MDPI

Article

Consciousness Detection on Injured Simulated Patients Using Manual and Automatic Classification via Visible and Infrared Imaging

Diana Queirós Pokee *, Carina Barbosa Pereira, Lucas Mösch, Andreas Follmann and Michael Czaplik

Acute Care Innovation Hub, Department of Anaesthesiology, RWTH Aachen University Hospital, 52074 Aachen, Germany; cbarbosapere@ukaachen.de (C.B.P.); lmoesch@ukaachen.de (L.M.); afollmann@ukaachen.de (A.F.); mczaplik@ukaachen.de (M.C.)
* Correspondence: dpokee@ukaachen.de

Abstract: In a disaster scene, triage is a key principle for effectively rescuing injured people according to severity level. One main parameter of the used triage algorithm is the patient's consciousness. Unmanned aerial vehicles (UAV) have been investigated toward (semi-)automatic triage. In addition to vital parameters, such as heart and respiratory rate, UAVs should detect victims' mobility and consciousness from the video data. This paper presents an algorithm combining deep learning with image processing techniques to detect human bodies for further (un)consciousness classification. The algorithm was tested in a 20-subject group in an outside environment with static (RGB and thermal) cameras where participants performed different limb movements in different body positions and angles between the cameras and the bodies' longitudinal axis. The results verified that the algorithm performed better in RGB. For the most probable case of 0 degrees, RGB data obtained the following results: Mathews correlation coefficient (MMC) of 0.943, F1-score of 0.951, and precision-recall area under curve AUC (PRC) score of 0.968. For the thermal data, the MMC was 0.913, F1-score averaged 0.923, and AUC (PRC) was 0.960. Overall, the algorithm may be promising along with others for a complete contactless triage assessment in disaster events during day and night.

Keywords: disaster medicine; triage; mass casualty incidents; remote imaging; infrared imaging; movement detection; image processing; visible imaging; mask R-CNN; UAV

1. Introduction

Disaster medicine frequently is characterized by lack of personal resources where an efficient and keen triage is essential for rescuing survivors. During triage, individuals are categorized according to injury severity and available medical resources [1]. In catastrophic situations, medical responders are overwhelmed and rationing healthcare resources is inevitable [1]. The focus is not on individuals but on populations, and these approaches' standard goal is to provide optimal services for the largest number of people [1,2].

Nowadays, triage is usually performed in three stages [3]: The primary triage is at the scene of the incident where an emergency technician carries it out, then aims the prompt assessment of the injured person, and conducts the rapid transfer to the treatment center. In most mass casualty incidents, triaging systems use tags or colored designations to categorize the victims according to their condition and their treatment need. Secondary triage is used when, due to the large extent of the incident and lack of resources in the pre-hospital, the transmission of the injured person has been prolonged in the scene and triage is performed at the hospital. The third triage is performed by a surgeon or a critical care specialist to prioritize and decide on receiving care services [3].

In the last decades, several technological progresses have been made to optimize triage in these scenarios by: including telemedical services on site with audio-video streaming through a mobile transmission unit [4], smart glasses and augmented reality [5,6], and

the development of an artificial intelligence (AI) classification for survival prediction via wearable devices and based on vital signs (while replacing medical judgment indicators) [7]. Other on-ground-based options are rescue robots, rescue workers, and rescue dogs [8]. The first requires invasive equipment and human operators with the possibility of limited access due to restrained mobility [8]. Rescue workers and rescue dogs, in turn, can overcome limited mobility, but need time and quick dislocation to distant locations [8]. As part of the project FALKE, which the German Federal Ministry of Education and Research funded, our research group investigates whether unmanned aerial vehicles (UAVs) equipped with high resolution cameras can be used for a (semi-)automatic triage. Here, a developed triage algorithm proposes a triage tag for each injured person based on their vital parameters' information obtained remotely with the cameras. Afterwards, a medical professional verifies the recommended triage tag. The speed and altitude of UAVs allied to imaging sensors may be especially adequate for an initial and fast assessment of scenarios with difficult or delayed access (due to physical barriers or Chemical, Biological, Radiological and Nuclear [CBRN] situations) and be suitable for searching for living survivors in the aftermath of a mass casualty incident [8].

In recent decades, different triage algorithms were developed and established to help rescue emergency professionals, avoid deficits in treatment and enable an early assessment at the scene [2,5,9]. Some highly integrated triage systems in the emergency department throughout the world are: Canadian Triage and Acuity Scale (cCTAS), the Australian Triage Score, the Manchester Triage Score, and Emergency Severity Index and Simple Triage and Rapid Treatment (START) [10]. In Germany, the Primary Ranking for Initial Orientation in Rescue Service (PRIOR) algorithm is often used [5]. Here, the patient's consciousness, breathing, and circulation are qualitatively addressed, which results in an individual triage and non-pathological evaluation [5]. PRIOR triage is divided into three categories as follows: severely injured with immediate treatment priority (category red or I), severely injured with appropriate transport priority (category yellow or II), and easily injured or uninjured (category green or III) [5]. The triage process also may be based on both patient's vital signs (respiratory rate, oxygen saturation in blood, heart rate, level of consciousness, and body temperature) and their main complaints [11]. Oxygen saturation along with age and level of consciousness were the three variables that best predicted mortality during hospitalization in two studies [11–13]. Another study showed that impaired consciousness was a potentially life-threatening emergency and an important factor for patients to enter the emergency department [14]. The Glasgow Coma Scale (CGS) is a scoring system to classify a patient's conscious state based on three scaled values (eye opening, verbal response, and motor response) [7]. Another triage algorithm Amberg-Schwandorf Algorithm for Primary Triage (ASAV) algorithm takes into account, among other variables, the body posture, consciousness, and if the patient is ambulating [15]. Other methods use the "ability to walk" as a criterion for the categorization, and though it is not a good discriminant, it is suited excellently for performing a rapid but coarse division of patients in tangled scenarios [9]. In fact, patients lacking walking ability must be immediately rescued from the injury zone. For subsequent triage and treatment priority allocation, this discriminate is then unsuitable [9]. In 2021, Álvarez-García and co-workers proposed the concept for an Aerial Remote Triage Algorithm [16]. This includes four steps for victims' assessment: (1) major bleeding, (2) walking, (3) consciousness, and (4) signs of life (e.g., spontaneous body movements, respiratory movements, coughing, etc.). According to the system, injured people can be classified into four categories: priority 1 (red), priority 2 (yellow), priority 3 (green), and priority * (violet).

Thus, independent of the triage method used, the level of consciousness or unconsciousness is a triage parameter and may be assessed by evaluating the patient's visual mobility capability or body activity. This work aims to visually detect patient unconsciousness and/or immobility through their mobility capacity and motion detection during daytime and nighttime using visible imaging (RGB) and infrared thermography (IRT, also known as thermal imaging) for each case, respectively. Hence, the current paper presents a

new algorithm for contactless detection of human immobility and unconsciousness based on the presumption that if a person is unconscious, they most likely will be motionless. Thus, an algorithm is proposed that combines deep learning with image processing techniques for detecting and tracking human bodies and further consciousness classification. To the best of our knowledge, there is no up-to-date algorithm in the literature for the contactless classification of patient consciousness and/or mobility on disaster survivors.

2. Materials and Methods

This section presents the algorithm developed for consciousness detection. It was optimized for two imaging techniques: RGB and IRT. The current approach is based on tracking the feature points throughout the videos. For both imaging techniques, the process is similar, differing only in the image processing of the reference frame F_0 in the first step. Here, in the reference frame, the IRT or RGB frame is preprocessed to have the most feature points dispersed in the region of interest (ROI), which is the human body. Forthwith, in a second step, the feature points will be tracked in the subsequent frames from F_1 to F_T where $0 < t \leq T$, $t \in \mathbb{Z}$, T = 30. All frames are converted in grayscale. Afterwards, the movement signal will be extracted and classified as "moving" or "nonmoving." The flowchart diagram describing the process for the RGB data is displayed in Figure 1. In Figure 2, only the first step for IRT data is described because it is the only one that differs.

