A System for the Detection of Persons in Intelligent Buildings Using Camera Systems—A Comparative Study

Miroslav Schneider®, Zdenek Machacek®, Radek Martinek®, Jiri Koziorek© and Rene Jaros *

Department of Cybernetics and Biomedical Engineering, Faculty of Electrical Engineering and Computer Science, VSB–Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava, Czech Republic; miroslav.schneider@vsb.cz (M.S.); zdenek.machacek@vsb.cz (Z.M.); radek.martinek@vsb.cz (R.M.); jiri.koziorek@vsb.cz (J.K.)
* Correspondence: rene.jaros@vsb.cz

Received: 18 May 2020; Accepted: 22 June 2020; Published: 23 June 2020

Abstract: This article deals with the design and implementation of a prototype of an efficient Low-Cost, Low-Power, Low Complexity—hereinafter (L-CPC) an image recognition system for person detection. The developed and presented methods for processing, analyzing and recognition are designed exactly for inbuilt devices (e.g., motion sensor, identification of property and other specific applications), which will comply with the requirements of intelligent building technologies. The paper describes detection methods using a static background, where, during the search for people, the background image field being compared does not change, and a dynamic background, where the background image field is continually adjusted or complemented by objects merging into the background. The results are compared with the output of the Horn-Schunck algorithm applied using the principle of optical flow. The possible objects detected are subsequently stored and evaluated in the actual algorithm described. The detection results, using the change detection methods, are then evaluated using the Saaty method in order to determine the most successful configuration of the entire detection system. Each of the configurations used was also tested on a video sequence divided into a total of 12 story sections, in which the normal activities of people inside the intelligent building were simulated.

Keywords: object detection; person detection; motion detection; background; static background; dynamic background

1. Introduction

Camera systems are increasingly found in the interior of not only intelligent buildings. These systems are most often used for obtaining a video signal with the aim of recording or online streaming to the control room of the security service, that is, in order to protect people or property, which is the subject of References [1,2]. However, such a video signal can be used for detection or identification of persons present in the scene being captured after certain processing [3–11].

This article deals with detection of people in a scene being captured of the interior space of a building. The aim of this study is to create our own laboratory solution for the detection of persons in the room with the subsequent analysis of the data applicable for the intelligent building control system [12–17]. Other approaches to dealing with the issue of person detection apply principles different from detection from the image signal. For example, References [18,19] discuss modern methods of 3D detection of persons using the N2-3DDV-Hop method. Such and other methods
described in References [18,20] offer position information in 3-dimensional space. In contrast to these methods, image detection is focused on information from the projection of space into a 2D image matrix. In Reference [21], a similar problem is addressed using real hardware. However, image processing is handled differently here. The solution described in this article is intended for implementation using a built-in system hardware type. Detection should work without the use of complex algorithms, using convolutional methods [22,23] such as neural networks [24–26], using neural classifiers for the purpose of detecting persons as in Reference [27], or methods of detection of persons using the Haar-cascade classifier as in Reference [28], with the simplest possible subsequent implementation on built-in devices [29–31]. Built-in devices for detecting persons can be based on the use of image sensors, similarly as described in solutions [32–34], in the form of a smart camera sensor with the function of preprocessing image data into binary images with white dots indicating the position of a person in the scene, a smart camera sensor for detecting the background of the scene and foreground of the scene as the position of the found person as in Reference [35], smart camera sensor with the function of neglecting the dynamic background as in Reference [36], a smart camera sensor for Histogram of Oriented Gradients (HOG) image data processing as in Reference [37], or a specialized solution of the System on Chip (SOC) coping with basic image processing tasks such as edge detection in References [38,39], as an edge detector [40], or a solution with a low-power smart CMOS image sensor used to detect persons for indoor and outdoor use as in Reference [41]. Other image processing solutions, such as edge detection using digital parallel pulse computation [42], a non-parallel Sobel edge detector addressed by a smart camera sensor [43], discuss similar solutions that serve as sources of information providing a wide range of possible alternative solutions.

Finding the persons of the solution stated consists in finding a suspicious object in the scene, checking that the object is of a sufficient size, and checking that it may be an existing finding. This is followed by evaluation and printing of the information obtained for further possible processing, for example by the intelligent building control system [44–46]. Detection and localization is described in References [47,48], which are texts on camera systems for finding objects in an image matrix; in References [49–51], the methods of locating objects/persons in the image are discussed. References [48,52] were used as sources in tracing, and text [2] was used as a source of information on tracing using calibration methods and inverse kinematics methods as another direction in dealing with the issue of tracing.

The system devised finds suspicious objects or potential persons by subtracting two frames from each other. For example, subtraction of the current and previous frame processed, when the methods creating a static or dynamic background (replacing the previous frame found in the basic model of the system) were tested. References [38,42,53] were used as sources for work with a dynamic background and subsequent subtraction of the area of interest from the background—addressing similar solutions as this work, but in a different way. The binarized difference in brightness levels of these frames is subsequently subjected to a check of the size of the findings. The difference is obtained based on the results of the difference frame given by the frame rate (fps) used and the thresholding algorithm sensitivity. Another solution applied is the use of the Horn-Schunck method of optical flow [54,55], whose output is, similarly as the difference method output, binarized and further processed by decision algorithms, see Reference [56]. The detection, localization and tracing solutions described in this text are addressed here using only a camera sensor, comparing to other methods using a laser scanner, ultrasonic sensors, PIR sensors, as in References [1,57–59].

Persons are detected based on the assumed dimensions of their silhouettes. Since a person’s silhouette can take different shapes and dimensions in different camera shots (depending on the optical distortion and the location of the lens in the room), the parameters of this detection filter must be set to a particular scene. They discuss, in more detail, the basic methods of image processing for systems with low energy demands [40,60,61]. A comparison of available camera sensors for low-power systems is provided in Reference [62], but another source of image data was used in this work. In this work, a scene from the ceiling camera (Type—ASUS ZenFone 2 ZE551ML mobile phone component)
located in the middle of the room using a 170° additional camera lens was applied. In this solution, an unpreprocessed video signal is used, that is, similar to the output of the specialized circuits mentioned in References [63,64]. The description of background detection algorithms, detection of suspicious findings and the algorithm of the final process of distinguishing the persons detected from other dynamic objects is provided in the following section.

The proposed methods in the final version of the solution are designed for low-power systems, ideally based on an SOC (System on a Chip) technology, which will be employed for the use in intelligent building applications, which is more suitable for this issue than using high-power PC graphics cards as in Reference [65]. Solutions using special chips addressing the basic tasks of image processing are provided, for example, in Reference [66], where such a system can be used for data preprocessing of a more complex system. The primary benefit of the study is the testing of the individual detection methods applied on a test video sequence including possible interfering light effects in the room. The solution proposed can be used for implementation on built-in devices.

This concept deals with the detection of persons from originally untransformed image data of an RGB camera for common use. The paper is focused on Low-Cost, Low-Power, Low Complexity—hereinafter (L-CPC) image recognition system research and it gives us limitation for possible colour model. RGB colour model for algorithms and methods comparison is chosen because of image CCD and CMOS sensors common output signal. There are other necessary additional calculations for transforming to HSV, HSB, HSL colour models [67–69]. Of course, these models could be better for recognition results, but it is not applicable for L-CPC image recognition system with request for the highest possible process speed and the lowest computational load [70,71]. The option of use in conjunction with the aforementioned methods would be possible for these purposes of the thermal infrared camera. The error factor entered in the form of brightness deviations (caused, for example, by sharp sunlight entering the room through the window) could thus be replaced by thermal deviations in the scanned part of the scene.

Section 2 describes differential methods for detecting persons from the input video signal. Our three detection methods using computation towards a static background, three methods using computation towards a dynamic background and a method using the principle of optical flow are described here.

Section 3 describes the applied methods of adjusting the detected findings and the method of their filtering, identification, and description.

Section 4 describes the methods of evaluating the quality of detection of persons of each of the described methods. The evaluated sections of the test video used are described here, followed by the evaluation of the detection quality.

Section 5 describes the overall evaluation of the work and the results contained therein.

2. Methods

The developed and implemented methods described in the article were specified according to the differences in the approach to detecting the image background of the original image sequence. This not quite traditional approach based on recognition algorithms only on the variability of background detection. This way was chosen by the authors of the research work mainly with regards to own experiences and from selected literature [54,55,72,73]. Traditionally from others, the procedure for background recognition is always described, but this is not a fundamental problem and the core of the research activities. From our point of view, the variability of background detection methods is crucial for human movements recognizing. During the research, we analysed several original background detection algorithms, which were inspired by the available literature. By adapting the methods for presented detection problem, we were able to design our own original solution for each of the algorithm. The individual designed and analysed methods are described in detail below and evaluated with respect to the object recognition accuracy and success rate.

The methods of differentiating the current frame from the background can be based on both a static background and a dynamic background, see References [72,74]. Although both methods may
have a plethora of variations of creation of the backgrounds used and their application in differential search for suspicious objects, there are a total of six custom solutions to this problem, three to find the difference towards the static background where each test frame is compared to a fixed frame, and three towards the dynamic background where each current test frame is compared to a background frame dynamically adjusted based on $n$-previous frames.

- The method of differentiating the current frame towards the static background, see Reference [72].
- The method of differentiating the current frame towards the dynamic background, see Reference [72].
- The method of optical flow, see Reference [55].

The changes detected in the frame are found to be appropriately thresholded and are then distributed to other parts of the program as a binary map of the objects found as described in the diagram of the activity applied in Figure 1.

![Image change detection method](image.png)

**Figure 1.** Diagram of frame change detection activity.

Appropriate selection of the threshold for adjusting the output image results in filtering out low-contrast changes. These changes may represent slight added shadows, or large changes in brightness may also be perceived as an undesirable finding. In order to increase the robustness of the filtration system, it would be advantageous to use two threshold levels, that is, multi-level thresholding. However, in the described solution, such filtration was not used in order to demonstrate the behaviour of the described detection methods.

