# Augmented Reality-Based Lung Ultrasound Scanning Guidance

Keshav Bimbraw(✉), Xihan Ma, Ziming Zhang, and Haichong Zhang

Worcester Polytechnic Institute, Worcester, MA, USA
{kbimbraw,xma4,zzhang15,hzhang10}@wpi.edu

**Abstract.** Lung ultrasound (LUS) is an established non-invasive imaging method for diagnosing respiratory illnesses. With the rise of SARS-CoV-2 (COVID-19) as a global pandemic, LUS has been used to detect pneumopathy for triaging and monitoring patients who are diagnosed or suspected with COVID-19 infection. While LUS offers a cost-effective, radiation-free, and higher portability compared with chest X-ray and CT, its accessibility is limited due to its user dependency and the small number of physicians and sonographers who can perform appropriate scanning and diagnosis. In this paper, we propose a framework of guiding LUS scanning featuring augmented reality, in which the LUS procedure can be guided by projecting the scanning trajectory on the patient's body. To develop such a system, we implement a computer vision-based detection algorithm to classify different regions on the human body. The DensePose algorithm is used to obtain body mesh data for the upper body pictured with a mono-camera. Torso sub-mesh is used to extract and overlay the eight regions corresponding to anterior and lateral chests for LUS guidance. To minimize the instability of the DensePose mesh coordinates based on different frontal angles of the camera, a machine learning regression algorithm is applied to predict the angle-specific projection model for the chest. ArUco markers are utilized for training the ground truth chest regions to be scanned, and another single ArUco marker is used for detecting the center-line of the body. The augmented scanning regions are highlighted one by one to guide the scanning path to execute the LUS procedure. We demonstrate the feasibility of guiding the LUS scanning procedure through the combination of augmented reality, computer vision, and machine learning.

**Keywords:** Lung ultrasound · POCUS · COVID-19 · Coronavirus · Augmented reality · Computer vision · Machine learning · Image processing

## 1 Introduction

Lung ultrasound (LUS) is a non-invasive diagnostic exam that produces ultrasound images of various components of the chest such as the lungs and pleural spaces. Artifacts in the ultrasound images arising from LUS have been used for the diagnosis of respiratory illnesses such as pneumothorax [1]. SARS-CoV-2 coronavirus has caused a worldwide pandemic of highly communicable respiratory illness known as Coronavirus Disease 2019 (COVID-19) [2]. LUS has been suggested and used to detect pneumopathy


Y. Hu et al. (Eds.): ASMUS 2020/PIPPI 2020, LNCS 12437, pp. 106–115, 2020.
https://doi.org/10.1007/978-3-030-60334-2_11

for triaging and monitoring patients who are diagnosed or suspected with COVID-19 because of its high specificity and ease of use in resource-constrained environments [3]. With COVID-19 and related respiratory illnesses claiming hundreds of thousands of lives around the world, there is an imminent need to develop time and resource-constrained diagnosis and monitoring techniques. While LUS with the point-of-care ultrasound (POCUS) offers a cost-effective, radiation-free, and higher portability solution compared with chest X-ray and CT, its accessibility is limited due to its user dependency and the small number of physicians and sonographers who can perform scanning and diagnosis [4]. There is a need to develop systems that can assist the LUS procedure for diagnosis and monitoring of lung illnesses, and primarily COVID-19. In this work, we present a solution to address this need.

We introduce a framework of guiding LUS scanning with augmented reality (AR), in which the LUS procedure can be guided by projecting the scanning trajectory. To develop such a system, we implement a computer vision-based detection algorithm to classify different regions on the human body. While 2-D/3-D cameras and depth sensors are more common for human pose estimation, we focus on monocular human pose estimation (HPE) [5] considering the resource-limited environment. A single person's chest is pictured to detect the scanning regions which define AR overlay areas to be displayed in the pictures. In the literature, several regression-based algorithms have been proposed for 2D single person pose estimation [6–9]. These algorithms map the detected coordinates of body joints or the parameters of human body models over the input image. Besides, the detection-based methods treat the body parts as detection targets based on two widely used representations: image patches and heatmaps of joint locations [10–12]. For example, PoseNet is a widely utilized algorithm for human pose estimation with monocular images [13]. While it has a high detection accuracy, it is not suitable for lung region extraction because of a smaller number of torso data points segmented. On the contrary, DensePose is another real-time approach for mapping a 3D surface-based model of the body to human pixels from 2D RGB images [14]. DensePose generates a mesh data of the human body, from which several sub-meshes including a torso sub-mesh can be extracted. We utilize DensePose to obtain the torso mesh data and to augment the "overlay", the scanning region information for the ultrasound probe tracking and the LUS guidance and navigation. Further, ArUco markers [15] are used for the detection and guidance of an ultrasound probe for regions on the chest to facilitate an effective LUS.