The difference in imaging processes include for RGB data: the existence of the equalization process applied to the skeletonized mask applied to the gray image, and the dimension of the white grid, which is 80×80 px. Relative to the IRT data, there is no equalization applied to the image, the dimension of the white grid is of 40×40 px, the input to the mask R-CNN (region-based convolutional neural networks) model has three channels (all gray color), and there is a merge process of all the masks obtained from the gray image in different image rotation levels ($0°$, $+90°$, $-90°$). These variations occur due to different image resolutions, image type, and camera lens between the infrared and visible cameras. Both the grid and skeletonization processes are applied in a further step for a better contrast when selecting the feature points inside the human body region.
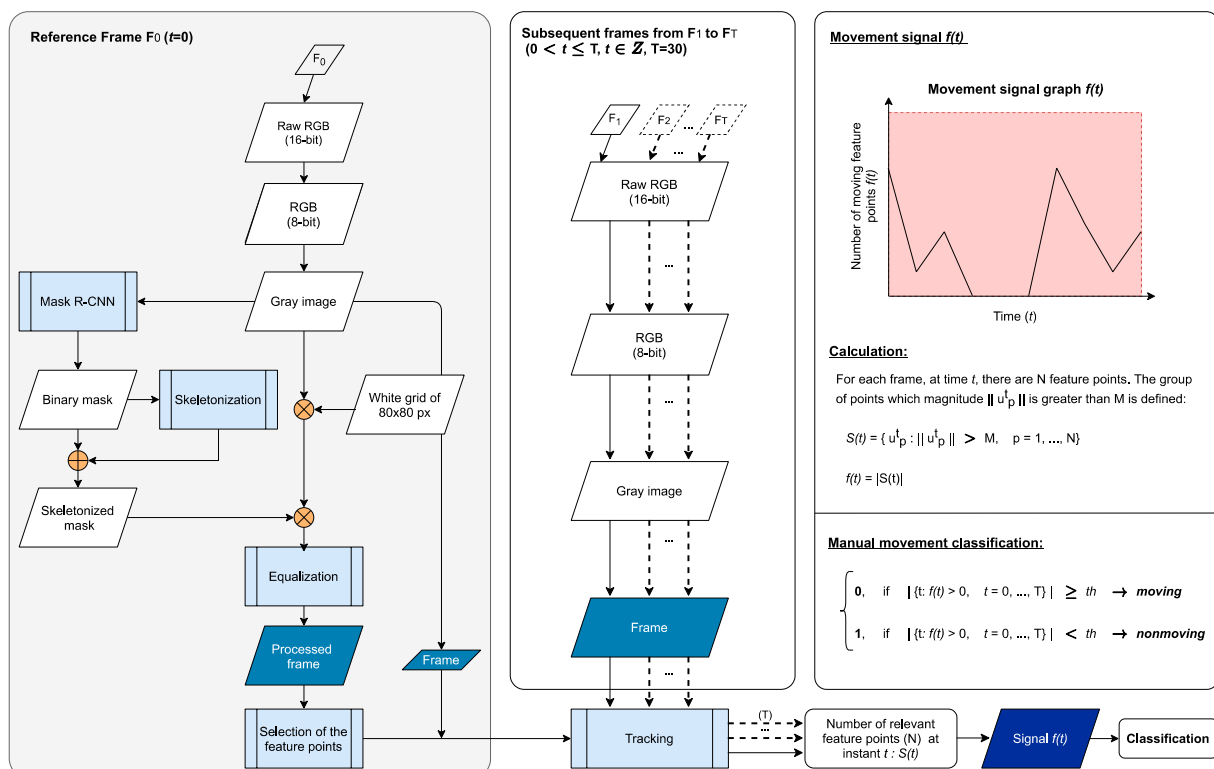

Figure 1. Flowchart of the algorithm steps for the human movement detection in RGB data.
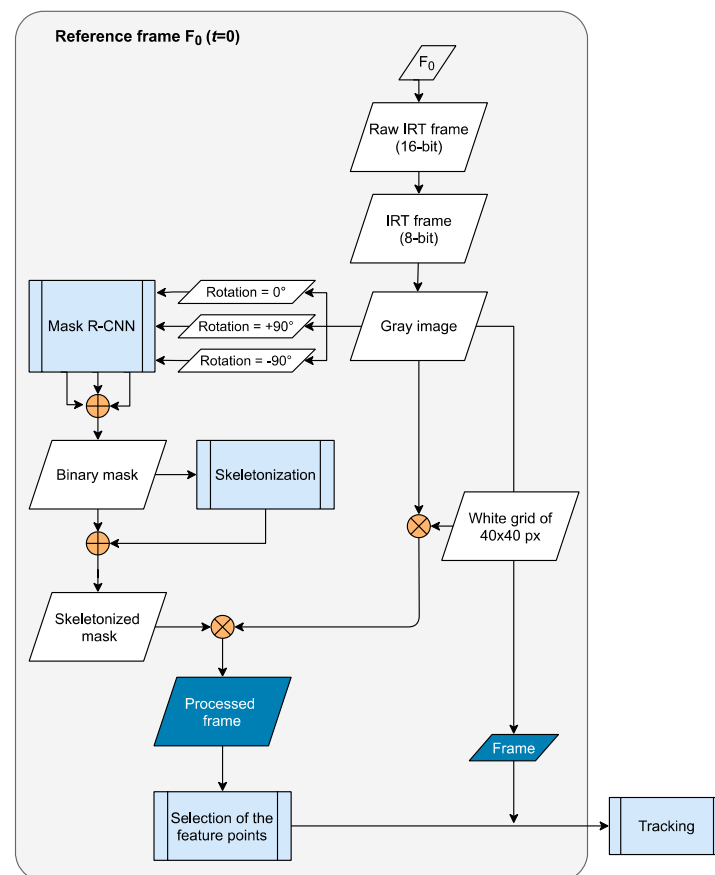
Figure 2. Flowchart of the reference frame's image processing method for the IRT data. Note there is no equalization process, the white grid applied to the gray image has a size of 40 px, and the different masks obtained from the mask R-CNN are merged (from different rotation angles of the gray image).

The present approach was implemented in Python 3.7 and tested on a 64-Bit Windows 10 computer with an octa-Core Intel® Core™ i5-8600K CPU @ 3.60 GHz, 16 GB RAM with an NVIDIA GTX 1070 (8 GB AMD GDDR5) manufactured by NVIDIA Corporate based in California, United States of America. The analysis of the data was carried out offline.

2.1. Preprocessing of the Reference Image and Subsequent Frames

The approach's first step consisted of converting the recorded frames (BAYER pattern (16-bit)) into RGB format (8-bit) and then into a gray image. Subsequently, the gray frame served as input for the mask R-CNN model. The mask R-CNN by He et al. [17] is a convolutional neural network (CNN) that detects objects in an image and generates a high-quality segmentation mask for each instance. It extends faster R-CNN by adding a branch, which predicts segmentation masks on each region of interest to the existing one for classification and bounding box regression. Faster R-CNN is composed of two stages: (1) region proposal network and (2) fast R-CNN model. The former is a neural network that proposes candidate object bounding boxes in a particular image. The latter extracts features using ROI pooling from each candidate box and performs classification and bounding-box regression. The final step of the mask R-CNN consisted in applying a fully connected network to each ROI to predict a segmentation mask. As pixel-level segmentation requires more fine-grained alignment than bounding boxes, mask R-CNN replaces ROI pooling with an ROI align layer. In the ROI align method, the bilinear interpolation was used to compute the image value on the pixel point whose coordinate is a floating-point number, thus mapping the ROI is more precise. Please note that the mask R-CNN model's pre-trained weights were based on the COCO dataset (http://cocodataset.org, accessed on 15 September 2021) [18]. This is a large dataset used for target detection and image

segmentation, including 328K images and 91 target categories. In sum, mask R-CNN served to not only recognize and classify a human being as a "person" in the images, but also compute its mask. The detection threshold was 70%, which meant that all the proposals with less than 0.7 confidence were ignored.