The selected methods are structured according to the type of work with the background image data of the detected persons. The individual versions were set up for laboratory testing, within the test video sequence, so that each of the versions achieved the highest possible success rate in detecting people. The purpose of the mutual comparison of methods very susceptible to changes in the image and sudden changes of immune methods is to point out their possible applicability in various applications. Due to the complexity of the analysis of object recognition methods, we decided to introduce as many versions and types of methods as possible. These allow embedded systems with limited computing capabilities to recognize object recognition in the image. The only complete analyses of different method versions give us complex and unquestionable answer to the most optimal algorithms.
2.1. The Method of Differentiating the Current Frame towards the Static Background

For clarity purposes, the method is hereinafter referred to as Method 1. This method uses the differences between the current frame and the background frame. A static background is considered to be a background fixed for the duration of the program run or for a period otherwise specified. Typically, such a background is compared to more than one frame of the scene. The principle of such a method is illustrated in the diagram in Figure 2. The advantage of using this method is a considerable saving of memory space when, after the static background is finally found, it is necessary to keep only one frame of the final background.

![Diagram of the difference method activity towards the static background.](image)

**Figure 2.** Diagram of the difference method activity towards the static background.

Within this difference method of frame change detection, three methods are applied to create a static background, where the background is:

- A single unedited reference frame (hereinafter referred to as Method 1.1).
- A weighted average of the first \( n \)-frames (hereinafter referred to as Method 1.2).
- A background map supplemented by invariable regions of two consecutive frames made up of \( n \)-frames (hereinafter referred to as Method 1.3).

2.1.1. A Single Unedited Reference Frame (Method 1.1)

A background created in this way is usually given by the first frame of the video sequence. The frame intended for the background should be based on the static scene in the image, ideally a scene devoid of objects dynamically moving at the time of taking the picture. Such a condition for capturing the background image ensures a minimum error rate of the subsequent detection of dynamically moving objects in future frames of the video sequence. The principle of the method is illustrated in the diagram in Figure 3. However, using this method of finding a background for change detection in the frame is very susceptible to changes in the lighting conditions in the scene of the frame compared to this background. An example of using this method is shown in a series of frames Figure 4.
2.1.2. A Weighted Average of the First \( n \)-Frames (Method 1.2)

The background thus created is based on the gradual weighted averaging of the first \( n \)-frames. Due to frame averaging, the effect of changing the lighting conditions in the scene is minimized, thereby minimizing the effect of unwanted changes in pixel brightness in the frames being analyzed during the subsequent detection. The purpose of this method of background creation is also neglect of possible dynamically moving objects in the scene at the beginning of the video sequence being captured. An example of using this method is shown in a series of frames Figure 5. The background can be practically created even when dynamically moving objects are present in the scene. However, such objects can introduce significant errors in the subsequent results of the detection system.

In the system code applied according to Figure 6, a background weight having a value of 5 is used. The final background is produced from the first 10 frames of the video sequence.
2.1.3. A Background Map Supplemented by Invariable Regions of Two Consecutive Frames Made Up of \( n \)-Frames (Method 1.3)

The background thus created is based on the assumption that there may be dynamically moving persons or other objects in the scene while taking pictures of the scene in order to create a static background. The principle of such a method is to gradually find fixed regions between successive frames that are gradually put together into the final background map. This algorithm is described in the diagram below in Figure 7. If this is an ideal scene that will be completely unchanged during the first two frames of the video sequence, specifically in the entire area of the frame checked, the final background will consist of pixels of the same brightness levels as the original frames. Since practically such ideal conditions cannot be achieved, there are different regions between every two frames, from tens of pixels per frame to almost the entire area of the frames compared (such an error rate may be caused by either evaluating the lighting conditions in the scene by the sensing device or by changing the lighting conditions in the scene being captured itself). Technically, by adding a static-looking region to the background map in the final background thus produced, a frame of a static-looking dynamic object may be introduced as an error in the background creation method. For example, persons sitting still in the room.
To make the background algorithm functionality applicable also in non-ideal conditions, adjustable tolerance of the match of brightness level of shades of grey was added for the reasons of making a decision on the match of the pixel brightness value of the frames compared. An example of using this method is shown in a series of frames Figure 8.

(a) Background frame. (b) Difference in brightness between the currently checked frame and the background frame. (c) Currently checked frame. (d) Binarized difference in brightness between the currently checked frame and the background frame. Findings that reach a value of logical 1 (white in the frame field) are considered potential findings of a detected person.

Since, in the initial state, the background frame has a value of 255 in each pixel, that is, a brightness level corresponding to white, in the final frame of the background acquired, indeterminate regions or regions with a constant change between frames attain a brightness level of just 255. Such regions...
will then be evaluated as findings of objects, but of an error nature, in the subsequent binarization by thresholding.

2.2. The Method of Differentiating the Current Frame towards the Dynamic Background

For the sake of clarity, the method is referred to as Method 2. A background that is not compared twice to any current frame (i.e., the background is adjusted, if not changed completely, for each frame checked) is considered to be a dynamic background or, also, a variable background.

These sections describe the individual methods for the purpose of their mutual comparison. By comparing the described assumptions and the real experiment behaviour of the implemented methods, there is presented with a set of image processing options for the detection of persons by low-performance embedded camera systems used in intelligent buildings. Different methods are applied to create a various dynamic background, because it is very important for comparison of the most optimal way to background presentation. Recognition algorithm success rate for person detection directly depends on the optimal background definition. Based on these theoretical assumptions and own research experiences, we decide to analyse three methods with the most dissimilar results for background definition in our research. The chosen background detection methods are comparison frame previous to the current frame, average of all previous frames, average of $n$-previous frames. Each method brings advantages and disadvantages, which are described in text.

The difference between the current frame towards the background is calculated only after the initial background is created. A new background is created only after the current background is used for the image change detection algorithm (in the first pass of the differential part of the program, the current background is the initial background), that is, after the termination of the part of the program that uses the current background, to prevent image data collision. The new background is either a brand new frame or an existing background frame modified according to the background creation method used. Such a new background is an image function of the background of the old and new frames. The algorithm depicted above is also described in the diagram attached, in Figure 9.

![Figure 9. Diagram of the difference method activity of the current frame towards the dynamic background.](image-url)
Within this difference method of frame change detection, three methods are applied to create a dynamic background, where the background is:

- A frame previous to the current frame (hereinafter referred to as Method 2.1).
- Average of all previous frames (hereinafter referred to as Method 2.2).
- Average of \( n \)-previous frames (hereinafter referred to as Method 2.3).

2.2.1. A Frame Previous to the Current Frame (Method 2.1)

As shown in the background creation diagrams above, this background creation method always uses the previous frame as the background for the current frame. Frame change detection based on the difference towards such a background is focused primarily on dynamically moving objects, namely objects showing a constant change between frames. The algorithm depicted is also described by the diagrams in Figure 10. The detection of an ideally constantly moving object, which can be the person being detected, is exceptionally reliable from the perspective of the theoretical assumption of the functionality of the change detection method. A great advantage of such change detection is the high resistance to changes in the lighting conditions in the scene between frames of the video sequence. The disadvantage of frame change detection, based on the principle of the method, is the subsequent increase in the error rate when the movement of the person (object) in the image scene is stopped. However, such an object, amongst the frames being dealt with, looking like a static object, cannot be found then by the detection algorithm as an object moving dynamically in the entire video sequence. The results of the behaviour of this method are shown graphically in a series of frames in Figure 11.

![Figure 10. Diagrams of Method 2.1 activity.](image-url)

![Figure 11. Examples of Method 2.1. functionality.](image-url)
2.2.2. Average of All Previous Frames (Method 2.2)

The method of creating a dynamic background as the arithmetic mean of all the previous frames is a dynamic analogy to the above-mentioned method of creating a static background. The method is described by the diagrams in Figure 12. The background thus created is based on the gradual weighted averaging of all the frames of the video sequence. Due to frame averaging, the effect of changing the lighting conditions in the scene is minimized, thereby minimizing the effect of unwanted changes in pixel brightness in the current frames during the subsequent detection.

![Diagram](image)

**Figure 12.** Diagrams of Method 2.2 activity.

The purpose of this method of background creation is also neglect of possible dynamically moving objects in the scene by setting a weight greater than 1, that is, the existing background weight having a value of 1 when creating a new background. By preferring the information from the previous background, neglect of the new frame is caused as well as changes in the scene associated with this neglect. Therefore, if the change in the scene is constant, it will become less important with each newly created background due to the subsequent difference of the current frame towards the background. However, the difference towards such a background is not resistant to significant sudden changes in lighting conditions in the scene. It is evident from the principle of the method that the change in the lighting conditions in the scene must be as small as possible in order to minimize errors (methods of detecting the change in the background). The weight of the previous background frame towards the current frame is a particularly important factor. The greater the weight of the previous background frame, the more negligible the errors of the newly added current frame will be in the new background. The aim of this method is to create an idealized restoring background in which the error noise caused by dynamically moving objects in the scene of the original frame is eliminated. An example of using this method is illustrated in a series of frames Figure 13.

![Frame Examples](image)

**Figure 13.** Examples of Method 2.2 functionality. (a) Background frame. (b) Difference in brightness between the currently checked frame and the background frame. (c) Currently checked frame. (d) Binarized difference in brightness between the currently checked frame and the background frame. Findings that reach a value of logical 1 (white in the frame field) are considered potential findings of a detected person.
However, the differential frame change detection in the current image from such a background is not resistant to sudden changes in the lighting conditions in the scene. Such a sudden change in lighting conditions cannot be evaluated other than an error of the method, indicating a high error rate in the entire frame.

2.2.3. Average of n-Previous Frames (Method 2.3)

Similarly to the method of difference towards the dynamic background created by averaging of all the previous frames, the functionality of the method using a background created by averaging n-previous frames is dependent on setting the weighting parameter of the previous frame. In addition to this adjustable parameter, it is also necessary to determine the number of frames that will form the dynamic background of the current frame when the difference is detected in the image. The weight parameter of the previous background frame, when creating a new background, affects the gradual disappearance of images of dynamically moving objects or persons in the scene in order to achieve the most accurate background for the subsequent differential detection of changes in the current frame towards the background frame. The method is described by the diagrams shown in Figure 14, describing the creation of the initial background, and Figure 15 describing the creation of each new background.

