3D graphics*. Kitware, 2006. 88

[171] S. Seifi, A. Rafighi, and O. Meruvia-Pastor. Derees: Real-time registration of rgbd images using image-based feature detection and robust 3d correspondence estimation and refinement. In *International Conference on Image and Vision Computing*, pages 136–141, 2014. 9, 17, 19, 79, 80

[172] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Computer Vision and Pattern Recognition (CVPR)*, pages 519–528, 2006. 15

[173] T. Seppanen, J. Kananen, K. Noponen, O.-P. Alho, and T. Seppanen. Accurate measurement of respiratory airflow waveforms using depth data. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 7857–7860, 2015. 6, 7, 20, 23, 25, 37

[174] D. Shao, Y. Yang, C. Liu, F. Tsow, H. Yu, and N. Tao. Noncontact monitoring breathing pattern, exhalation flow rate and pulse transit time. *IEEE Transactions on Biomedical Engineering*, 61(11):2760–2767, 2014. 6, 7, 20, 22, 36

[175] C. Sharp, V. Soleimani, S. Hannuna, M. Camplani, D. Damen, J. Viner, M. Mirmehdi, and J. Dodd. Remote pulmonary function testing–computer gaming in the respiratory world. *Thorax*, 70(Suppl 3):A117–A118, 2015. 10, 131

[176] C. Sharp, V. Soleimani, S. Hannuna, M. Camplani, D. Damen, J. Viner, M. Mirmehdi, and J. Dodd. Towards respiratory assessment using depth measurements from a time-of-flight sensor. *Frontiers in Physiology*, 8(65), 2017. 10, 131

[177] S. Shunta, M. Kouchi, M. Mochimaru, and Y. Aoki. Building 3d deformable body model and torso shape estimation system. In *2nd International Digital Human Modeling Symposium*, 2013. 132

[178] M. Siudak and P. Rokita. A survey of passive 3d reconstruction methods on the basis of more than one image. *Machine Graphics and Vision*, 23(3/4):57–117, 2014. 15

[179] S. W. Smith. *Digital signal processing: a practical guide for engineers and scientists*. Newnes, 2003. 54

[180] V. Soleimani, M. Mirmehdi, D. Damen, S. Hannuna, M. Camplani, J. Viner, and J. Dodd. Remote pulmonary function testing using a depth sensor. In *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2015. 8, 80

[181] V. Soleimani, M. Mirmehdi, D. Damen, S. Hannuna, and M. Camplani. 3d data acquisition and registration using two opposing kinects. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 128–137, 2016. 9

[182] V. Soleimani, M. Mirmehdi, D. Damen, J. Dodd, S. Hannuna, C. Sharp, M. Camplani, and J. Viner. Remote, depth-based lung function assessment. *IEEE Transactions on Biomedical Engineering*, 64(8):1943–1958, 2017. 8, 80

[183] V. Soleimani, M. Mirmehdi, D. Damen, M. Camplani, S. Hannuna, J. Dodd, and C. Sharp. Depth-based whole body photoplethysmography in remote pulmonary function testing. *IEEE Transactions on Biomedical Engineering*, 65(6):1421–1431, 2018. 9

[184] E. A. Spriggs. John hutchinson, the inventor of the spirometer–hisnorthcountrybackground, life in london, and scientific achievements. *Medical history*, 21(4):357, 1977. 3

[185] E. A. Spriggs. The history of spirometry. *British journal of diseases of the chest*, 72:165–180, 1978. 3

[186] R. Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010. 13

[187] F. Tahavori, M. Alnowami, and K. Wells. Marker-less respiratory motion modeling using the microsoft kinect for windows. In *Proceedings SPIE 9036, Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling*, pages 1–10, 2014. vii, 6, 7, 20, 23, 25, 26, 37

[188] K. Tan, R. Saatchi, H. Elphick, and D. Bruke. Real-time vision based respiration monitoring system. In *International Symposium on Communication Systems Networks and Digital Signal Processing (CSNDSP)*, pages 770–774, 2010. 6, 7, 20, 21, 22, 36

[189] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological Measurement*, 35(5):807, 2014. 23, 36

[190] E. Tola, V. Lepetit, and P. Fua. A fast local descriptor for dense matching. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008. 14

[191] P. H. S. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78: 138–156, 2000. 89