The next step of the algorithm consisted of skeletonizing the binary mask (output of the mask R-CNN). There are three major categories of skeletonization techniques: (1) skeleton is extracted based on distance transforms or region in an image, (2) calculating the Voronoi diagram the boundary points generate, and (3) through morphological thinning techniques based on erosion operations [19]. For this aim, image morphological operations were performed by using binary structuring elements (SE), that is, a small matrix or kernel [20]. For instance, erosion is used to shrink objects in an image and remove noise, while dilation adds pixels to the objects' boundaries. Let A be the set of foreground pixels (i.e., objects) and B the SE, with A and B sets in z^2, the erosion of A by B, denoted $A \ominus B$, is defined as:

$$A \ominus B = \{z | (B)_z \subseteq A\}, \tag{1}$$

where the z's are foreground values (ones). In other words, the erosion of A by B is the set of all points z such that B, translated by z, is contained in A [20]. As for dilating A by B, denoted as $A \oplus B$, with A and B as sets in z^2, it is defined as:

$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \varnothing\}. \tag{2}$$

Dilation is based on reflecting B in its origin and translating the reflection by z as in erosion. In other words, dilating A by B is the set of all displacements z, such that the foreground elements of $\hat{B}$ overlap at least one element A. These morphological operations take part in the skeletonization process according to Lee et al. [21]. First, an erosion followed by dilation known as opening, is performed to the original image. Then, the opening resulting image is subtracted from the original image, resulting in a temporary image. Forthwith, the original image is eroded, and the skeleton is refined by computing the union of the current skeleton and the temporary image. This step creates the skeletonized mask.

To the gray image, a white grid is applied before the multiplication process. Afterwards, to adjust image intensities and enhance contrast, the resulted image f from the multiplication of the skeletonized mask was equalized with the grided frame. Please note that no equalization was performed for the IRT frames. Let image f be represented as m_r by m_c matrix of integer pixel intensities (n) ranging from 0 to $L - 1$ [20]. Let p denote the normalized histogram of f with a bin for each possible intensity so that:

$$p_n = \frac{number\ of\ pixels\ with\ intensity\ n}{m_r \times m_c}, \ n = 0,\ 1,\ \ldots,\ L - 1 \tag{3}$$

The histogram of the equalized image g is defined by:

$$g_{i,f} = floor\ ((L - 1) \sum_{n=0}^{f_{i,j}} p_n), \tag{4}$$

where *floor* () rounds down to the nearest integer. This is equivalent to transforming the pixel intensities, k, of f by the function:

$$T(k) = floor\ ((L - 1) \sum_{n=0}^{K} p_n). \tag{5}$$

Thinking of the intensities of f and g as continuous random variables X and Y on $[0, L - 1]$, Y is defined by:

$$Y = T(X) = (L - 1) \int_0^X p_X(x)dx, \tag{6}$$

where p_X is the probability density function (PDF) of a random variable X. The function T is the cumulative distributive function (CDF) of X multiplied by $(L - 1)$ [20]. The function T is always positive, and it corresponds to the uniform histogram of the resulting equalized image [20]. Y corresponds to the image resultant of the equalization process.

2.2. Selection and Tracking of the Feature Points

In the next step, the feature points were selected and then tracked using the Shi-Tomasi corner detector [22]. For a window (*W*) located at (*x,y*) with a pixel intensity of *I* (*x,y*), the scoring function calculates the score for each pixel to be considered a corner by:

$$f(x,y) = \sum(I(x_k + y_k) - I(x_k + \Delta x, y_k + \Delta y)^2 \ where \ (x_k + y_k) \ \epsilon \ W. \tag{7}$$

When scanning an image with a window and an area is above a minimum value, a corner is selected at that point through:

$$R = \min(\lambda_1, \lambda_2). \tag{8}$$

Thus, *R* is the score, where λ_1 and λ_2 are eigenvalues of the scoring function. If *R* is greater than a certain predefined value, it can be marked as a corner. This works as follows: if both eigenvalues are greater than a certain value, the pixel is considered a corner; if either λ_1 or λ_2 is less than the required minimum, the pixel is considered an edge; or if both eigenvalues are less than the required minimum, the pixel is considered "flat" [23].

The algorithm finds *N* (set to a maximum of 90) strongest corners in the gray image with a distance of at least 10 pixels. After selecting the corners, the aim is to compute their trajectories in consecutive frames. Using the Lucas–Kanade optical flow method [24], the selected feature points were iteratively tracked throughout the video frames. The algorithm identifies the feature points' location in a specifically sized window around the feature points in the next frame. Displacing the feature points creates feature vectors. Considering the intensity of a pixel *I(x,y,t)* at instant *t* in time, that moves by distance *(dx,dy)* in next frame at *dt* time, and considering that the pixels do not change intensity, it can be stated that:

$$I(x,y,t) = I(x + dx, \ y + dy). \tag{9}$$

Using the Taylors series approximation, removing the common terms and dividing by *dt*, results in the following equation:

$$f_x u + f_y v + f_t = 0. \tag{10}$$

Here, f_x and f_y are image gradients and f_t is the gradient along time. To obtain the vector *(u,v)* values, the least square fit method is used, resulting in the equation:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_i f_{x_i}^2 & \sum_i f_{x_i} f_{y_i} \\ \sum_i f_{x_i} f_{y_i} & \sum_i f_{y_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i f_{x_i} f_{t_i} \\ \sum_i f_{y_i} f_{t_i} \end{bmatrix}, \tag{11}$$

where *u* and *v* are vector values in *x* and *y*, respectively. Let vector *u* be derived through the gradient in *x* and *y* coordinates, the image velocity at *x* is also known as optical flow at *x* [25].

2.3. Extracting the Movement Signal and Classification

The movement signal was obtained through all the read frames along the recorded videos of period *T* = 30 s. First, for each frame, the algorithm computed the feature vectors' vertical and horizontal gradients. The gradient's magnitude was also calculated for each feature vector. Let the two-dimensional vector be $u = (u_1, u_2)$ and its magnitude be $||u||$. The tracking points (*N*) which magnitude at instant *t* $\left\lVert u_p^t \right\rVert$ was higher than a certain magnitude *M* (distance/dt) in between the frames and was selected resulting in *S(t)*. The latter is the group of feature points in which magnitude is higher than *M*. The value of *M* was set to 100 for visible data and 450 for infrared data. The *S(t)* function at instant *t* is defined by:

$$S(t) = \left\{ u_p^t : \left\lVert u_p^t \right\rVert > M, \ p = 1, \dots, \ N \right\}, \ (0 < t \le T, \ t \in \mathbb{Z}). \tag{12}$$

Then, the signal *f* along the period of 30 s was the counting of each instant's relevant feature points, or the number of the selected feature points in the group *S(t)* for each instant:

$$f(t) = |S(t)| \tag{13}$$

Finally, the number of peaks of the signal during the video duration of 30 s ($N = 30$) was counted and was only considered if it was lower than the threshold *th* (set to 3). Therefore, if the number of peaks was lower than *th*, the algorithm interpreted the signal as a nonmovement and stated that the person was "nonmoving". Otherwise, it considered as "moving." The "nonmoving" or potentially unconscious person was labeled as 1 and "nonmoving" or conscious person was labeled as 0:

$$Classification = \begin{cases} 0, & if\ |\{t : f(t) > 0,\ t = 0, \ldots, T\ \}| \ \geq\ 3\ \rightarrow moving \\ 1, & if\ |\{t : f(t) > 0,\ t = 0, \ldots, T\ \}|\ <\ 3\ \rightarrow nonmoving \end{cases} \quad (14)$$

where $(0 < t < T,\ t \in \mathbb{Z})$. At the end, the algorithm returned a value as "moving" or "nonmoving," which was considered either as conscious or potentially unconscious, respectively. For a clearer and more practical follow-up and understanding, the labels "moving" and "movement" were used for mobility/consciousness detection (negative prediction) and "nonmoving" and "no movement" for immobility/unconsciousness detection (positive prediction). This labeling is due to how unconsciousness detection is our aim, and thus, a positive prediction.