![Diagram](image.png)

**Figure 14.** Diagram of Method 2.3 activity—creation of the initial background.

This method is designed to eliminate the effects of dynamic object images in the scene of the current frame involved in creating a new background frame. In contrast to the aforementioned method of averaging of all the previous frames of the current frame, a limited number of previous frames are used for the current dynamic background of the current frame. This prevents the past error from being brought into the current background. The weighting calculation setting is also adjusted so that the oldest background frame took the highest weight, thus bringing the highest error rate to the resulting frame. In this way, the error detection of dynamically moving objects in the scene, whose position dynamic change is expected in the scene of the current frame, is limited. Figure 16 shows an example of the output of this method.
Sensors 2020, 20, 3558

Methode 2.3 Activity
- Creating a next background

\[
TF(1:k) = \left( TF(1:k) \times w + NF \right) / \left( w + 1 \right)
\]

\[Bi++;
B = TF(Bi)
\]

\[TF(Bi) = NF
Bi == k
Last TF?
\]

\[Bi = 0;
\]

Figure 15. Diagram of Method 2.3 activity—creation of a new background.

Figure 16. Examples of Method 2.3 functionality. (a) Background frame. (b) Difference in brightness between the currently checked frame and the background frame. (c) Currently checked frame. (d) Binarized difference in brightness between the currently checked frame and the background frame. Findings that reach a value of logical 1 (white in the frame field) are considered potential findings of a detected person.

Of the above-mentioned methods, this method is the most memory-intensive method, because it is necessary to preserve just \(n\)-frames of the temporary background for the background created from \(n\)-previous frames (background frames are gradually completed according to the algorithms described above).

2.3. The Method of Optical Flow (Method 3)

For the sake of clarity, the method is referred to as Method 3. This is a method developed to detect dynamically moving objects in a scene. The program calculates a velocity vector that reflects the direction and magnitude of the velocity of motion in the image using the optical flow method and the resulting brightness function calculation. However, from this velocity vector, information about the velocity vector magnitude at each point of the image examined is taken for further processing. This information of potentially moving objects is then thresholded in order to obtain binary images of the findings. The application of the optical flow method is described in the diagram below in Figure 17.
This is a method that is very sensitive to changes in the lighting conditions in the scene of the input frames. In contrast to the difference methods, which deal with the change in the image consisting of greyscale input frames, the method used deals with finding the change between the input frames, the current frame, and the background frame, in the original RGB colour model. Since a third-party code is used, the algorithm of this method will not be described in detail in this work. The Horn-Schunck optical flow solution algorithm is available on the MathWorks MATLAB support site [54,55,73]. An example of applying the method described to a moving object detection algorithm is shown in a series of frames in Figure 18.
3. Adjustment of the Objects Found

This part of adjustment of the block of the objects found is designed to adjust the dimensions and integrity of the objects found. This adjustment is achieved by means of the following steps:

1. Filter erosion of the objects found to remove image noise findings.
2. Dilatation of the objects found.
3. Erosion of the object found.

Each of these steps has its reason in the final shape and size of the object found. Filtering the findings caused by noise in the image using erosion is a method of minimizing the potential for increasing error rates in the subsequent dilatation step of the objects found. The work uses erosion by three pixels for this purpose, as shown in Figure 19, that is, each of the objects found is reduced line by line, always by one pixel of a positive finding.

![Figure 19. Example of adjustment of objects found in the binary image by erosion filtering. Erosion filtering of objects by 3 pixels. (a) The original binary image of the findings after binarization of the difference finding. (b) Image (a) filtered by noise objects using erosion.](image)

During dilatation, the size of each object finding is extended by a user-adjustable factor of the region around the particular finding point. This step is used to combine discontinuous object findings into an undivided object. The frame dilatation effect in Figure 19b is shown in Figure 20a. However, the adjustment of the aforementioned dilatation region factor is different for the resulting binary image of each of the methods of frame change detection used, as needed. For example, when using the change detection method by means of optical flow, a significant expansion of the findings is necessary, since the findings belonging to the location of the person in the scene of the current frame may occur separately from the findings of objects belonging to the same person in the scene.

![Figure 20. Example of adjustment of objects found in the binary image by dilatation and reverse erosion of objects by a rectangular area of 20 × 20 pixels, both in the horizontal and vertical directions. (a) The result of the filtered finding dilatation. (b) The result of the reverse erosion of the binary image (a).](image)
If added shadows occur, the success rate of the methods selected for the correct detection of persons in the image is often reduced. In the case of differential methods, such as Method 1.1, Method 1.2, Method 1.3, Method 2.1, Method 2.2, Method 2.3, the shadow could be perceived as a new detected person. To evaluate a sufficient change in the image, sufficient to evaluate a new person, it is thus crucial to set a threshold level for filtering the input data. To clarify this possible phenomenon, a text was added to the article (Section 2) explaining the behaviour of the described methods in the case of the occurrence of excessive shadows. In addition, the practical implementation envisages the addition of an algorithm comprising definition of areas where newcomers can appear and where they can leave the room. The article no longer deals with this functional superstructure, but it will significantly contribute to the elimination of the above-mentioned phenomenon.

3.1. Identification of Object Findings

The binary image of the objects adjusted is then subjected to the identification of the individual object findings.

In the first stage of identification, the positive pixels (belonging to the potential person) of each continuous pixel cluster according to the 4-neighbour rule are grouped into regions identified by the original identifier.

The next stage of the identification of the object found is marking into the delimited regions and their possible joining into larger units. The identified object regions found are further distributed by the vector description of the rectangular areas, that is, the description of two pixels in the image. The first, initial, point is given by the smallest horizontal pixel coordinate and the smallest vertical pixel coordinate of the object described. The second point is then given by the largest horizontal pixel coordinate and the largest vertical pixel coordinate of the object described. The principle of marking the object found is explained in the attached illustration in Figure 21.

![Figure 21](image-url) Illustration describing the marking of the object found by a rectangular marking field.

The objects found that are thus delimited by the rectangular areas are then checked for intersection with another delimited region in the delimitation area. If they have intersection areas, they are then unified into a common region of the object, as shown in the illustration in Figure 22.

The information about the object found is inserted in a two-dimensional description field of the objects found in the frame currently checked for the reasons of further analysis. After the check and unification of the objects into marking regions are completed, the applied image is no longer used by the algorithm, as other parts of the algorithm only work with the so-called rectangular area marking vectors.
### 3.2. Filtration of Error Findings from Findings of Persons

The object findings adjusted and identified are filtered according to the dimensions of the custom rectangular delimitation and the actual pixel area of the positive pixels in the marking field.

In the above-mentioned illustration in Figure 23, vectors representing regions indicating object findings are plotted. Vectors denoted as $\vec{a}_1$ and $\vec{a}_2$ represent regions smaller than the minimum dimensions of $x_{\text{min}}$ and $y_{\text{min}}$, specified by the condition for filtering object findings from findings of persons. Vectors denoted as $\vec{b}_1$ and $\vec{b}_2$ represent regions, one dimension of which is less than the minimum dimension limit of $x_{\text{min}}$ or $y_{\text{min}}$, specified by the condition for filtering object findings from findings of persons. Vectors denoted as $\vec{d}_1$ and $\vec{d}_2$ represent regions, at least one dimension of which is more than the maximum dimension limit of $x_{\text{max}}$ or $y_{\text{max}}$, specified by the condition for filtering object findings from findings of persons. Thus, the findings delimited by the regions represented by vectors denoted as $\vec{a}_1$, $\vec{a}_2$, $\vec{b}_1$, $\vec{b}_2$, $\vec{d}_1$, and $\vec{d}_2$ do not meet the requirements of the filter criteria set according to the dimensions and are, therefore, not further processed as findings of persons. In contrast, vectors denoted as $\vec{c}_1$ and $\vec{c}_2$ represent regions acquiring both dimensions greater than the minimum dimensions of $x_{\text{min}}$ and $y_{\text{min}}$ and smaller than the dimensions of $x_{\text{max}}$ or $y_{\text{max}}$, specified by the condition for filtering object findings from findings of persons. Such object findings are further analyzed as findings of persons.

**Figure 22.** Illustrations describing the unification of the intersecting objects found into the final object.

**Figure 23.** Illustration of filtering of the objects found according to the dimensions of their rectangular delimitations described by a descriptive vector.
3.3. Cataloguing of Persons Found

In the block of cataloguing, the persons found, represented by a description of their areas of location and size, are identified, and subsequently monitored to see if these findings are still current in the scene. It is in this part of the system for the detection of persons in intelligent buildings that the findings of persons from current frames are set in context with the findings of persons from the previous frames. The cataloguing block works with a two-dimensional array of information, in which each current finding is assigned with one row of this array of information. The array of information is the bearer of information about the persons currently found as well as error findings in the scene, which are assigned with a lifespan value of each of the findings within this block.

The index that is assigned to each person newly found or a person currently restored at the maximum possible user-selectable value, called the “maximum lifespan of a person finding”, is designated as the lifespan of person findings as shown in the frame in Figure 24a. Each time the program loop passes through, the lifespan value of each person found is decremented as shown in the frame in Figure 24b. If the person finding in the array of current findings is no longer restored during the period of his/her non-zero lifespan and the person finding has a lifespan of zero, the person finding is removed from the array of current findings as an outdated or error finding.

![Frame number 42.](image1) ![Frame number 43.](image2)

**Figure 24.** Example of reducing the lifespan at an error finding. The illustration shows the real results of Method 1.1, with a lifespan of 1s, i.e., 4 frames at a frame rate of 4 fps.

As part of its identification, each of the new person findings is compared to the person findings listed in the array of information about the currently occurring person findings in the scene. The evaluation of conformity of the person findings is conducted by means of detecting the intersections of the mutually checked rectangular areas of the delimitations of the findings. The effect of detecting the intersections of the areas found is shown in the frames in Figure 25. In the case of finding the intersection of the areas checked, the existing finding is overwritten by the new finding and supplemented with the maximum lifespan of the person finding.

In the example below, illustrating the structure of the individual output data of the final system, the information characterizing the two objects from Figure 24 is shown, namely Figure 24b, when one of the objects/potential-persons found acquires a 100% lifespan, wherein the lifespan of the other one is already reduced compared to the frame in Figure 24a, where both findings have a 100% lifespan.