In this paper, we present the AR-based LUS guidance with the outline described in the following order. First, Sect. 2 describes the implementation of the utilization of DensePose along with the ultrasound probe tracking using ArUco markers and the regression-based overlay region estimation. Next, Sect. 3 describes the experimental setup. Finally, Sect. 4 presents and discusses the results for overlay generation, probe tracking, and regression-based region estimation.

## 2 Materials and Methods

### 2.1 Lung Ultrasound Scanning Protocol

LUS is performed by scanning pre-defined regions of the chest. In a common clinical workflow, the operator scans 8 regions in a supine position with 4 locations on the front and 2 on each side of the chest [16, 18]. For the LUS procedure also including the prone position, an additional 2 to 6 regions are scanned on the back, increasing the total number of regions to be 10 to 14 [17, 19–23]. Table 1 summarizes the scanning regions of different LUS protocol. While more regions provide more information for the diagnosis, LUS in the supine position is more focused especially in emergency ultrasound because the prone position is not always accessible due to the immobility of patients. To avoid the complexity of guidance and workflow, this paper focuses on 4 anterior and 4 lateral chest regions for a corresponding AR overlay. The landmark method is used to obtain LUS data from specific landmarks on the chest.

**Table 1.** Regions and locations of different LUS protocols.

| Ref. no. | # of regions | Region location | Guidance |
|---|---|---|---|
| [16] | 8 | 4 anterior, 4 lateral | Landmark |
| [17] | 12 | 12 zones | Rib |
| [18] | 8 | 4 anterior, 4 lateral | Linear and curved |
| [19] | 12 | 4 anterior, 4 lateral, 4 posterior | Exhaustive |
| [20] | 12 | Complex regions | Landmark |
| [21] | 12 | 4 anterior, 4 lateral, 4 posterior | Landmark |
| [22] | 14 | 4 anterior, 4 lateral, 6 posterior | Landmark |
| [23] | 14 | 4 anterior, 4 lateral, 6 posterior | Exhaustive |

### 2.2 DensePose-Based Upper Body Detection and Region Estimation

DensePose is a deep learning algorithm which establishes dense correspondences from 2D images to a 3D surface-based model of the human body using a single RGB image [20]. The algorithm uses the architecture of Mask-RCNN (Region-Based Convolutional Neural Network) [22] with FPN (Feature Pyramid Network) Features [23] and ROI (Region of Interest) Align pooling to obtain part labels and coordinates within each region. The supervised deep learning algorithm has the convolutional network on the top of the ROI pooling which generates per-pixel classification results for the selection of the surface part and regresses local coordinates within the part [20].

The region of interest for LUS is defined through the torso surface partition obtained after implementing the DensePose algorithm on a test image (Fig. 1(a)). The 2D surface partition thus obtained is partitioned into half to consider two hemithoraces (Fig. 1(c)). A new image is created by extracting all pixels with value 2 (relating to the torso region)

in layer I. The U and V coordinates of this sub-region are used to draw an overlay based on [1] as shown in Fig. 1(b). The augmented scanning regions are used as a reference to assist the operator to perform the LUS based on a landmark-based method, as described in Sect. 2.1. Figure 1(c) shows I, U, and V layers.

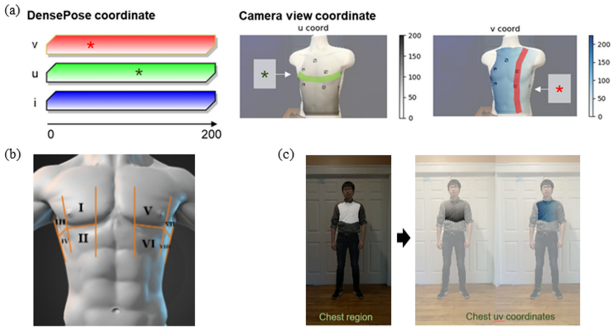



**Fig. 1.** DensePose-based torso region estimation: (a) DensePose Coordinates, (b) targeted LUS scanning regions augmented on the body based on the protocol presented in [16], & (c) post-processed IUV images for chest segmentation.