All steps to the final classification of the human movement as negative or positive in the RGB are displayed in Figure 3. First, from the mask obtained from the mask R-CNN, its skeleton and a white grid were added to the same mask, which was then applied to the gray image. This first step was preprocessing the videos' first frame. Afterwards, the feature points from the frame 0 were selected and then tracked in the gray subsequent frames. Finally, processing the frame data resulted in a movement signal along the 30 s recording time and later in classifying a moving/nonmoving person.
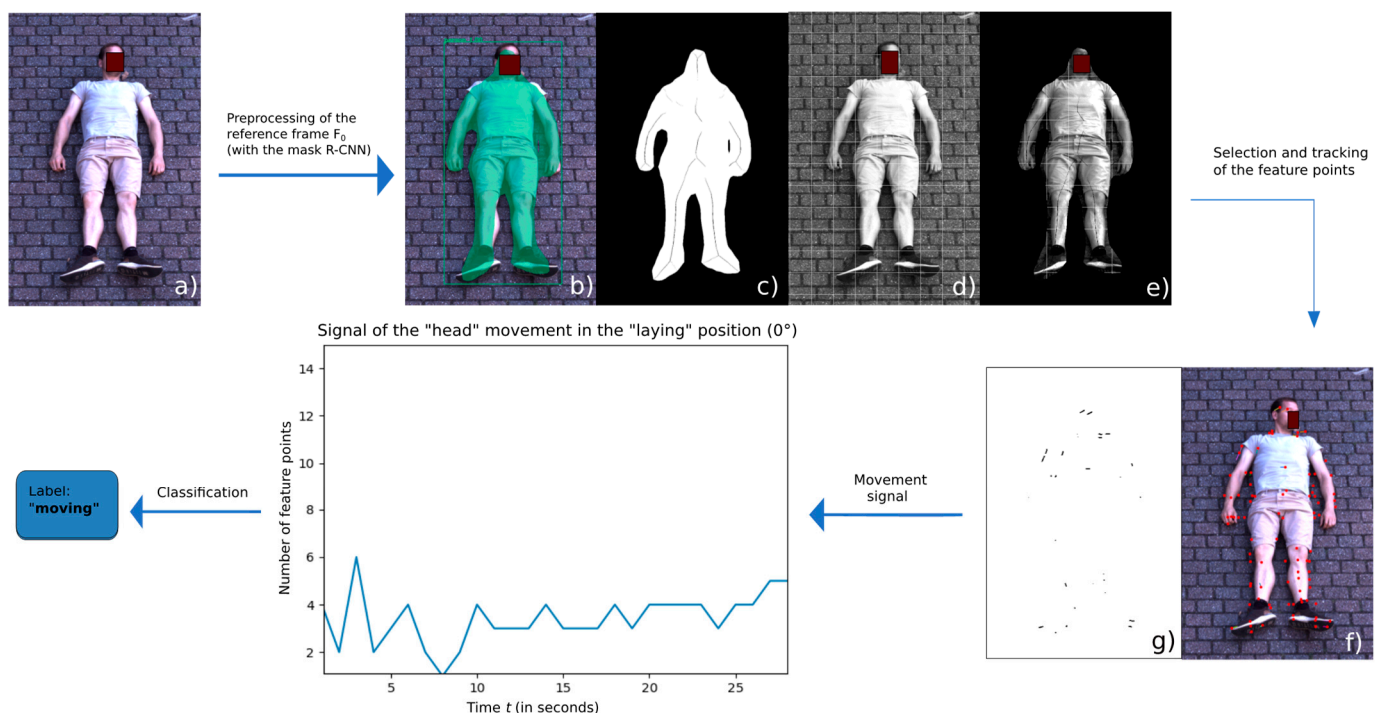
Figure 3. Intermediary RGB images throughout the algorithmic process: (**a**) the original RGB frame, (**b**) the mask and bounding box given from mask R-CNN, (**c**) the binary skeletonized mask, (**d**) the white grid on the gray image, (**e**) the processed image prepared for the selection of feature points, (**f**) the selection of feature points, and (**g**) the movement lines in between two frames of the feature points (essentially "head" movement). The obtained movement signal through time is displayed in the graphic, and, in this case, the person is classified as "moving".

2.4. Statistical Methods

For the human movement detection, a binary classification using the labels movement/consciousness and no movement/unconsciousness were used. For each of the different movements the patients performed, the positive label and the negative label corresponded to "no movement" and "movement," respectively. Thus, the ground truth of the data was no movement when there was no movement, and movement for each of the limbs' movement performed (head, one arm, both arms, one leg, both legs, and slight movement). As one may note, the dataset was imbalanced because the ratio of positives to negatives was 1:6. For a clearer understanding and more appropriate analysis of the algorithm's performance, various statistical parameters were investigated and the following were chosen: F1-score, AUC (PRC), MCC, Cohen's kappa or kappa, and balance error rate (BER).

Let "TN," "TP," "FN," and "FP" represent "true negative," "true positive," "false negative," and "false positive," respectively. The total number of samples in the classification is N (=TN + TP + FN + FP). In such an imbalanced dataset, where the number of negative samples is large, the false positive rate increases slower. Typical calculated metrics include precision, recall, and accuracy. However, these alone are not a good measure when evaluating an imbalanced dataset, thus leading to an imbalanced precision and recall. Therefore, for a comprehensive evaluation of the classification model metrics, the harmonic mean of precision and recall (F_{β}-score) was calculated by [26]:

$$CF_{\beta} = \left(1 + \beta^2\right) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot \ precision) + recall}. \tag{15}$$

The beta parameter allowed us to control the tradeoff of importance between precision and recall. $\beta < 1$ focuses more on precision, while $\beta > 1$ focuses more on recall metrics [26]. Regarding the F1-score, β equals 1, and both recall and precision have the same weight on evaluating the process.

Another statistical metric evaluated was Cohen's kappa *(k)*, which is another popular consensus metric of interrater reliability that estimates the degree of consensus between two judges and determinates whether the agreement level exceeds what would be expected to be observed by chance alone [27,28]. A kappa value of 0 does not indicate that the two judges did not agree at all but rather that the two judges did not agree with each other any more than would have been predicted by chance alone. Again, the kappa coefficient is a highly useful statistic when most observations fall into a single category. Supposing the correct recognition rates (CR_1 and CR_2) and error rates (E_1 and E_2) from a confusion matrix are defined by

$$CR_1 = \frac{TN}{N}, \quad CR_2 = \frac{TP}{N}, \\ E_1 = \frac{FP}{N}, \quad E_2 = \frac{FN}{N}, \tag{16}$$

and supposing the relationship of the population rates p_1 and p_2 is

$$p_1 = CR_1 + E_1, \text{ and } p_2 = CR_2 + E_2, \tag{17}$$

Cohen's kappa is described as follows and within the range $[-1, 1]$:

$$k = (Pa - Pc)/(1 - Pc), \text{ where} \tag{18}$$

$$P_a = \frac{TN + TP}{N} \text{ and} \tag{19}$$

$$P_c = p_1 \frac{TN + FN}{N} + p_2 \frac{TP + FP}{N}. \tag{20}$$

Kappa or k is a metric used to measure the strength of inter-rater agreement, for which the maximum agreement is 1 and 0 represents a random agreement [29,30].