Table 1 provides an example of all current person findings, in the frame in Figure 24b, including their current lifespan. Since, when producing these examples, the maximum, that is, 100% lifespan, is set to 4, that is, 4 frames, the most current person finding, mentioned above in the tables of the initial findings and unified person findings, is correctly evaluated by the lifespan index in the first row. The next finding, having a lifespan index value of 3, is the finding that was last updated during the previous program pass, that is, in the previous input frame. Furthermore, the table of output information transferred to the superior system includes data on: the number of custom pixels belonging to the person found and information on location of the individual findings in the matrix of the image checked.

---

**Table 1.** Example of all current person findings, in the frame in Figure 24b, including their current lifespan.
4. Results

For each of the methods used, the person detection system was applied to the same video sequence, which was designed to simulate the normal operating conditions of a camera in intelligent buildings. The test video was divided into 12 partial event sections, which were individually evaluated and statistically processed.

Table 2 depicting the test video event sections describes the activities of the persons in each section. These test video sections are designed to capture the most common possible behaviour of persons in a scene being captured by the camera system of an intelligent building. It includes entrance, exit, movement of persons, occasional stopping of persons and shifting of objects in the scene.

| Video Section | Section Interval (Video Frames) | Actions in the Section                                                                 |
|---------------|---------------------------------|----------------------------------------------------------------------------------------|
| 1             | 1 to 6                           | empty room with no movement                                                            |
|               | 7 to 30                          | door opening, entrance of person no.1, door closing                                     |
| 2             | 31 to 40                         | person no.1—walking across the room                                                   |
|               | 41 to 48                         | person no.1—grabbing an object (a folder on the bed)                                   |
| 3             | 49 to 70                         | person no.1—sitting down, end of movement                                              |
| 4             | 71 to 73                         | person no.1—at rest                                                                   |
|               | 74 to 84                         | person no.1—at rest, door opening, entrance of person no.2                             |
|               | 87 to 96                         | person no.1—at rest, door opening, entrance of person no.2                             |
| 5             | 97 to 134                        | person no.1—at rest, person no.2—shifting a chair and sitting down                    |
|               | 135 to 142                       | person no.1—at rest, person no.2—at rest                                             |
| 6             | 143 to 161                       | person no.1—walking across the room, person no.2—at rest                              |
|               | 162 to 172                       | person no.1—walking across the room, person no.2—at rest, intersecting of persons in the image |
|               | 173 to 188                       | person no.1—at rest, person no.2—at rest, intersecting of persons in the image        |
|               | 189 to 190                       | person no.1—walking across the room, person no.2—at rest, intersecting of persons in the image |
Table 2. Cont.

| Video Section | Section Interval (Video Frames) | Actions in the Section |
|---------------|---------------------------------|-------------------------|
| 7             | 191 to 215 person no.1—walking across the room, person no.2—at rest | 216 to 229 person no.1—pushing a toy car, person no.2—at rest |
|               | 230 to 239 person no.1—at rest, person no.2—at rest | 216 to 229 person no.1—pushing a toy car, person no.2—at rest |
| 8             | 240 to 268 person no.1—at rest, person no.2—walking across the room | 269 to 272 person no.1—at rest, person no.2—at rest |
|               | 273 to 281 person no.1—at rest, person no.2—at rest, darkening of the room using blinds | 282 to 296 person no.1—at rest, person no.2—at rest, brightening of the room using blinds |
| 9             | 273 to 281 person no.1—at rest, person no.2—at rest, darkening of the room using blinds | 297 to 319 person no.1—walking across the room (exit), person no.2—at rest |
| 10            | 297 to 319 person no.1—walking across the room (exit), person no.2—at rest | 320 to 339 person no.2—walking across the room (exit) |
| 11            | 320 to 339 person no.2—walking across the room (exit) | 340 to 346 door closing |
|               | 347 to 350 empty room with no movement | 347 to 350 empty room with no movement |

4.1. Verification of the Results

In order to determine the correct function of the algorithms applied, it was necessary to create a verification template of the results. The results of the verification had to be generated for each of the frames tested so as to sufficiently assess the quality of the detection of the persons. In the test video sequence, the verification of the results achieved was thus conducted on the frame-by-frame basis.

The final weighted detection quality assessment \( K \) is a result of the sum of the individual weights of the quality criteria \( K \) multiplied by a percentage of the quality criteria met \( p_K \). The weights of the quality criteria are determined by the Saaty multi-criteria evaluation method according to the source cited below.

The values of the criteria weights are determined using geometric means of the rows of the Saaty matrix, plotted in the Table 3. If these row geometric means are standardized, the standardized weights of the set of criteria being dealt with are obtained [15].

\[
v_i = \frac{G_i}{\sum_{i=1}^{n} G_i}, \tag{1}\]

where \( v_i \) is the standardized weight of the i-th criterion, \( G_i \) is the geometric mean of the i-th criterion, and \( n \) is the total number of criteria.

The work uses five quality criteria. Each of these criteria stated expresses the correctness of person findings detected in the scene being captured.

- \( K_1 \) Criterion: Average percentage of custom pixels of person findings in the field for marking persons
- \( K_2 \) Criterion: Average lifespan of person findings
- \( K_3 \) Criterion: Average percentage intersection of the areas of marking of the person finding with the verification pattern
- \( K_4 \) Criterion: Average error of location of marking of person findings with the intersection in the verification pattern
- \( K_5 \) Criterion: Success rate of the detection of correct person findings in relation to all current findings.
The values provided in the rows and columns of Table 3 are based on the importance of one criterion in relation to another. There are values of 1 in the diagonal because criterion $K_1$ in the row is equivalent to criterion $K_1$ in the column. The highest criterion is a value of 9, wherein the criterion acquiring the value of 9 is absolutely more significant than the criterion in the relevant column.

**Table 3.** Saaty matrix with geometric means calculated, and standardized weights of the individual criteria.

|   | $K_1$ | $K_2$ | $K_3$ | $K_4$ | $K_5$ | $G_i$ | $v_i$ |
|---|---|---|---|---|---|---|---|
| $K_1$ | 1 | 0.2 | 0.111 | 0.333 | 0.111 | 0.241 | 0.0351 |
| $K_2$ | 1 | 5 | 0.166 | 3 | 1 | 1.200 | 0.1746 |
| $K_3$ | 6 | 1 | 3 | 2 | 3.178 | 0.4626 |
| $K_4$ | 3 | 0.333 | 0.333 | 1 | 0.5 | 0.699 | 0.1017 |
| $K_5$ | 9 | 1 | 0.5 | 2 | 1 | 1.552 | 0.2260 |

The preferences between the individual criteria are determined expertly according to the requirement for strictness of the final weighted detection quality assessment. Since it is necessary to include the information on whether or not the quality assessment criteria are met in such a system of weights devised in order to conduct a cumulative assessment of the system detection quality results, the resulting weighted quality assessment is obtained by the following calculation:

$$K = \sum_{i=1}^{n} \left( v_i \cdot \frac{p_{ki}}{100} \right),$$

where the weighted detection quality assessment is denoted as $K$, $v_i$ is the standardized weight of the $i$-th criterion, and $p_{ki}$ is the percentage of the $i$-th criterion.

By using $p_{ki}$ as a percentage of meeting the $i$-th criterion, the resulting value of the weighted quality assessment $K$ is in the range of 0 to 1. The weighted detection quality assessment $K$ equals zero if each of the percentages of the $i$-th criterion equals zero, or the person finding detected is negative, or, also, ideally incorrect. Conversely, $K$ equals 1 if the ideal person finding detected is correct. As it is obvious from the following relationships:

$$K_{\text{max}} = \sum_{i=1}^{n} \left( v_i \cdot \frac{p_{k_i \text{max}}}{100} \right) = 1,$$

$$K_{\text{min}} = \sum_{i=1}^{n} \left( v_i \cdot \frac{p_{k_i \text{min}}}{100} \right) = 0.$$ 

In practice, however, the ideal state where $K$ equals 1 is very difficult to achieve. This is because the percentage of meeting the $i$-th criterion, $p_{ki}$, is below 100% of the ideal state for each person detection quality assessment criterion.

The example shown in the figure (Figure 26) indicates the distribution and importance of the individual weights of the weighted detection quality rating to the final rating ($K$). Table 4 is shown as an example of the summary results of the selected method (Method 1.1) in the individual sections, including the average measurement error.
The following graphs of the results of Methods 1.1, 1.2, 1.3, 2.1, 2.2, 2.3, 3 and the results of each of them when changing the lifespan parameter of the persons found show that the optimal value of the lifespan parameter of the persons found plays a significant role for each of the detection algorithm configurations applied.

As it is obvious from the example of the results of the average weighted rating of Method 1.1 shown in Table 4, this functional configuration achieved the best rating with a lifespan of the person found of 12 frames (i.e., a lifespan of 3 s at 4 fps), with an average measurement error of 7.5%. The results shown in Table 4 are indicated in the attached figures (Figures 27a, 28b and 29a).

Many methods, as illustrated in the resulting graphs, from Figures 27–47, had a significantly increased error rate in the first and last section. This is caused by the inconsistent presence of detectable persons in the room being checked and their erroneous assessment, either due to the background creation delay or the delay caused by the excessively long lifespan of the persons detected.