### 2.3 Probe Tracking and Scanning Region Training Using ArUco Markers

ArUco markers are binary square fiducial markers used to find correspondences to obtain the marker pose for the camera frame [21]. $6 \times 6$ marker dictionary is used which has 36 bits in the inner binary matrix. ArUco library is a part of the OpenCV python module cv2 which uses Numpy array data types. The *aruco.detectMarkers* function is used to detect the marker, its features, and identification information. Once the marker has been detected, the camera pose can be estimated given the internal camera matrix utilizing the *aruco.estimatePoseSingleMarkers* function.

To enable tracking of the ultrasound probe and to train the LUS scanning regions, 8 different ArUco markers were placed at 8 different locations on the chest based on regions shown in Fig. 1(b). Besides, an ArUco marker was also placed on top of the ultrasound probe, and the transformation matrix was calculated between the ArUco marker on the ultrasound probe and its tip. Then, based on the sequence of navigation to different regions on the chest, an arrow points towards the marker in the desired region to facilitate navigation of the markers. We obtained the 3 by 3 internal camera matrix, *cmat* for Logitech C920 webcam after calibrating with 20 different images of a checkerboard. This is used to obtain the tip position in *XYZ* coordinates as explained in Sect. 3.

### 2.4 Improving the Overlay Generation

While implementing the DensePose algorithm, there are two challenges: centering of torso mesh model irrespective of the camera angle, and inaccuracies in the mesh overlay (described in Fig. 1(b)) for varying camera angles (Fig. 2(b)). The angle was measured

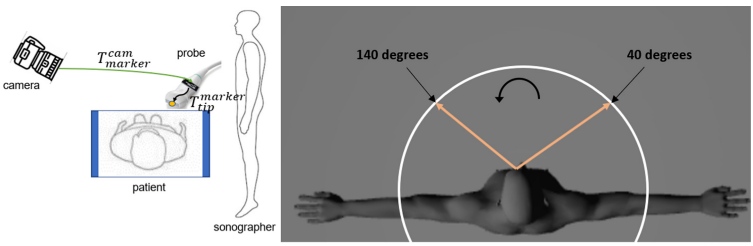
laterally assuming the anterior part of the body to be perpendicular to the testbed. In this subsection, we explain solutions to these issues.

**Centering Torso Model.** To center the torso overlay for varying camera angles, an ArUco marker was placed at the center of the torso near the neck. Center 'X' and 'Y' coordinates from the marker and the corresponding 'V' center coordinates given by DensePose are extracted. The 'V' coordinates are then used to center the torso overlay.

**Overlay Correction.** To have a corrected overlay for different camera angles, regression models are trained. The trained data are based on the 'U' and 'V' locations of the ArUco markers on the chest. The regions of the markers are based on Fig. 1(a). Models are trained for angles ranging from 40°–140° at an increment of 5° each. The Support Vector Regression (SVR) algorithm is used to define the model based on the data [24]. SVR is chosen for the data because it has been proven to be an effective tool in real-value function estimation [25]. The SVR uses a supervised learning approach and trains using a symmetrical loss function which equally penalizes high and low misestimates. Using SVR, the 'U' & 'V' coordinates which more precisely reflecting the true locations are predicted by using the camera angle as input to the machine learning model.

## 3 Experimental Implementation

The tracked ultrasound probe is used to navigate its scanning path from one region to the next through the combination with the AR-based guidance. Figure 2(a) shows the setup for probe tracking utilizing a camera. An ArUco marker is mounted on the top of the ultrasound probe for pose tracking.




(a) (b)

**Fig. 2.** (a) Ultrasound probe tracking and guidance setup. (b) Angle measurement from 40°–140° at an increment of 5° with the camera along the circle in white.

Denoting the 4 by 4 homogeneous transformation matrix from camera frame to marker frame as $T^{cam}_{marker}$ and from marker frame to transducer tip as $T^{marker}_{tip}$, $T^{marker}_{cam}$ can be obtained using the *aruco.estimatePoseSingleMarkers* function. Equation 1 denotes the 3D coordinates of the tip location $tip_{xyz} = \left[tip_x, tip_y, tip_z\right]^T$ which is obtained as:

$$\begin{bmatrix} tip_{xyz} \\ 1 \end{bmatrix} = T^{cam}_{marker}.T^{marker}_{tip}. \tag{1}$$

To determine the tip position $tip_{uv} = \left[tip_u, tip_v\right]^T$ for the regions, camera $XYZ$ coordinates are converted to UV coordinates by applying the transformation defined in Eq. 2 as

$$\begin{bmatrix} tip_{uv} \\ 1 \end{bmatrix} = cmat.T_{cam}^{world}.\begin{bmatrix} tip_{xyz} \\ 1 \end{bmatrix}. \tag{2}$$

$T_{cam}^{world}$ is the homogeneous transformation matrix from the world frame to the camera frame. Since the camera frame is defined as the world frame, $T_{cam}^{world}$ is 3 by 3 identity matrix concatenated with $[0, 0, 0]^T$ to the right.