The Matthews correlation coefficient (MCC) is a measure that is not influenced by imbalance test sets since it considers accuracies and error rates mutually on both classes and involves all values of the confusion matrix. The MCC ranges from 1 for a perfect prediction to −1 for the worst possible prediction. A value of 0 indicates a model that

performs randomly [31]. Some authors consider this metric as the best singular assessment metric and especially suitable to the case of imbalanced data [32]. It is defined by:

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{p_1 p_2 N^2 (TN + FN)(TP + FP)}}. \tag{21}$$

The balance error rate (BER) is measured rather than the often used error rate since previous studies have shown that the error rate does not change in datasets with unequal numbered classes [33]. The BER is considered a "proper" measure for processing class-imbalanced problems [31]. It is defined as it follows:

$$\text{BER} = \frac{1}{2}\left(\frac{E1}{p_1} + \frac{E2}{p_2}\right). \tag{22}$$

In a study carried by [34] showed that accuracy, F1-score, kappa, and ROC curve were are skewed distributions in imbalanced data [34]. Cohen's kappa was affected by skew in either direction, while F1-score was affected in only one direction.

Finally, commonly used measures of classifier performance for visual graphical data are accuracy, error rate, and area under the receiver operating characteristics (ROC) curve (AUC). However, recent studies [33,35,36] have shown that ROC can be misleading when this last applied to imbalance scenarios, due to its overly optimistic performance. In fact, ROC plots can be deceptive with respect to conclusions about the reliability of classification performance, due an intuitive but wrong interpretation of specificity in imbalanced datasets [32,33,36]. Studies have shown that only precision, PRC, MMC, and F-beta scores vary between datasets, while most other measures stay unchanged (ROC, concentrated ROC [CROC], cost curves [CCs], error rate, accuracy, FPR, and accuracy) [32,33,36]. As compared to the traditional ROC, PRC changes not only with the ratio of positive and negatives in the dataset but also with the AUC score of the PRC as well. The latter is denoted as AUC (PRC) and will be evaluated in this study [33,34,37].

2.5. Experimental Protocol

An experiment was carried out in an outdoor uncontrolled environment with static infrared and visible cameras: a FLIR BOSON 640 (640 × 512 px image resolution with spectral range 7.5–14 μm) and Allied Vision Mako G-234C (1080 × 1920 px image resolution) with a frame rate of 30 and 15 fps, respectively. These were sat atop a tripod and positioned so that the angle between them and the floor was 45°, and the diagonal distance to the patients was approximately 12 m, as depicted in Figure 4. Here, it was important to have a full human body in the field of view of the cameras. The setup was dimensioned so that it could be as similar as possible to the setup using a UAV with cameras hovering over the wounded people.
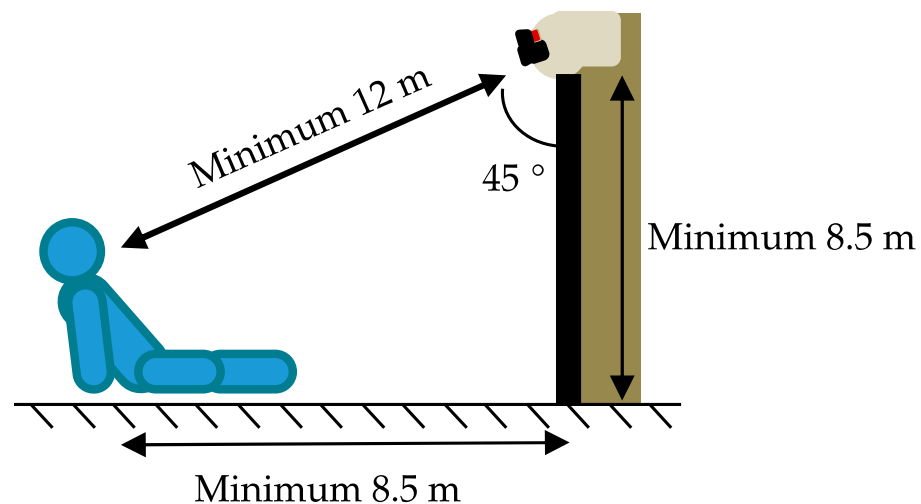

Figure 4. Sketch of the experimental setup on the floor (using static cameras).

The tripod was placed on the second floor of the Acute Care Innovation Hub Institute to replicate the camera's angle and height toward the floor when using an UAV. In this cast, 20 healthy volunteers (six women and 14 men) between the ages of 22 and 36 (27.40 years ± 3.84) participated in the current experiment. Each participant's temperature was checked (using an infrared FLIR E95 camera), and basic individual and weather parameters were collected. The subject group had an average height of 180.40 cm ± 9.55, average weight of 78.3 kg ± 15.02, and average body temperature of 36.69 °C ± 1.28. The average weather conditions were 26.50 °C ± 4.34 temperature, 31.15% ± 4.21 relative humidity, and 22 km/h ± 6.94 wind velocity.

The same protocol was performed using three body angles toward the camera (clockwise rotation on the longitudinal axis of the human body at 0°, 45°, and 90°, as depicted in Figure 5a). For each angle, the subjects posed in two positions: "sitting" and "laying" (as displayed in Figure 5b). For each position, a set of seven movements was performed ("no movement," "head," "one arm," "both arms," "one leg," "both legs," and "slight" movement) for 30 s each.

Figure 5. (**a**) Laying positions seen from different angles between the longitudinal axis of the human body and the cameras (0°, 45°, and 90° clockwise); (**b**) sitting and laying positions in a 0° body sagittal plan rotation toward the camera.

3. Results

Most of the videos were recorded according to the study protocol. However, due to technical issues in the image calibration of the thermal cameras, only IRT videos from 12 probands had sufficient quality to be evaluated.

All of the human subjects performed seven movements in two positions (sitting and laying) for each angle (0°, 45° and 90°) for a 30 s duration. Thus, 42 video recordings were made per subject. Overall, for the group of 20 individuals, there were 280 video samples for each angle, of which 40 videos labeled as "no movement" and the remaining 240 categorized as "movement." In sum, 840 total RGB videos were collected. The statistical metrics from the RGB data are presented in Table 1.

The ratios of correctly predicted classes for each position, angle, and movement are displayed in Table 2. The ratio values of correctly predicted classes is 1 when there is the greatest and perfect correspondence between the actual and predicted value, and 0 for correspondence at all. Notably, the algorithm performed worse for the "slight" movement, especially in the "laying" position, where it predicted "nonmoving" instead of "moving." As for the "no movement" pose, it failed for the "sitting" positions at 0° and 45°. Additionally, the algorithm failed once for the "one arm" movement detection.

All the positive cases were correctly predicted at both 0° and 45°, while the number of FPs was similar for all the angles. Recall values was 1.0 and higher at 0° and 45°, while precision was higher for 0° than it was for the other two angles. Accuracy values ranged between 0.958 and 0.976, with 0° having the greatest value. The F1-score ranged from 0.889 to 0.923 and was the highest at 0°. AUC (ROC) scores were similar, ranging from 0.936 to 0.960, and was the highest at 0° as well. Finally, kappa and MMC scores had higher

variance, between 0.938 and 0.909 as well as 0.840 and 0.913, respectively. The BER was the lowest for 0°, followed by 45° and then 90°; it ranged between 0.014 and 0.059.

Table 1. Statistical metrics for the RGB data from the whole study group for three angles between the longitudinal axis of the body and the cameras.

Metrics	Angles		
	0°	45°	90°
Confusion matrix $\begin{pmatrix} TN & FP \\ FN & TP \end{pmatrix}$	$\begin{pmatrix} 237 & 3 \\ 1 & 39 \end{pmatrix}$	$\begin{pmatrix} 236 & 4 \\ 3 & 37 \end{pmatrix}$	$\begin{pmatrix} 239 & 1 \\ 0 & 40 \end{pmatrix}$
F1-score	0.951	0.914	0.988
AUC (PRC)	0.968	0.965	0.994
kappa	0.943	0.899	0.986
MMC	0.943	0.899	0.986
BER	0.019	0.046	0.002

TN: true negatives, *FP*: false positives, *FN*: false negatives, *TP*: true positives.

Table 2. Ratio of correctly predicted classes for each position, angle, and movement for the RGB data from the 20-subject group (each 0.05 in the value accounts for a corrected predicted video).