Table 4. The results of the weighted person detection assessment using Method 1.1.

| Section Number | Weighted Average Rating | The Mean Error of Measurement Rating |
|----------------|-------------------------|--------------------------------------|
|                | Lifespan of 1 Frame     | Lifespan of 12 Frames                | Lifespan of 40 Frames                | Lifespan of 1 Frame     | Lifespan of 12 Frames                | Lifespan of 40 Frames                |
| 1              | 0.477                   | 0.477                                | 0.477                                | 0.294                   | 0.294                                | 0.294                                |
| 2              | 0.623                   | 0.620                                | 0.541                                | 0.030                   | 0.038                                | 0.058                                |
| 3              | 0.595                   | 0.552                                | 0.533                                | 0.030                   | 0.027                                | 0.029                                |
| 4              | 0.504                   | 0.493                                | 0.479                                | 0.040                   | 0.030                                | 0.030                                |
| 5              | 0.505                   | 0.478                                | 0.479                                | 0.028                   | 0.030                                | 0.030                                |
| 6              | 0.505                   | 0.478                                | 0.479                                | 0.028                   | 0.030                                | 0.030                                |
| 7              | 0.337                   | 0.411                                | 0.424                                | 0.156                   | 0.059                                | 0.062                                |
| 8              | 0.555                   | 0.552                                | 0.553                                | 0.030                   | 0.027                                | 0.029                                |
| 9              | 0.459                   | 0.487                                | 0.488                                | 0.077                   | 0.045                                | 0.044                                |
| 10             | 0.469                   | 0.475                                | 0.391                                | 0.071                   | 0.056                                | 0.040                                |
| 11             | 0.262                   | 0.456                                | 0.432                                | 0.202                   | 0.096                                | 0.120                                |
| 12             | 0.262                   | 0.456                                | 0.432                                | 0.202                   | 0.096                                | 0.120                                |

Average of rating of all sections 0.481 0.487 0.470 0.091 0.075 0.073

Final ranking, from top-rated 2 1 3

Figure 26. Distribution of the individual criterion weights in the practical result of the video section.
4.1.1. Evaluation of Method 1.1

By comparing the results of Method 1.1, evaluation at a lifespan of 1, 12, and 40 frames, it is ascertained that the best detection results were achieved within video sections 2 to 5. A graphical representation of the evaluation of the configurations used is shown, for results with a lifespan of 1 frame, in Figure 27, for results with a lifespan of 12 frames, in Figure 28, for results with a lifespan of 40 frames in Figure 29. In subfigure (b), the selected figures show boxplots of the statistical distribution of the weighted evaluation in selected sections into quartiles. The final summary of the results of all the methods in Table 5 includes the weighted average evaluation results for all the sections of the video, wherein the best weighted and scaled scores were achieved with Method 1.2 having a lifespan of 12 frames.

Figure 27. Results of Method 1.1 with a lifespan of 1 frame. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.

Figure 28. Results of Method 1.1 with a lifespan of 12 frames. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.

Figure 29. Results of Method 1.1 with a lifespan of 40 frames. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.
4.1.2. Evaluation of Method 1.2

By comparing the results of Method 1.2, evaluation at a lifespan of 1, 12, and 40 frames, it is ascertained that the best detection results were achieved within video sections 2 to 4. A graphical representation of the evaluation of the configurations used is shown, for results with a lifespan of 1 frame, in Figure 30, for results with a lifespan of 12 frames, in Figure 31, for results with a lifespan of 40 frames in Figure 32. In subfigure (b), the selected figures show boxplots of the statistical distribution of the weighted evaluation in selected sections into quartiles. The final summary of the results of all the methods in Table 5 includes the weighted average evaluation results for all the sections of the video, wherein the best weighted and scaled scores were achieved with Method 1.2 having a lifespan of 12 frames.

![Figure 30](image1.png)

Figure 30. Results of Method 1.2 with a lifespan of 1 frame. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.

![Figure 31](image2.png)

Figure 31. Results of Method 1.2 with a lifespan of 12 frames. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.

![Figure 32](image3.png)

Figure 32. Results of Method 1.2 with a lifespan of 40 frames. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.
4.1.3. Evaluation of Method 1.3

By comparing the results of Method 1.3, evaluation at a lifespan of 1, 12, and 40 frames, it is ascertained that the best detection results were achieved within video sections 2 to 4. A graphical representation of the evaluation of the configurations used is shown, for results with a lifespan of 1 frame, in Figure 33, for results with a lifespan of 12 frames, in Figure 34, for results with a lifespan of 40 frames in Figure 35. In subfigure (b), the selected figures show boxplots of the statistical distribution of the weighted evaluation in selected sections into quartiles. The final summary of the results of all the methods in Table 5 includes the weighted average evaluation results for all the sections of the video, wherein the best weighted scores were achieved with Method 1.3 having a lifespan of 12 frames; the best scaled scores were achieved with a lifespan of 40 frames.

Figure 33. Results of Method 1.3 with a lifespan of 1 frames. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.

Figure 34. Results of Method 1.3 with a lifespan of 12 frames. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.

Figure 35. Results of Method 1.3 with a lifespan of 40 frames. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.
4.1.4. Evaluation of Method 2.1

By comparing the results of Method 2.1, evaluation at a lifespan of 1, 12, and 40 frames, it is ascertained that, within the lifespan increase, an increase in the stability of the results of the weighted detection assessment in the individual sections was achieved. A graphical representation of the evaluation of the configurations used is shown, for results with a lifespan of 1 frame, in Figure 36, for results with a lifespan of 12 frames, in Figure 37, for results with a lifespan of 40 frames in Figure 38. In subfigure (b), the selected figures show boxplots of the statistical distribution of the weighted evaluation in selected sections into quartiles. The final summary of the results of all the methods in Table 5 includes the weighted average evaluation results for all the sections of the video, wherein the best weighted scores were achieved with Method 2.1 having a lifespan of 12 frames; the best scaled scores were achieved with a lifespan of 40 frames. As it is obvious from the results in both Figure 36 and Table 5, Method 2.1 achieves very poor results with a lifespan of findings lasting 1 frame.

![Figure 36](image1.png)

**Figure 36.** Results of Method 2.1 with a lifespan of 1 frames. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.

![Figure 37](image2.png)

**Figure 37.** Results of Method 2.1 with a lifespan of 12 frames. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.
4.1.5. Evaluation of Method 2.2

By comparing the results of Method 2.2, evaluation at a lifespan of 1, 12, and 40 frames, it is ascertained that, within the lifespan increase, an increase in the stability of the results of the weighted detection assessment in the individual sections was achieved. A graphical representation of the evaluation of the configurations used is shown, for results with a lifespan of 1 frame, in Figure 39, for results with a lifespan of 12 frames, in Figure 40, for results with a lifespan of 40 frames in Figure 41. In subfigure (b), the selected figures show boxplots of the statistical distribution of the weighted evaluation in selected sections into quartiles. The final summary of the results of all the methods in Table 5 includes the weighted average evaluation results for all the sections of the video, wherein the best weighted scores were achieved with Method 2.2 having a lifespan of 12 frames; the best scaled scores were achieved with a lifespan of 40 frames. As it is obvious from the results in both Figure 39 and Table 5, Method 2.2 achieves very poor results with a lifespan of findings lasting 1 frame.

Figure 39. Results of Method 2.2 with a lifespan of 1 frames. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.
4.1.6. Evaluation of Method 2.3

By comparing the results of Method 2.3, evaluation at a lifespan of 1, 12, and 40 frames, it is ascertained that, within the lifespan increase, an increase in the stability of the results of the weighted detection assessment in the individual sections was achieved. A graphical representation of the evaluation of the configurations used is shown, for results with a lifespan of 1 frame, in Figure 42, for results with a lifespan of 12 frames, in Figure 43, for results with a lifespan of 40 frames in Figure 44. In subfigure (b), the selected figures show boxplots of the statistical distribution of the weighted evaluation in selected sections into quartiles. The final summary of the results of all the methods in Table 5 includes the weighted average evaluation results for all the sections of the video, wherein the best weighted and scaled scores were achieved with Method 2.3 having a lifespan of 12 frames. As it is obvious from comparing the results in Figures 43 and 44, Method 2.3 achieves the best score when applying a lifespan of findings of 12 frames. The error rate of the assessment increases as the lifespan extends.
By comparing the results of Method 3, evaluation at a lifespan of 1, 12, and 40 frames, it is ascertained that, within the lifespan increase, an increase in the stability of the results of the weighted detection assessment in the individual sections was achieved. A graphical representation of the evaluation of the configurations used is shown, for results with a lifespan of 1 frame, in Figure 45, for results with a lifespan of 12 frames, in Figure 46, for results with a lifespan of 40 frames in Figure 47.
In subfigure (b), the selected figures show boxplots of the statistical distribution of the weighted evaluation in selected sections into quartiles. The final summary of the results of all the methods in Table 5 includes the weighted average evaluation results for all the sections of the video, wherein the best weighted scores were achieved with Method 3 having a lifespan of 12 frames; the best scaled scores were achieved with a lifespan of 40 frames.

Figure 45. Results of Method 3 with a lifespan of 1 frames. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.

Figure 46. Results of Method 3 with a lifespan of 12 frames. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.

Figure 47. Results of Method 3 with a lifespan of 40 frames. (a) the weighted average rating with the average mean absolute error of the weighted rating and (b) boxplots of the distribution of the statistical sets of the results of the individual sections evaluated indicating: minimum, 1st quartile, median, 3rd quartile, maximum.
4.2. Mean Error Calculation

The results of the frame average assessment in the sections shown in column (a) in the figures (from Figures 27–47) were plotted together with error bars having a size of $\epsilon_i$, both in the positive and negative direction. $\epsilon_i$ represents the mean error of the i-th section of the video. $\epsilon_i$ is calculated using the arithmetic mean ($\epsilon_k$) where $k$ is the index of frames belonging to the i-th section. $\epsilon_k$ is the absolute value of the difference between $K_i$ and $K_k$ where $i$ is the section index, $k$ is the index of the frame being dealt with, $n$ is the number of frames in the section, $K_k$ is the weighted rating of the selected frame and $K_i$ is the weighted rating of the selected section.

$$K_i = \frac{\sum_{k=1}^{n} K_k}{n}, \quad (5)$$
$$\epsilon_k = |K_i - K_k|, \quad (6)$$
$$\epsilon_i = \frac{\sum_{k=1}^{n} \epsilon_k}{n}. \quad (7)$$

4.3. Assessment of the Best Working Method for the Test Video Scene

Since the assessment of the results in the sections using the Saaty method is not a sufficiently transparent assessment for expressing the incorrect or correct functionality of the detection algorithm, all the results of the weighted assessment are further evaluated by the so-called scaled scores.