Figure 2(b) introduces the data collection process from the subjects. While laying in the supine position, the project member takes pictures of the human chest, with angles measured from 40°–140° at an increment of 5° with the camera along the circle in white.

## 4 Results and Discussion

Figure 3 presents the DensePose verification where the camera is set at the chest level to get the body images from 20 different angles. 8 ArUco markers (ID# from 1 to 8) were attached to the center of eight regions on the chest. An additional marker (ID# 9) was attached to the lower-neck region for locating the center of the overlay. 20 sub-figures corresponding to 20 angles were processed using DensePose and the overlay was generated.

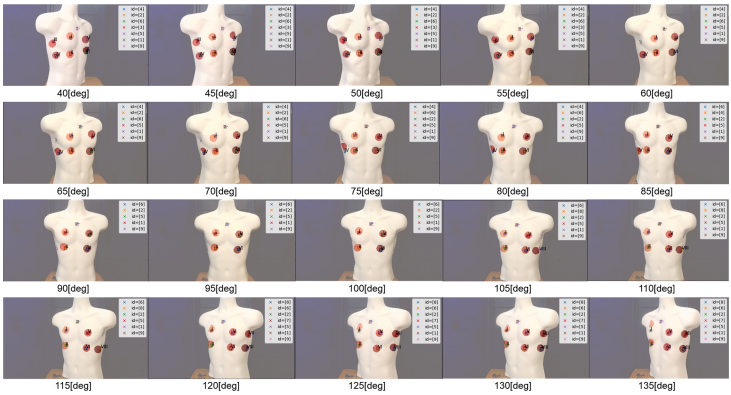



**Fig. 3.** DensePose verification for angles ranging from 40°–135° at an increment of 5°.

'U' and 'V' data corresponding to the different markers taken from camera angles from 40 to 140° at an increment of 5° were collected. To provide the best fit for the various streams of data (8 different u and v points), machine-learning based regression algorithms were utilized. Data preprocessing was done to eliminate the angles for which the marker 'U' and 'V' coordinates were zero. A total of 8 different pairs of regression models were trained, each pair corresponding to 'U' and 'V' coordinates per marker. $R^2$ score was utilized as a goodness of fit measure to quantify the performance of the machine learning

models. Table 2 shows the $R^2$ score for 'U' and 'V' coordinate predictions for different camera angles for the 8 markers. Superscript 'w' has been used for 'U' predictions for marker 1 and 6 to denote the weighted data used to generate the final score. This was done to minimize the effect of outliers obtained during support vector regression modeling.

**Table 2.** $R^2$ score for U and V value predictions for the 8 markers.

| Marker number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| U | $0.88^w$ | 0.77 | 0.85 | 0.91 | 0.76 | $0.84^w$ | 0.99 | 0.84 |
| V | 0.98 | 0.98 | 0.89 | 0.95 | 0.98 | 0.97 | 0.99 | 0.85 |

To evaluate whether the segmentation based on DensePose matches the desired segmentation, the error vector $e(i)$ in pixel space was defined as the Cartesian distance between the center of the overlay region and the center of the corresponding ArUco marker as defined in Eq. 4 as

$$e(i) = \frac{||P_m(i) - P_d(i)||}{\sqrt{w^2 + h^2}} \tag{4}$$

where $i \in [1, 8]$ is the region index, $P_m = \left[P_{mx}, P_{my}\right]^T$ is the two-dimensional marker center position in pixels and $P_d = \left[P_{dx}, P_{dy}\right]^T$ is the two-dimensional region center generated from DensePose. $w$ and $h$ are the frame width (1920 pixels) and frame height (1080 pixels) respectively. Figure 4 shows the region segmentation error for the marker position and presents a clinically acceptable range of 1–5%. Extremities are observed in markers 1 and 6 because of the outliers observed during the training of the model using the SVR algorithm.