	No Movement	Head	One Arm	Both Arms	One Leg	Both Legs	Slight
Sitting 0°	0.95	1.00	1.00	1.00	1.00	1.00	1.00
Laying 0°	1.00	1.00	1.00	1.00	1.00	1.00	0.85
Sitting 45°	0.85	1.00	1.00	1.00	1.00	1.00	1.00
Laying 45°	1.00	1.00	0.95	1.00	1.00	1.00	0.85
Sitting 90°	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Laying 90°	1.00	1.00	1.00	1.00	1.00	1.00	0.95

Overall, the algorithm had better statistical metrics at 0° and 90°. The number of FNs was also lower for these two angles, while the number of FPs was greater for 0° and 45°. The F1-score was higher for 0° and 90°, ranging from 0.914 to 0.988, and the AUC (PRC) varied from 0.965 to 0.994. The kappa and MMC scores presented equal values at each angle, varying from 0.899 to 0.986. The BER values were between 0.002 and 0.046. At 0° and 90°, the F1-score, kappa, and MMC scores were higher and the BER was lower.

Regarding the IRT results of the data available from 12 individuals, 168 videos samples were collected from each angle, with 24 corresponding to the "no movement" class and the remaining 144 to the "movement" class. Table 3 shows the statistical metrics for the infrared data.

Table 4 displays the ratios of correctly predicted classes for each position, angle, and movement. The incorrectly predicted classes mainly involved the "slight" movement, with a small tendency toward the "sitting" position. For the "no movement" pose, the algorithm failed only for 0° at the "sitting" position. In one video, the classifier categorized the "one arm" movement as "nonmoving" instead of "moving".

For comparison purposes, the statistical metrics for the RGB data were calculated using only the subjects for whom infrared videos were available. These results are displayed in Table 5. Overall, similar results were obtained to those presented in Table 5. The F1-scores ranged from 0.980 to 0.917 and were the highest at 90°. AUC (PRC) had similar scores for all angles, from 0.982 to 0.988. Kappa and MMC ranged from 0.903–0.976 and 0.903–0.976, respectively. Lastly, BER was lowest at 90°, followed by 0° and 45°, ranging between 0.003 and 0.049.

Table 3. Statistical results for the IRT data of the group of 12 participants for three angles between the longitudinal axis of the body and the cameras.

Metrics	Angles		
	0°	45°	90°
Confusion matrix $\begin{pmatrix} TN & FP \\ FN & TP \end{pmatrix}$	$\begin{pmatrix} 140 & 4 \\ 0 & 24 \end{pmatrix}$	$\begin{pmatrix} 138 & 6 \\ 0 & 24 \end{pmatrix}$	$\begin{pmatrix} 139 & 5 \\ 2 & 22 \end{pmatrix}$
F1-score	0.923	0.889	0.863
AUC (PRC)	0.960	0.945	0.936
kappa	0.909	0.868	0.838
MMC	0.913	0.876	0.840
BER	0.014	0.021	0.059

TN: true negatives, *FP*: false positives, *FN*: false negatives, *TP*: true positives.

Table 4. Ratios of correctly predicted classes for each position, angle, and movement for the IRT data from the 12-subject group (approx. each 0.08 accounts for a correctly predicted video).

	No Movement	Head	One Arm	Both Arms	One Leg	Both Legs	Slight
Sitting 0°	1.00	1.00	1.00	1.00	1.00	1.00	0.75
Laying 0°	1.00	1.00	1.00	1.00	1.00	1.00	0.92
Sitting 45°	1.00	1.00	1.00	1.00	1.00	1.00	0.67
Laying 45°	1.00	1.00	1.00	1.00	1.00	1.00	0.83
Sitting 90°	0.92	1.00	1.00	1.00	1.00	1.00	0.83
Laying 90°	0.92	1.00	0.92	1.00	1.00	1.00	0.83

Table 5. Statistical results from the RGB data of the 12-participant group for three angles between the longitudinal axis of the body and the cameras.

Metrics	Angles		
	0°	45°	90°
Confusion matrix $\begin{pmatrix} TN & FP \\ FN & TP \end{pmatrix}$	$\begin{pmatrix} 142 & 2 \\ 0 & 24 \end{pmatrix}$	$\begin{pmatrix} 142 & 2 \\ 2 & 22 \end{pmatrix}$	$\begin{pmatrix} 143 & 1 \\ 0 & 24 \end{pmatrix}$
F1-score	0.96	0.917	0.98
AUC (PRC)	0.985	0.982	0.988
kappa	0.953	0.903	0.976
MMC	0.954	0.903	0.976
BER	0.007	0.049	0.003

TN: true negatives, *FP*: false positives, *FN*: false negatives, *TP*: true positives.

Table 6 displays the ratios of correctly predicted classes for each position, angle, and movement. Here, the wrongly predicted labels involved the "slight" movement in the "laying" position, for which the algorithm predicted "nonmoving" instead of "moving." Additionally, two "no movement" videos were wrongly predicted as "moving" videos at 45°. There were no incorrectly predicted videos for the "moving" videos except for "slight" movement.

Table 6. Ratio of correctly predicted classes for each position, angle, and movement for the RGB data from the 12-subject group (each approx. 0.08 accounts for a corrected predicted video).

	No Movement	Head	One Arm	Both Arms	One Leg	Both Legs	Slight
Sitting 0°	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Laying 0°	1.00	1.00	1.00	1.00	1.00	1.00	0.83
Sitting 45°	0.83	1.00	1.00	1.00	1.00	1.00	1.00
Laying 45°	1.00	1.00	1.00	1.00	1.00	1.00	0.83
Sitting 90°	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Laying 90°	1.00	1.00	1.00	1.00	1.00	1.00	0.92

4. Discussion

In general, the presented algorithm obtained good results with both the IRT and RGB data for detecting immobility on people using upper static cameras.

In the RGB 20-subject data results, the mask R-CNN detected and segmented all the subjects with very good accuracy. However, in some videos, the mask R-CNN could not segment the bodily extremities (hands and feet) with great precision. From the 12 wrongly predictions, 4 were FNs and 8 were FPs. The four videos wrongly predicted as "moving" instead of "nonmoving" (FN) were due to some clothing movement caused by the wind or to sudden sunlight changes. The very intense bright light led to a greater number of homogeneous pixels, consequently hindering the tracking of the feature points. The FNs were mainly in the "sitting" position; and, the algorithm performed worst at 45°. In turn, the FPs were mainly in the "laying" position, and no major differences were observed between all three angles. Additionally, the majority of the FPs were among the "slight" movement videos, where the subjects performed "slight movements," such as moving their hand, fingers, or feet. These videos were expected to have fewer moving feature points and with less magnitude, making these movements more prone to being predicted incorrectly. The main reasons for the FPs and FNs were non-detection of fully body extremities by the mask R-CNN, insufficient magnitude of the feature points for them to be considered relevant moving ones, and an insufficient number of feature points. There were more FPs than FNs. In fact, having more FPs is preferable. It is more important to minimize FNs or the cases where unconscious people are classified as conscious, for a greater impact on their survival in a disaster scenario.

Overall, the developed algorithm predicted "movement" and "non-movement" with high kappa, MMC, and F1-score values and low BER. The kappa and MMC values were lower than the F1-scores were. The statistical metrics show that the MMC values are symmetrical (they consider accuracies and error rates mutually on both classes) and involve all values of the confusion matrix. The MMC values correspond to the correlation between both true and predicted values and translates the probability of having a perfect prediction, ranging from [−1 to 1]. Data evaluation showed equal MMC and Kappa values of 0.943, 0.899 and 0.986 for 0°, 45°, and 90°, respectively. According to the interpretation of the metric's creators Landis and Koch [29], Kappa values are considered almost perfect as values belong to interval [0.81–1.00]. AUC (PRC) values were also good for all three angles, ranging from 0.965 to 0.994. Thus, the algorithm appears to be promising for detecting immobility in people and potential unconsciousness.