The limit values dividing the range of the weighted quality rating value ($K$) into the following intervals are:

- Lower limits of the correct finding ($Q_1$).
- Upper limits of the incorrect finding ($Q_2$).

$$Q_1 = \sum_{k=1}^{n} \left( v_i \cdot \frac{P_{k, \text{minimum evaluated purely correct finding}}}{100} \right) = 0.0351 \cdot \frac{1}{100} + 0.1746 \cdot \frac{1}{100} + 0.4626 \cdot \frac{1}{100} + 0.1017 \cdot \frac{50}{100} + 0.2260 \cdot \frac{100}{100} = 0.3344, \quad (8)$$

$$Q_2 = \sum_{k=1}^{n} \left( v_i \cdot \frac{P_{k, \text{maximum evaluated purely correct finding}}}{100} \right) = 0.0351 \cdot \frac{100}{100} + 0.1746 \cdot \frac{100}{100} + 0.4626 \cdot \frac{100}{100} + 0.1017 \cdot \frac{50}{100} + 0.2260 \cdot \frac{0}{100} = 0.3114. \quad (9)$$

Each of the assessment intervals is assigned with a value of mark $\gamma$ (note: Greek letter gamma) used for the resulting calculation of the averaging detection quality of the results thus obtained. The result is a rigorous detection quality assessment that can provide better information than the previous weighted detection quality assessment relating to the success or failure of person detection in intelligent buildings using the above-mentioned system setting configurations applied.

$$(K \geq 0) \land (K < Q_2) \Rightarrow \gamma = -1, \quad (10)$$
$$(K \geq Q_2) \land (K < Q_1) \Rightarrow \gamma = 0, \quad (11)$$
$$(K \geq Q_1) \land (K \leq 1) \Rightarrow \gamma = 1. \quad (12)$$
Thus, the rating ($\Gamma$) of the section of the test video containing $k$-frames assessed will be as follows:

$$\Gamma = \frac{\sum_{i=1}^{k} \gamma_i}{k}. \quad (13)$$

The arithmetic mean of the person detection results of the individual frames at the configuration of the system of detecting persons in intelligent buildings. If the average rating ($\Gamma$) is positive or near 1, this is a good detection. If the average rating ($\Gamma$) is negative or close to 0, this is a poor detection.

Selected top-rated methods from the section comprising person detection results of the individual methods applied at different lifespans of person findings are selected from Tables 5 and 6 as the most successful system configuration for detecting persons. In this table, the system configurations are ranked from the top-rated method to the bottom-rated method, according to the three methods for evaluating the results of the scaled scores, designated as (A), (B) and (C). Wherein score (A) is the assessment of the ranking of the results according to the average scaled scores in all sections of the video; score (B) is the assessment of the ranking of the results according to the sum of the rankings in all sections of the video; scoring (C) is the assessment of the ranking of the results according to the ranking in all sections of the video, except for the first and last section. The ranking of the methods selected thus assessed is more representative in terms of detection quality assessment because of the omission of the results from the first and last section, in which there is an unevenly increased detection error rate in all the methods due to the inconsistent presence of persons in the frames of these sections and, therefore, there is the absence of verification patterns. The best-rated method achieves mark 1; the less successful methods receive lower marks.

**Table 5.** Table of sorting the tested system configurations according to the average scaled rating in the test video sections, from top-rated.

| System Configuration Used (Change Detection Method and Lifespan of Person Finding) | Result of the Weighted Rating—$c$ | Result of the Scaled Rating—$\Gamma$ | Ranking of the Results According to the Average Scaled Rating in all Sections of the Test Video |
|---|---|---|---|
| Method 1.1 with a lifespan of 1 frame | 0.481 | 0.650 | 9 |
| Method 1.1 with a lifespan of 12 frames | 0.487 | 0.780 | 2 |
| Method 1.1 with a lifespan of 40 frames | 0.470 | 0.775 | 4 |
| Method 1.2 with a lifespan of 1 frame | 0.473 | 0.622 | 13 |
| Method 1.2 with a lifespan of 12 frames | 0.476 | 0.762 | 6 |
| Method 1.2 with a lifespan of 40 frames | 0.460 | 0.747 | 7 |
| Method 1.3 with a lifespan of 1 frame | 0.479 | 0.643 | 10 |
| Method 1.3 with a lifespan of 12 frames | 0.485 | 0.772 | 5 |
| Method 1.3 with a lifespan of 40 frames | 0.469 | 0.776 | 3 |
| Method 2.1 with a lifespan of 1 frame | 0.400 | 0.215 | 20 |
| Method 2.1 with a lifespan of 12 frames | 0.467 | 0.574 | 15 |
| Method 2.1 with a lifespan of 40 frames | 0.438 | 0.639 | 11 |
| Method 2.2 with a lifespan of 1 frame | 0.432 | 0.349 | 18 |
| Method 2.2 with a lifespan of 12 frames | 0.468 | 0.715 | 8 |
| Method 2.2 with a lifespan of 40 frames | 0.453 | 0.789 | 1 |
| Method 2.3 with a lifespan of 1 frame | 0.412 | 0.278 | 19 |
| Method 2.3 with a lifespan of 12 frames | 0.463 | 0.623 | 12 |
| Method 2.3 with a lifespan of 40 frames | 0.424 | 0.552 | 16 |
| Method 3 with a lifespan of 1 frame | 0.328 | -0.015 | 21 |
| Method 3 with a lifespan of 12 frames | 0.437 | 0.539 | 17 |
| Method 3 with a lifespan of 40 frames | 0.405 | 0.608 | 14 |
Table 6. Table ranking the best results (A) Ranking of the results according to the average scaled rating in all sections of the video; (B) Ranking of the results according to the rankings in all sections of the video; (B) Ranking of the results according to the rankings in all sections of the video, except for the first and last section.

| System Configuration Used | (A)  | (B)  | (C)  |
|---------------------------|------|------|------|
| Method 1.1 with a lifespan of 12 frames | 2    | 1    | 1    |
| Method 1.3 with a lifespan of 12 frames | 3    | 2    | 2    |
| Method 2.2 with a lifespan of 40 frames | 1    | 4    | 3    |
| Method 1.2 with a lifespan of 12 frames | 4    | 3    | 4    |
| Method 2.3 with a lifespan of 12 frames | 6    | 5    | 5    |
| Method 2.1 with a lifespan of 40 frames | 5    | 6    | 6    |
| Method 3 with a lifespan of 40 frames | 7    | 7    | 7    |

5. Conclusions

The aim of the research was to analyze and evaluate the standard and original methods of detection of the dynamic movement of persons when a person is perceived as an object moving in the scene being captured and the solution described emphasizes the appropriateness of the method providing data for the subsequent finding of a potential person in the scene being captured. Each of the original background creation solutions analyzed and the subsequent detections have their own specifics that must be taken into account and it is necessary to proceed from the research results achieved.

Appropriate weighted criteria listed in Table 3 are an important criterion for the correct evaluation of the results depending on the system requirements. The results in the final evaluation Table 6, including the best results of each of the methods from Table 5, are ranked not only according to the weighted and scaled rating in the entire range of the test video sections checked (Table 6, columns (A), (B)), but also according to the scaled rating in the range of the video sections, except for the first and last section, which resulted in significant detection errors in the rating (Table 6, column (C)). Specific methods increase errors due to detection of redundant objects in the sections, or they did not detect persons who, according to a set of verification data based on the number and possible occurrence of persons according to Table 2, do not occur in the scene being captured.

From the results shown in Table 6 (of the top-rated results), it is obvious that the top-rated configuration is the configuration of Method 1.1 having a lifespan of 12 frames, where the best-weighted score ($K = 0.487$) was achieved compared to the worst rated method (Method 3) with the following score: $K = 0.328$. The best-scaled rating was achieved by Method 2.2 with a lifespan of 40 frames and a result of $\Gamma = 0.789$, and aforementioned Method 1.1 with a lifespan of 12 frames with a result of $\Gamma = 0.78$; the worst-scaled method is Method 3 with a lifespan of 1 frame and a result of $\Gamma = -0.015$. In the attached weighted rating graphs (Figures 27–47), precisely the influence and importance of using the appropriate value of the lifespan of persons, which neglects the momentary “loss” of the person found in the results of the detection of the current change in the image, are obvious. Method 1.1 with a lifespan of 12 frames is a method that uses the difference between the current image towards the static background created by the frame taken at the beginning of the test video. Errors confirming the theoretical functionality of this detection method were entered in the evaluation results (Figures 27–29). These errors include longer-lasting relocation of the objects in the scene, which are then mistakenly considered as a person found; this error could be resolved by occasional restarting of the detection system, thus creating a new static background.

In this best-rated configuration, the lifespan is set to 12 frames, that is, at a frame rate of 4 fps of the test video, the lifespan of the person found is 3 seconds. Such a system setup is therefore ideal for scenes being captured where people are moving rather than acquiring static conditions in the scene.

The lowest score of the scaled rating was achieved with Method 3, which is a method that uses only memoryless optical flow. According to the results in Figures 45–47, this method achieved the best detection results in section 7 shown in Table 2. This method provides information about the person
detected only in pixels that are moving between the frames tested in the scene. A significant increase in the error rate was thus achieved in the test sections where no or only slight movement, insufficient for the detection of the person, was performed in the scene. This increase was compensated by the lifespan factor of the findings that were detected correctly when static conditions were acquired in the scene.

Other methods applied mentioned in Section 2 using both static and dynamic backgrounds achieved results confirming the theoretical assumptions. An important role in their application was played by the suitably selected lifespan parameter of the persons detected. As it is obvious from Figures 27–47 with the plotted results of the individual sections assessment at the given system configurations, the average rating further evaluated with an error entered does not always correspond with the statistical distribution of the individual statistical files—the video sections. Outlying values are also included in the average in order to make the evaluation as strict as possible.

The research has proved experimentally and specified mathematically which of the methods is the most suitable and further applicable and usable in sophisticated computational applications implementable in intelligent buildings.

**Author Contributions:** Conceptualization, M.S. and Z.M.; methodology, M.S. and Z.M.; software, M.S.; validation, Z.M., R.M., J.K. and R.J.; formal analysis, M.S. and R.J.; investigation, Z.M., R.M., J.K. and R.J.; resources, R.M. and J.K.; data curation, M.S.; writing—original draft preparation, M.S. and Z.M.; writing—review and editing, M.S. and Z.M.; visualization, Z.M., R.M. and R.J.; supervision, Z.M., R.M., and R.J.; project administration, R.M. and J.K.; funding acquisition, R.M. and J.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Education of the Czech Republic, Project No. SP2020/42 and by the European Regional Development Fund in the Research Centre of Advanced Mechatronic Systems project, Project Number CZ.02.1.01/0.0/0.0/16_019/0000867.