[192] S. Transue, P. Nguyen, T. Vu, and M. H. Choi. Real-time tidal volume estimation using iso-surface reconstruction. In *International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pages 209–218, 2016. vii, 6, 7, 20, 23, 24, 25, 37

[193] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *International Workshop on Vision Algorithms: Theory and Practice*, pages 298–372, 2000. 17

[194] A. O. Ulusoy, F. Calakli, and G. Taubin. Robust one-shot 3d scanning using loopy belief propagation. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 15–22. IEEE, 2010. 14

[195] V. Vaish, M. Levoy, R. Szeliski, C. L. Zitnick, and S. B. Kang. Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2331–2338. IEEE, 2006. 14

[196] R. J. Valkenburg and A. M. McIvor. Accurate 3d measurement using a structured light system. *Image and Vision Computing*, 16(2):99–110, 1998. 14

[197] T. Vos, R. M. Barber, B. Bell, A. Bertozzi-Villa, S. Biryukov, I. Bolliger, F. Charlson, A. Davis, L. Degenhardt, D. Dicker, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: A systematic analysis for the global burden of disease study 2013. *The Lancet*, 386(9995):743–800, 2015. 1

[198] C.-C. Wang, S. B. Trivedi, F. Jin, S. Stepanov, Z. Chen, J. Khurgin, P. Rodriguez, and N. S. Prasad. Human life signs detection using high-sensitivity pulsed laser vibrometer. *IEEE Sensors Journal*, 7(9):1370–1376, 2007. 6, 7, 29, 30, 31, 37

[199] O. Wasenmüller and D. Stricker. Comparison of kinect v1 and v2 depth images in terms of accuracy and precision. In *Asian Conference on Computer Vision*, pages 34–45. Springer, 2016. vii, 16

[200] M. Weinmann, R. Ruiters, A. Osep, C. Schwartz, and R. Klein. Fusing structured light consistency and helmholtz normals for 3d reconstruction. In *British Machine Vision Conference*, pages 1–12, 2012. 15

[201] S. Wiesner and Z. Yaniv. Monitoring patient respiration using a single optical camera. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2740–2743, 2007. 23, 36

[202] U. Wijenayake and S.-Y. Park. Pca based analysis of external respiratory motion using an rgb-d camera. In *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6, 2016. 6, 20, 23, 27, 37, 92

[203] U. Wijenayake and S.-Y. Park. Real-time external respiratory motion measuring technique using an rgb-d camera and principal component analysis. *Sensors*, 17 (8), 2017. vii, 6, 7, 20, 23, 27, 37, 92

[204] J. Xia and R. A. Siochi. A real-time respiratory motion monitoring system using kinect: proof of concept. *Medical Physics*, 39(5):2682–2685, 2012. 6, 7, 20, 23, 27, 29, 37, 53

[205] W. Xu, M. Salzmann, Y. Wang, and Y. Liu. Deformable 3d fusion: From partial dynamic 3d observations to complete 4d models. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2183–2191, 2015. 132

[206] M.-C. Yu, J.-L. Liou, S.-W. Kuo, M.-S. Lee, and Y.-P. Hung. Noncontact respiratory measurement of volume change using depth camera. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2371–2374, 2012. 6, 7, 20, 23, 28, 37, 63, 80, 92

[207] P. Zanuttigh, G. Marin, C. Dal Mutto, F. Dominio, L. Minto, and G. M. Cortelazzo. *Time-of-flight and structured light depth cameras*. Springer, 2016. 14, 15

[208] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International journal of computer vision*, 13(2):119–152, 1994. 16, 85

[209] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International journal of computer vision*, 27(2):161–195, 1998. 14

[210] Z. Zhang. Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2):4–10, 2012. 14

[211] F. Zhao, M. Li, Y. Qian, and J. Z. Tsien. Remote measurements of heart and respiration rates for telemedicine. *PLOS ONE*, 8(10):1–14, 10 2013. 23, 36

[212] M. Zollhöfer, M. Nießner, S. Izadi, C. Rehmann, C. Zach, M. Fisher, C. Wu, A. Fitzgibbon, C. Loop, C. Theobalt, et al. Real-time non-rigid reconstruction using an rgb-d camera. *ACM Transactions on Graphics (TOG)*, 33(4):156, 2014. 16, 79