Regarding the IRT data (including only videos from 12-subjects), the mask R-CNN computed and segmented human bodies with less precision. By comparison, for the RGB data, body extremities were less segmented, and full limbs sometimes were not segmented. This might be due to the training of the neural model with RGB images instead of IRT images. Overall, the algorithm developed for thermal images had good performance but was less accurate compared to the results obtained for the RGB data. Among the 504 videos, there were 17 wrongly predicted values, 2 FNs (both at 90°) and 15 FPs (mostly for "slight" movement at the three angles). The FPs were in the "sitting" and "laying"

positions, while the FPs tended to be in the "sitting" position. The algorithm performed the worst at 90° and the best at 0°. The results from 45° were better relatively to those of 90° by presenting no FNs. The algorithm mainly failed for the following reasons for the IRT data and for FPs. The mask failed to detect the bodily extremities as part of the human body, so the number of moving feature points was insufficient for the signal to be considered motion for the recorded time period. The remaining FNs occurred when the movement was so slight that the magnitude of the feature points was insufficient for the points to be considered moving feature points. Both incorrect predictions might be due to the grayscale from the infrared information (in comparison to the RGB scale from visible frames) and the feature points consequently being harder to track, lower image resolution/contrast, and the fact that the mask R-CNN was pre-trained with RGB data and consequently less suitable for infrared human detection. Regarding the metrics, videos recorded in 0° had the best kappa (0.909), MMC (0.913), F1-score (0.923), and BER (0.014) results. For the other angles, the kappa scores were between 0.838 and 0.868, MMC was between 0.840 and 0.876, F1-scores were between 0.863 and 0.889, and BER was between 0.021 and 0.059. Most of the incorrectly predicted values were FPs, which again has less impact in real scenarios. Although the algorithm performed worse with IRT videos during the daylight experiment, the advantage of using infrared thermography, rather than visible imaging, is the possibility of utilizing it at nighttime or scenarios with limited visibility or natural light, such as dusty and foggy conditions.

An evaluation using the RGB data from 12 subjects was performed to compare both imaging technologies. Globally, the algorithm performed better on the RGB data, with MMC scores of 0.943, 0.899, and 0.986 at 0°, 45°, and 90°, respectively. However, the MMC score obtained from the IRT videos was sufficiently good, with 0.913, 0.876, and 0.840 at 0, 45°, and 90°, respectively. The number of FPs and FNs was greater than 10 for the IRT data group. This can be due to the lower contrast among the IRT images, which can lead to less accurate segmentation, less prominent edges in the images, and a lower number of selected feature points, which compromise the tracking performance and classification process. In general, the algorithm performed worse at detecting and recognizing "slight" movement as movement. Most of the FNs were in the "sitting" position, while most of the FPs were in the "laying" position.

For further algorithm development, the mask R-CNN may be refined and be trained with IRT data, or another model pre-trained on infrared images can be applied. The magnitude threshold for feature points to be considered moving ones could be investigated. Additionally, different environments and ground temperatures may be studied, which may influence mask segmentation and classification in IRT images using mask R-CNN. In addition, to the best of our knowledge, there is not up to now any open dataset which meets our requirements relatively to acquisition condition (distance and angle of the cameras toward participants). Thus, to guarantee a more reliable analysis of the algorithm's performance, more subjects should be enrolled in future studies. The algorithm was tested using higher-level static cameras to validate the algorithm. Testing is planned using cameras placed in a hovering drone over simulated injured people in a set-up catastrophe scenario. In this case, some future main obstacles include the drone's movement and camera calibration toward the image background as well as how the wind provoked by the drone influences the movement of the probands' clothes. Other interesting experimental setups would include testing on partially occluded human bodies, testing in different laying positions (supine, prone, or side-lying), or identifying moving limbs. This could lead to more precise information being delivered to rescuers.

Lastly, to the best of our knowledge, there is no other study or method in literature to assess consciousness detection based on human motion and remote imaging in both infrared and visible imaging, and also specifically when using an UAV. Therefore, this work can be used as preliminary study and a staircase for future algorithm development to be applied in remote and efficient triage algorithms in catastrophic medicine, as well

as in other medical fields. In fact, it can provide information about one common triage parameters, i.e., consciousness or/and mobility when categorizing injured people.

5. Conclusions

In this paper, an algorithm was proposed that proved capable of detecting human body immobility through visible and infrared technologies in an outside environment using static cameras. This could be useful for detecting unconsciousness in a disaster scenario. The system is a combination of image-processing techniques along with deep learning. It assesses, for two image modalities, one of the four main parameters (major bleeding, walking, consciousness, and signs of life) essential for triage, according to Álvarez-García et al. [16]. It could be used during daytime and nighttime with RGB and IRT cameras.

In general, the classifier performed better for RGB imaging but still performed well for thermal imaging. It had high MMC and kappa scores for both imaging technologies, especially for the most expected case of a $0°$ camera angle toward the human body, when a drone evaluates each person during a rescue. Clearly, the algorithm detected more FPs than FNs, which is preferable because classifying wrongly moving people has less of an impact on rescuing people than wrongly identifying injured people who are not moving would. It mainly failed when the participants performed "slight" random movements. As for the FNs, most of them were in the "sitting" position for FNs and in the "laying" position for FPs. For the most expected angle between each human body and the camera ($0°$), the algorithm had high MMC values of 0.943 and 0.913 with RGB and IR imaging, respectively.

In conclusion, this preliminary study evaluated the performance of the proposed algorithm, using two image modalities, to detect unconscious victims in an outside environment using static cameras. In a future study, the algorithms should be tested using video data acquired from a UAV. The purpose is to use RGB imaging for daytime rescues and IR imaging for nighttime rescues. Indeed, graphic techniques as in the form of smart photographs have still a lot to offer, especially in disaster medicine, either by using solely or merging both AI and image processing techniques. Some existing detectors in literature are already being applied in the health monitoring field including heart rate, respiratory rate, and eye blink detection algorithms. With the presented algorithm, we aim to evaluate automatically and objectively (instead of subjectively) one common triage parameter (consciousness) for the purpose of triage and rescue optimization, while easing the work of paramedics. In this wise, this algorithm along other vital parameters detectors, could provide a quicker and more reliable information to rescuers and efficiently prioritize the most severely ill patients. Specifically, using an UAV, such system could remotely suggest an holistic patient categorization and well-being report before the medical team arrives the disaster site. Lastly, remote imaging systems can be applied to areas of expertise other than the medical field.