**Acknowledgments:** This work was supported by the project SP2020/42, “Development of algorithms and systems for control, measurement and safety applications VI” of Student Grant System, VSB-TU Ostrava. This work was supported by the European Regional Development Fund in the Research Centre of Advanced Mechatronic Systems project, Project Number CZ.02.1.01/0.0/0.0/16_019/0000867 within the Operational Programme Research, Development and Education.

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

**References**

1. Luppe, C.; Shabani, A. Towards reliable intelligent occupancy detection for smart building applications. In Proceedings of the 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, Canada, 30 April–3 May 2017; pp. 1–4. [CrossRef]

2. Kahily, H.M.; Sudheer, A.P.; Narayanan, M.D. RGB-D sensor based human detection and tracking using an armed robotic system. In Proceedings of the 2014 International Conference on Advances in Electronics Computers and Communications, Bangalore, India, 10–11 October 2014; pp. 1–4. [CrossRef]

3. Ahmad, M. Human motion detection and segmentation from moving image sequences. In Proceedings of the 2008 International Conference on Electrical and Computer Engineering, Dhaka, Bangladesh, 20–22 December 2008; pp. 407–411.

4. Horn-Schunck Optical Flow Method—File Exchange—MATLAB Central. Library Catalog. Available online: www.mathworks.com (accessed on 16 June 2020).

5. a/l Kanawathi, J.; Mokri, S.S.; Ibrahim, N.; Hussain, A.; Mustafa, M.M. Motion detection using Horn Schunck algorithm and implementation. In Proceedings of the 2009 International Conference on Electrical Engineering and Informatics, Banj, Malaysia, 5–7 August 2009; pp. 83–87. [CrossRef]

6. Karch, P.; Zolotova, I. An experimental comparison of modern methods of segmentation. In Proceedings of the 2010 IEEE 8th International Symposium on Applied Machine Intelligence and Informatics (SAMI), Herlany, Slovakia, 28–30 January 2010; pp. 247–252. [CrossRef]

7. Kotyza, J.; Machacek, Z.; Koziorek, J. Detection of Directions in an Image as a Method for Circle Detection. IFAC-PapersOnLine 2018, 51, 496–501. [CrossRef]

8. Ondraczka, T.; Machacek, Z.; Koziorek, J.; Ondraczka, L. Autonomous Recognition System for Skin Pigment Evaluation. IFAC-PapersOnLine 2018, 51, 196–201. [CrossRef]
9. Chang, X.; Ma, Z.; Lin, M.; Yang, Y.; Hauptmann, A.G. Feature Interaction Augmented Sparse Learning for Fast Kinect Motion Detection. *IEEE Trans. Image Process.* 2017, 26, 3911–3920. [CrossRef]

10. Cai, G.; Wang, J.; Qian, K.; Chen, J.; Li, S.; Lee, P.S. Extremely Stretchable Strain Sensors Based on Conductive Self-Healing Dynamic Cross-Links Hydrogels for Human-Motion Detection. *Adv. Sci.* 2017, 4, 1600190. [CrossRef]

11. Yang, Z.; Pang, Y.; Han, X.L.; Yang, Y.; Ling, J.; Jian, M.; Zhang, Y.; Yang, Y.; Ren, T.L. Graphene Textile Strain Sensor with Negative Resistance Variation for Human Motion Detection. *ACS Nano* 2018, 12, 9134–9141. [CrossRef] [PubMed]

12. Minoli, D.; Sohraby, K.; Occhiogrosso, B. IoT Considerations, Requirements, and Architectures for Smart Buildings—Energy Optimization and Next-Generation Building Management Systems. *IEEE Internet Things J.* 2017, 4, 269–283. [CrossRef]

13. Lin, K.; Chen, M.; Deng, J.; Hassan, M.M.; Fortino, G. Enhanced Fingerprinting and Trajectory Prediction for IoT Localization in Smart Buildings. *IEEE Trans. Autom. Sci. Eng.* 2016, 13, 1294–1307. [CrossRef]

14. Plageras, A.P.; Psannis, K.E.; Stergiou, C.; Wang, H.; Gupta, B. Efficient IoT-based sensor BIG Data collection–processing and analysis in smart buildings. *Future Gener. Comput. Syst.* 2018, 82, 349–357. [CrossRef] [PubMed]

15. Samad, T.; Koch, E.; Stluka, P. Automated Demand Response for Smart Buildings and Microgrids: The State of the Practice and Research Challenges. *Proc. IEEE* 2016, 104, 726–744. [CrossRef]

16. Bravo, J.C.; Castilla, M.V. Geometric Objects: A Quality Index to Electromagnetic Energy Transfer Performance in Sustainable Smart Buildings. *Symmetry* 2018, 10, 676. [CrossRef]

17. Mocnej, J.; Seah, W.K.; Pekar, A.; Zolotova, I. Decentralised IoT Architecture for Efficient Resources Utilisation. *IFAC-PapersOnLine* 2018, 51, 168–173. [CrossRef]

18. Cai, X.; Wang, P.; Du, L.; Cui, Z.; Zhang, W.; Chen, J. Multi-Objective Three-Dimensional DV-Hop Localization Algorithm With NSGA-II. *IEEE Sens. J.* 2019, 19, 10003–10015. [CrossRef]

19. Cui, Z.; Sun, B.; Wang, G.; Xue, Y.; Chen, J. A Novel Oriented Cuckoo Search Algorithm to Improve DV-Hop Performance for Cyber–Physical Systems. *J. Parallel Distrib. Comput.* 2017, 103, 42–52. [CrossRef]

20. Wang, G.G.; Cai, X.; Cui, Z.; Min, G.; Chen, J. High Performance Computing for Cyber Physical Social Systems by Using Evolutionary Multi-Objective Optimization Algorithm. *IEEE Trans. Emerg. Top. Comput.* 2017. [CrossRef]

21. Gao, M.; Tian, J.; Ai, L.; Zhang, F. Sewage Image Feature Extraction and Turbidity Degree Detection Based on Embedded System. In Proceedings of the 2008 International Conference on MultiMedia and Information Technology, Yichang, China, 30–31 December 2008; pp. 357–360. [CrossRef]

22. Pang, Y.; Yuan, Y.; Li, X.; Pan, J. Efficient HOG human detection. *Signal Process.* 2011, 91, 773–781. [CrossRef]

23. Pang, Y.; Yan, H.; Yuan, Y.; Wang, K. Robust CoHOG Feature Extraction in Human-Centered Image/Video Management System. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* 2012, 42, 458–468. [CrossRef]

24. Lin, H.L.; Hsu, M.H.; Chen, W.K. Human hand gesture recognition using a convolution neural network. In Proceedings of the 2014 IEEE International Conference on Automation Science and Engineering (CASE), Taipei, Taiwan, 18–22 August 2014; pp. 1038–1043. [CrossRef]

25. Haddadnia, J.; Ahmadi, M. N-feature neural network human face recognition. *Image Vis. Comput.* 2004, 22, 1071–1082. [CrossRef]

26. Yoo, J.H.; Hwang, D.; Moon, K.Y.; Nixon, M.S. Automated Human Recognition by Gait using Neural Network. In Proceedings of the 2008 First Workshops on Image Processing Theory, Tools and Applications, Sousse, Tunisia, 23–26 November 2008; pp. 1–6. [CrossRef]

27. Robinson, R.M.; Lee, H.; McCourt, M.J.; Marathe, A.R.; Kwon, H.; Ton, C.; Nothwang, W.D. Human-autonomy sensor fusion for rapid object detection. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 305–312. [CrossRef]

28. Setjo, C.H.; Achmad, B.; Faridah. Thermal image human detection using Haar-cascade classifier. In Proceedings of the 2017 7th International Annual Engineering Seminar (InAES), Yogyakarta, Indonesia, 1–2 August 2017; pp. 1–6. [CrossRef]

29. A Motion Detection Approach based on UAV Image Sequence. *KSII Trans. Internet Inf. Syst.* 2018, 12. [CrossRef]
30. Yoon, M.-K.; Mohan, S.; Choi, J.; Kim, J.-E.; Sha, L. SecureCore: A multicore-based intrusion detection architecture for real-time embedded systems. In Proceedings of the 2013 IEEE 19th Real-Time and Embedded Technology and Applications Symposium (RTAS), Philadelphia, PA, USA, 9–11 April 2013; pp. 21–32. [CrossRef]

31. Rahmatian, M.; Kooti, H.; Harris, I.G.; Bozorgzadeh, E. Hardware-Assisted Detection of Malicious Software in Embedded Systems. IEEE Embed. Syst. Lett. 2012, 4, 94–97. [CrossRef]

32. Paral, P.; Chatterjee, A.; Rakshit, A. Vision Sensor-Based Shoe Detection for Human Tracking in a Human–Robot Coexisting Environment: A Photometric Invariant Approach Using DBSCAN Algorithm. IEEE Sens. J. 2019, 19, 4549–4559. [CrossRef]

33. Yin, C.; Chiu, C.F.; Hsieh, C.C. A 0.5 V, 14.28-kframes/s, 96.7-dB Smart Image Sensor With Array-Level Image Signal Processing for IoT Applications. IEEE Trans. Electron. Devices 2016, 63, 1134–1140. [CrossRef]

34. Maryam Noorbakhsh, S.; Habibi, M. Design of a low power multiple object tracking vision sensor for use in robot localization and surveillance tasks. In Proceedings of the 2016 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS), Tehran, Iran, 14–15 December 2016; pp. 1–5. [CrossRef]

35. Tanaka, J.; Imamoto, H.; Seki, T.; Oba, M. Low power wireless human detector utilizing thermopile infrared array sensor. In Proceedings of the IEEE SENSORS 2014 Proceedings, Valencia, Spain, 2–5 November 2014; pp. 462–465. [CrossRef]

36. Lan, J.; Li, J.; Xiang, Y.; Huang, T.; Yin, Y.; Yang, J. Fast Automatic Target Detection System Based on a Visible Image Sensor and Ripple Algorithm. IEEE Sens. J. 2013, 13, 2720–2728. [CrossRef]