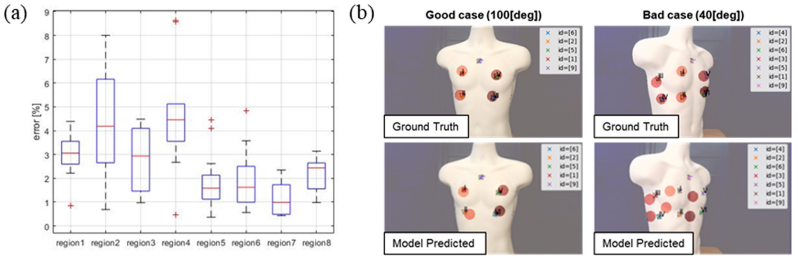



**Fig. 4.** (a) Scanning region segmentation error values from the true marker locations. Percentile errors were computed for the camera view dimension (frame size: 1080 * 1920 pixels). (b) Examples of the scanning region overlay with ground truth locations and locations predicted by the SVR model. (Color figure online)

The workflow of the real-time scanning procedure of the proposed guidance system is shown in Fig. 5. First, an overlay is generated using DensePose (area marked in red).

Due to high deploying requirements and relatively low running speed of DensePose, for 'real-time' scanning of the patient, assuming that the patient does not move during the ultrasound test, an overlay was generated once at the left and right sides, respectively. The distance between the ultrasound probe tip and the marker location is monitored and used to determine the switch of guidance from one region to another. The system guides the operator to move the probe from the center of the visible region (anterior chest) at the side to the center of the last visible region (lateral chest) using the green arrow. The developed system can be integrated with the head-mounted display or hand-held phone-based guidance for AR implementation. Future works include incorporation of LUS with the prone position as well as improving the pose estimation algorithms to achieve better region identification.

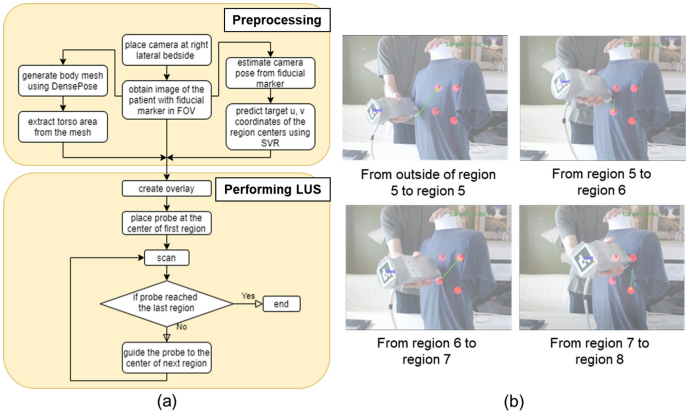



**Fig. 5.** (a) The workflow of the AR-based LUS guidance system. (b) Screenshots of real-time demonstration of the right chest as an example (right chest covers region 5 to 8).

## 5 Conclusions

In this paper, an AR-based LUS guidance framework is introduced by utilizing the scanning region identification through a combination of DensePose and ArUco markers for assisting diagnosis and subsequent monitoring of COVID-19 patients. An AR overlay can assist the sonographer to carry out the LUS for the diagnosis of respiratory illnesses such as COVID-19. Machine learning algorithms have been used to further improve the LUS region augmentation by minimizing the camera angle-associated inaccuracy. This system presents a demonstration of the capability to utilize AR, machine learning, and computer vision to help POCUS-based diagnosis and monitoring of lung-related illnesses such as COVID-19.

**Acknowledgment.** The financial support was provided through the Worcester Polytechnic Institute's internal fund; in part by the National Institute of Health (DP5 OD028162).