Author Contributions: Conceptualization, A.F. and M.C.; methodology, D.Q.P., C.B.P.; software, D.Q.P.; validation, D.Q.P. and L.M.; formal analysis, D.Q.P.; investigation, D.Q.P. and A.F.; resources, D.Q.P., M.C.; data curation, D.Q.P. and C.B.P.; writing—original draft preparation, D.Q.P.; writing—review and editing, all authors; visualization, D.Q.P.; supervision, C.B.P. and A.F.; project administration, A.F. and M.C.; funding acquisition, A.F. and M.C. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the local ethics committee Klinisches Ethik Komitee (KEK) and the Center for Translational and Clinical Research (CTC-A), with ethical approval number EK 322/21 (protocol code CTC-A, #21-028, approved 31 August 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ghanbari, V.; Ardalan, A.; Zareiyan, A.; Nejati, A.; Hanfling, D.; Bagheri, A. Ethical Prioritization of Patients during Disaster Triage: A Systematic Review of Current Evidence. *Int. Emerg. Nurs.* **2019**, *43*, 126–132. [CrossRef] [PubMed]
2. Bazyar, J.; Farrokhi, M.; Salari, A.; Khankeh, H. The Principles of Triage in Emergencies and Disasters: A Systematic Review. *Prehospital Disaster Med.* **2020**, *35*, 305–313. [CrossRef]
3. Bazyar, J.; Farrokhi, M.; Khankeh, H. Triage Systems in Mass Casualty Incidents and Disasters: A Review Study with A Worldwide Approach. *Open Access Maced. J. Med. Sci.* **2019**, *7*, 482–494. [CrossRef]
4. Follmann, A.; Rossaint, R.; Brokmann, J.C.; Beckers, S.K.; Czaplik, M. Remote Monitoring in Emergency Medical Services: A Report from so-Called Tele-Emergency Services Physician in Aachen. *Curr. Dir. Biomed. Eng.* **2017**, *3*, 479–481. [CrossRef]
5. Follmann, A.; Ohligs, M.; Hochhausen, N.; Beckers, S.K.; Rossaint, R.; Czaplik, M. Technical Support by Smart Glasses During a Mass Casualty Incident: A Randomized Controlled Simulation Trial on Technically Assisted Triage and Telemedical App Use in Disaster Medicine. *J. Med. Internet Res.* **2019**, *21*, e11939. [CrossRef] [PubMed]
6. Broach, J.; Hart, A.; Griswold, M.; Lai, J.; Boyer, E.W.; Skolnik, A.B.; Chai, P.R. Usability and Reliability of Smart Glasses for Secondary Triage During Mass Casualty Incidents. *Proc. Annu. Hawaii Int. Conf. Syst. Sci. Annu. Hawaii Int. Conf. Syst. Sci.* **2018**, *2018*, 1416–1422.
7. Kim, D.; You, S.; So, S.; Lee, J.; Yook, S.; Jang, D.P.; Kim, I.Y.; Park, E.; Cho, K.; Cha, W.C.; et al. A Data-Driven Artificial Intelligence Model for Remote Triage in the Prehospital Environment. *PLoS ONE* **2018**, *13*, e0206006. [CrossRef]
8. Al-Naji, A.; Perera, A.G.; Chahl, J. Remote Monitoring of Cardiorespiratory Signals from a Hovering Unmanned Aerial Vehicle. *Biomed. Eng. OnLine* **2017**, *16*. [CrossRef]
9. Heller, A.R.; Salvador, N.; Frank, M.; Schiffner, J.; Kipke, R.; Kleber, C. Diagnostic Precision of Triage Algorithms for Mass Casualty Incidents. English Version. *Anaesthesist* **2019**, *68*, 15–24. [CrossRef]
10. Curran-Sills, G.; Franc, J.M. A Pilot Study Examining the Speed and Accuracy of Triage for Simulated Disaster Patients in an Emergency Department Setting: Comparison of a Computerized Version of Canadian Triage Acuity Scale (CTAS) and Simple Triage and Rapid Treatment (START) Methods. *Can. J. Emerg. Med.* **2017**, *19*, 364–371. [CrossRef]
11. Farrohknia, N.; Castrén, M.; Ehrenberg, A.; Lind, L.; Oredsson, S.; Jonsson, H.; Asplund, K.; Göransson, K.E. Emergency Department Triage Scales and Their Components: A Systematic Review of the Scientific Evidence. *Scand. J. Trauma Resusc. Emerg. Med.* **2011**, *19*, 42. [CrossRef]
12. Olsson, T.; Terent, A.; Lind, L. Rapid Emergency Medicine Score: A New Prognostic Tool for in-Hospital Mortality in Nonsurgical Emergency Department Patients. *J. Intern. Med.* **2004**, *255*, 579–587. [CrossRef] [PubMed]
13. Goodacre, S.; Turner, J.; Nicholl, J. Prediction of Mortality among Emergency Medical Admissions. *Emerg. Med. J. EMJ* **2006**, *23*, 372–375. [CrossRef] [PubMed]
14. Völk, S.; Koedel, U.; Pfister, H.-W.; Schwankhart, R.; Op den Winkel, M.; Mühlbauer, K.; Klein, M. Impaired Consciousness in the Emergency Department. *Eur. Neurol.* **2018**, *80*, 179–186. [CrossRef]
15. Wolf, P.; Bigalke, M.; Graf, B.M.; Birkholz, T.; Dittmar, M.S. Evaluation of a Novel Algorithm for Primary Mass Casualty Triage by Paramedics in a Physician Manned EMS System: A Dummy Based Trial. *Scand. J. Trauma Resusc. Emerg. Med.* **2014**, *22*, 50. [CrossRef]
16. Development of the Aerial Remote Triage System Using Drones in Mass Casualty Scenarios: A Survey of International Experts. Available online: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0242947 (accessed on 22 July 2021).
17. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *arXiv* **2018**, arXiv:170306870.
18. Lin, T.-Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. *arXiv* **2015**, arXiv:14050312.
19. Hu, X.; Sun, B.; Zhao, H.; Xie, B.; Wu, H. Image Skeletonization Based on Curve Skeleton Extraction. In *Proceedings of the Human-Computer Interaction. Design and Development Approaches*; Jacko, J.A., Ed.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 580–587.
20. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*; Pearson: New York, NY, USA, 2018; ISBN 978-0-13-335672-4.
21. Lee, T.C.; Kashyap, R.L.; Chu, C.N. Building Skeleton Models via 3-D Medial Surface Axis Thinning Algorithms. *CVGIP Graph. Models Image Process.* **1994**, *56*, 462–478. [CrossRef]
22. Shi, J.; Tomasi. Good Features to Track. In Proceedings of the 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 21–23 June 1994; pp. 593–600.
23. Harris, C.; Stephens, M. A Combined Corner and Edge Detector. In Proceedings of the Fourth Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988; pp. 147–152.
24. Tomasi, C.; Kanade, T. Detection and Tracking of Point Features. *Int. J. Comput. Vis.* **1991**, *9*, 137–154. [CrossRef]
25. Bouguet, J. Pyramidal Implementation of the Affine Lucas Kanade Feature Tracker Description of the Algorithm. *Intel Corp.* **2001**, *5*, 4.
26. Álvarez, F.M.; Lora, A.T.; Muñoz, J.A.S.; Quintián, H.; Corchado, E. *International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on EUropean Transnational Education (ICEUTE 2019)*; Springer: Berlin/Heidelberg, Germany, 2019.
27. Osborne, J.W. *Best Practices in Quantitative Methods*; SAGE: Thousand Oaks, CA, USA, 2008; ISBN 978-1-4129-4065-8.

28. Cyr, L.; Francis, K. Measures of Clinical Agreement for Nominal and Categorical Data: The Kappa Coefficient. *Comput. Biol. Med.* **1992**, *22*, 239–246. [CrossRef]

29. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [CrossRef] [PubMed]

30. Wardhani, N.W.S.; Rochayani, M.Y.; Iriany, A.; Sulistyono, A.D.; Lestantyo, P. Cross-Validation Metrics for Evaluating Classification Performance on Imbalanced Data. In Proceedings of the 2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA), Tangerang, Indonesia, 23–24 October 2019; pp. 14–18.

31. Hu, B.-G.; Dong, W.-M. A Study on Cost Behaviors of Binary Classification Measures in Class-Imbalanced Problems. *arXiv* **2014**, arXiv:14037100.

32. Bekkar, M.; Djemaa, D.H.K.; Alitouche, T.A. Evaluation Measures for Models Assessment over Imbalanced Data Sets. *J. Inf. Eng. Appl.* **2013**, *13*, 27–38.

33. Saito, T.; Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* **2015**, *10*, e0118432. [CrossRef]

34. Jeni, L.A.; Cohn, J.F.; Torre, F.D.L. Facing Imbalanced Data–Recommendations for the Use of Performance Metrics. In Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, 2–5 September 2013; pp. 245–251.

35. Weiss, G.M. Foundations of Imbalanced Learning. In *Imbalanced Learning*; John Wiley & Sons, Ltd: Hoboken, NJ, USA, 2013; pp. 13–41, ISBN 978-1-118-64610-6.

36. Fu, G.; Yi, L.; Pan, J. Tuning Model Parameters in Class-Imbalanced Learning with Precision-Recall Curve. *Biom. J.* **2019**, *61*, 652–664. [CrossRef] [PubMed]

37. He, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [CrossRef]