37. Young, C.; Omid-Zohoor, A.; Lajevardi, P.; Murmann, B. A Data-Compressive 1.5/2.75-bit Log-Gradient QVGA Image Sensor With Multi-Scale Readout for Always-On Object Detection. IEEE J. Solid-State Circuits 2019, 54, 2932–2946. [CrossRef]

38. Rusci, M.; Rossi, D.; Lecca, M.; Gottardi, M.; Farella, E.; Benini, L. An Event-Driven Ultra-Low-Power Smart Visual Sensor. IEEE Sens. J. 2016, 16, 5344–5353. [CrossRef]

39. Zou, Y.; Gottardi, M.; Perenzoni, D.; Perenzoni, M.; Stoppa, D. A 1.6 mW 320 × 240-pixel vision sensor with programmable dynamic background rejection and motion detection. In Proceedings of the 2017 IEEE Sensors, Glasgow, UK, 29 October–1 November 2017; pp. 1–3. [CrossRef]

40. Carey, S.J.; Barr, D.R.W.; Dudek, P. Demonstration of a low power image processing system using a SCAMP3 vision chip. In Proceedings of the 2011 Fifth ACM/IEEE International Conference on Distributed Smart Cameras, Ghent, Belgium, 22–25 August 2011; pp. 1–2. [CrossRef]

41. Habibi, M. A low power smart CMOS image sensor for surveillance applications. In Proceedings of the 2010 6th Iranian Conference on Machine Vision and Image Processing, Isfahan, Iran, 27–28 October 2010; pp. 1–4. [CrossRef]

42. Lacassagne, L.; Manzanera, A.; Dupret, A. Motion detection: Fast and robust algorithms for embedded systems. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 3265–3268. [CrossRef]

43. Fernandez-Berni, J.; Carmona-Galan, R.; Rodriguez-Vazquez, A. Ultralow-Power Processing Array for Image Enhancement and Edge Detection. IEEE Trans. Circuits Syst. II Express Briefs 2012, 59, 751–755. [CrossRef]

44. Wang, Z.; Yang, R.; Wang, L.; Green, R.C.; Dounis, A.I. A fuzzy adaptive comfort temperature model with grey predictor for multi-agent control system of smart building. In Proceedings of the 2011 IEEE Congress of Evolutionary Computation (CEC), New Orleans, LA, USA, 5–8 June 2011; pp. 728–735. [CrossRef]

45. Ock, J.; Issa, R.R.A.; Flood, I. Smart Building Energy Management Systems (BEMS) simulation conceptual framework. In Proceedings of the 2016 Winter Simulation Conference (WSC), Washington, DC, USA, 11–14 December 2016; pp. 3237–3245. [CrossRef]

46. O’Keeffe, J.T. System and Method for User Profile Enabled Smart Building Control. U.S. Patent US20170055126A1, 23 February 2017.

47. Mituyosi, T.; Yagi, Y.; Yachida, M. Real-time human feature acquisition and human tracking by omnidirectional image sensor. In Proceedings of the Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, MFI2003, Tokyo, Japan, 1 August 2003; pp. 258–263. [CrossRef]

48. Sanoj, C.S.; Vijayaraj, N.; Rajalakshmi, D. Vision approach of human detection and tracking using focus tracing analysis. In Proceedings of the 2013 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, India, 21–22 February 2013; pp. 64–68. [CrossRef]
49. White, I.; Borah, D.K.; Tang, W. Robust Optical Spatial Localization Using a Single Image Sensor. *IEEE Sens. Lett.* 2019, 3, 1–4. [CrossRef]

50. Alhilal, M.S.; Soudani, A.; Al-Dhelaan, A. Low power scheme for image based object identification in wireless multimedia sensor networks. In Proceedings of the 2014 International Conference on Multimedia Computing and Systems (ICMCS), Marrakech, Morocco, 14–16 April 2014; pp. 927–932. [CrossRef]

51. Uhm, T.; Park, H.; Seo, D.; Park, J.I. Human-of-interest tracking by integrating two heterogeneous vision sensors. In Proceedings of the 2010 IEEE Virtual Reality Conference (VR), Boston, MA, USA, 20–24 March 2010; pp. 309–310. [CrossRef]

52. Shetty, A.D.; Shubha, B.; Suryanarayana, K. Detection and tracking of a human using the infrared thermopile array sensor—“Grid-EYE”. In Proceedings of the 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, India, 6–7 July 2017; pp. 1490–1495. [CrossRef]

53. Zou, Y.; Gottardi, M.; Perenzoni, D.; Perenzoni, M.; Stoppa, D. Live demonstration: Motion detection vision sensor with dynamic background rejection. In Proceedings of the 2017 IEEE Sensors, Glasgow, UK, 29 October–1 November 2017. [CrossRef]

54. Ondraczka, T.; Machacek, Z.; Kubicek, J.; Bryjova, I.; Cerny, M.; Koziorek, J.; Prochazka, V. Automated image detection of the main bloodstream stem from CT angiography image. *IFAC-PapersOnLine* 2016, 49, 109–114. [CrossRef]

55. Hercik, R.; Slaby, R.; Machacek, Z. Image processing for low-power microcontroller application. *Prz. Elektrotech.* 2012, 88, 343–346.

56. Miaou, S.G.; Sung, P.-H.; Huang, C.-Y. A Customized Human Fall Detection System Using Omni-Camera Images and Personal Information. In Proceedings of the 1st Transdisciplinary Conference on Distributed and Home Healthcare (D2H2), Virginia, VA, USA, 2–4 April 2006; pp. 39–42. [CrossRef]

57. Ong, K.S.; Hsu, Y.H.; Fu, L.C. Sensor fusion based human detection and tracking system for human-robot interaction. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura-Algarve, Portugal, 7–12 October 2012; pp. 4835–4840. [CrossRef]

58. Chang, Y.T.; Shih, T.K. Human fall detection based on event pattern matching with ultrasonic array sensors. In Proceedings of the 2017 10th International Conference on Ubi-media Computing and Workshops (Ubi-Media), Pattaya, Thailand, 1–4 August 2017; pp. 1–4. [CrossRef]

59. Nakamura, R.; Hadama, H. Target localization using multi-static UWB sensor for indoor monitoring system. In Proceedings of the 2017 IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet), Phoenix, AZ, USA, 15–18 January 2017; pp. 37–40. [CrossRef]

60. Verdant, A.; Dupret, A.; Mathias, H.; Villard, P.; Lacassagne, L. Adaptive Multiresolution for Low Power CMOS Image Sensor. In Proceedings of the 2007 IEEE International Conference on Image Processing, San Antonio, TX, USA, 16 September–19 October 2007; pp. V-185–V-188. [CrossRef]

61. Klein, J.; Lacassagne, L.; Mathias, H.; Moutault, S.; Dupret, A. Low Power Image Processing: Analog Versus Digital Comparison. In Proceedings of the Seventh International Workshop on Computer Architecture for Machine Perception (CAMP’05), Palermo, Italy, 4–6 July 2005; pp. 111–115. [CrossRef]

62. Choi, J. Review of low power image sensors for always-on imaging. In Proceedings of the 2016 International SoC Design Conference (ISOC), Jeju, Korea, 23–26 October 2016; pp. 11–12. [CrossRef]

63. Ait-Kaci, F.; Mathias, H.; Zhang, M.; Dupret, A. Exploration of analog pre-processing architectures for coarse low power motion detection. In Proceedings of the 2011 IEEE 9th International New Circuits and systems conference, Bordeaux, France, 26–29 June 2011; pp. 45–48. [CrossRef]

64. Vornicu, I.; Goraş, L. Image processing using a CMOS analog parallel architecture. In Proceedings of the CAS 2010 Proceedings (International Semiconductor Conference), Sinaia, Romania, 11–13 October 2010; pp. 461–464. [CrossRef]

65. Kim, Y.S.; Park, J.I.; Lee, D.J.; Chun, M.G. Real time detection of moving human based on digital image processing. In Proceedings of the SICE Annual Conference 2007, Takamatsu, Japan, 17–20 September 2007; pp. 2030–2033. [CrossRef]

66. Yang, B.; Yang, M.; Plaza, A.; Gao, L.; Zhang, B. Dual-Mode FPGA Implementation of Target and Anomaly Detection Algorithms for Real-Time Hyperspectral Imaging. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2015, 8, 2950–2961. [CrossRef]

67. Basilio, J.A.M.; Torres, G.A.; Perez, G.S.; Medina, L.K.T.; Meana, H.M.P. Explicit Image Detection Using YCbCr Space Color Model as Skin Detection. *Appl. Math. Comput. Eng.* 2011, 123–128.
68. Jayaraman, S.; Esakkirajan, S.; Veerakumar, T. *Digital Image Processing*; Tata McGraw-Hill Education: New Delhi, India; London, UK, 2011.

69. Vollmer, M.; Möllmann, K.P. *Infrared Thermal Imaging: Fundamentals, Research and Applications*; Wiley: Hoboken, NJ, USA, 2011.

70. Osman, M.Z.; Maarof, M.A.; Rohani, M.F. Improved Dynamic Threshold Method for Skin Colour Detection Using Multi-Colour Space. *Am. J. Appl. Sci.* **2016**, *13*, 135–144. [CrossRef]

71. Jones, M.J.; Rehg, J.M. Statistical Color Models with Application to Skin Detection. *Int. J. Comput. Vis.* **2002**, *46*, 81–96. [CrossRef]

72. Mate, S.; Bouchet, L.; Atteia, J.L.; Claret, A.; Cordier, B.; Dagoneau, N.; Godet, O.; Gros, A.; Schanne, S.; Triou, H. Simulations of the SVOM/ECLAIRs dynamic background: A fast, accurate and general approach for wide-field hard X-ray instruments. *Exp. Astron.* **2019**, *48*, 171–198. [CrossRef]

73. Duncan, J.; Chou, T.C. On the detection of motion and the computation of optical flow. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 346–352. [CrossRef]

74. Damarla, T.; Kaplan, L.; Chan, A. Human infrastructure & human activity detection. In Proceedings of the 2007 10th International Conference on Information Fusion, Quebec City, QC, Canada, 9–12 July 2007; pp. 1–8. [CrossRef]