## References


1. Lichtenstein, D., Mezière, G., Biderman, P., Gepner, A.: The comet-tail artifact: an ultrasound sign ruling out pneumothorax. Intensiv. Care Med. **25**(4), 383–388 (1999). https://doi.org/10.1007/s001340050862
2. WHO: Coronavirus Disease 2019 (COVID-19) Situation Reports, 1 April 2020. WHO Situation Report 2019(72), 1–19. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200324-sitrep-64-covid-19.pdf?sfvrsn=703b2c40_2%0Ahttps://www.who.int/docs/default-source/coronaviruse/situation-reports/20200401-sitrep-72-covid-19.pdf?sfvrsn=3dd8971b_2
3. Soldati, G., et al.: Is there a role for lung ultrasound during the COVID-19 pandemic? J. Ultrasound Med. Off. J. Am. Inst. Ultrasound Med., 1–4 (2020) https://doi.org/10.1002/jum.15284Ads
4. Lichtenstein, D.A., Mezière, G.A.: Relevance of lung ultrasound in the diagnosis of acute respiratory failure the BLUE protocol. Chest **134**(1), 117–125 (2008). https://doi.org/10.1378/chest.07-2800
5. Chen, Y., Tian, Y., He, M.: Monocular human pose estimation: a survey of deep learning-based methods. Comput. Vis. Image Underst. **192**, 1–23 (2020). https://doi.org/10.1016/j.cviu.2019.102897
6. Toshev, A., Szegedy, C.: DeepPose: Human pose estimation via deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1653–1660 (2014). https://doi.org/10.1109/CVPR.2014.214
7. Carreira, J., Agrawal, P., Fragkiadaki, K., Malik, J.: Human pose estimation with iterative error feedback. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 4733–4742, December 2016. https://doi.org/10.1109/CVPR.2016.512
8. Sun, C., Shrivastava, A., Singh, S., Gupta, A.: Revisiting unreasonable effectiveness of data in deep learning era. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 843–852, October 2017. https://doi.org/10.1109/ICCV.2017.97
9. Luvizon, D.C., Tabia, H., Picard, D.: Human pose regression by combining indirect part detection and contextual information. Comput. Graph. (Pergamon) **85**, 15–22 (2019). https://doi.org/10.1016/j.cag.2019.09.002
10. Fourure, D., Emonet, R., Fromont, E., Muselet, D., Tremeau, A., Wolf, C.: Residual conv-deconv grid network for semantic segmentation. In: British Machine Vision Conference, BMVC 2017 (2017). https://arxiv.org/pdf/1707.07958.pdf
11. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 5686–5696, June 2019. https://doi.org/10.1109/CVPR.2019.00584
12. Tang, W., Wu, Y.: Does learning specific features for related parts help human pose estimation? In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1107–1116, June 2019. https://doi.org/10.1109/CVPR.2019.00120
13. Chen, Y., Shen, C., Wei, X.S., Liu, L., Yang, J.: Adversarial PoseNet: a structure-aware convolutional network for human pose estimation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1221–1230, October 2017. https://doi.org/10.1109/ICCV.2017.137
14. Guler, R.A., Neverova, N., Kokkinos, I.: DensePose: dense human pose estimation in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7297–7306 (2016). https://doi.org/10.1109/CVPR.2017.280

15. Romero-Ramirez, F.J., Muñoz-Salinas, R., Medina-Carnicer, R.: Speeded up detection of squared fiducial markers. Image Vis. Comput. **76**, 38–47 (2018). https://doi.org/10.1016/j.imavis.2018.05.004
16. Volpicelli, G., et al.: Bedside lung ultrasound in the assessment of alveolar-interstitial syndrome. Am. J. Emerg. Med. **24**(6), 689–696 (2006). https://doi.org/10.1016/j.ajem.2006.02.013
17. Manivel, V., Lesnewski, A., Shamim, S., Carbonatto, G., Govindan, T.: CLUE: COVID-19 lung ultrasound in emergency department. Emerg. Med. Australas., EMA (2020). https://doi.org/10.1111/1742-6723.13546
18. Moore, S., Gardiner, E.: Point of care and intensive care lung ultrasound: a reference guide for practitioners during COVID-19. Radiography (2020). https://doi.org/10.1016/j.radi.2020.04.005
19. Bouhemad, B., Mongodi, S., Via, G., Rouquette, I.: Ultrasound for "lung monitoring" of ventilated patients. Anesthesiology **122**(2), 437–447 (2015). https://doi.org/10.1097/ALN.0000000000000558
20. Lee, F.C.Y.: Lung ultrasound-a primary survey of the acutely dyspneic patient. J. Intensiv. Care **4**(1) (2016). https://doi.org/10.1186/s40560-016-0180-1
21. Via, G., et al.: Instrument to Respiratory Monitoring Tool, August 2012
22. Soldati, G., et al.: Proposal for international standardization of the use of lung ultrasound for patients with COVID-19: a simple, quantitative, reproducible method. J. Ultrasound Med. (2020). https://doi.org/10.1002/jum.15285
23. Moro, F., Buonsenso, D., et al.: How to perform lung ultrasound in pregnant women with suspected COVID-19. Ultrasound Obstet. Gynecol. Off. J. Int. Soc. Ultrasound Obstet. Gynecol. **55**(5), 593–598 (2020). https://doi.org/10.1002/uog.22028
24. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
25. Awad, M., Khanna, R.: Support vector regression. In: Efficient learning machines, pp. 67–80. Apress, Berkeley (2015)