Appl. Sci. **2014**, *4*, 331-350; doi:10.3390/app4030331

OPEN ACCESS

applied sciences

ISSN 2076-3417

www.mdpi.com/journal/applsci

Article

Selection of a Visible-Light *vs.* Thermal Infrared Sensor in Dynamic Environments Based on Confidence Measures

Juan Serrano-Cuerda [1], Antonio Fernández-Caballero [1,2,*] and María T. López [1,2]

[1] Instituto de Investigación en Informática de Albacete, Albacete 02071, Spain;
E-Mails: jserranocuerda@gmail.com (J.S.-C.); Maria.LBonal@uclm.es (M.T.L.)

[2] Departamento de Sistemas Informáticos, Universidad de Castilla-La Mancha, Albacete 02071, Spain

* Author to whom correspondence should be addressed; E-Mail: Antonio.Fdez@uclm.es;
Tel.: +34-967-599-200; Fax: +34-967-599-224.

Received: 16 June 2014; in revised form: 26 July 2014 / Accepted: 4 August 2014 /
Published: 8 August 2014

Abstract: This paper introduces a confidence measure scheme in a bimodal camera setup for automatically selecting visible-light or a thermal infrared in response to natural environmental changes. The purpose of the setup is to robustly detect people in dynamic outdoor scenarios under very different conditions. For this purpose, two efficient segmentation algorithms, one dedicated to the visible-light spectrum and another one to the thermal infrared spectrum, are implemented. The segmentation algorithms are applied to five different video sequences recorded under very different environmental conditions. The results of the segmentation in both spectra allow one to establish the best-suited confidence interval thresholds and to validate the overall approach. Indeed, the confidence measures take linguistic values LOW, $MEDIUM$ and $HIGH$, depending on the reliability of the results obtained in visible-light, as well as in thermal infrared video.

Keywords: visible-light sensor; thermal infrared sensor; people detection; confidence measures

1. Introduction

Visual monitoring, including people detection, tracking, recognition and activity interpretation [1], is a key component of intelligent video surveillance systems [2,3]. The contribution of a camera to the observation of a scene depends on its viewpoint and on the scene configuration. This is a dynamic property, as the scene content is subject to change over time. There are two major types of security cameras when setting up a security system: visual-light (or color) and thermal infrared sensors. Recently, the robustness of a new thermal infrared pedestrian detection system has been investigated under different outdoor environmental conditions [4]. This paper offers a step forward towards enhancing the performance of visual monitoring systems by adding a visual-light sensor.

After studying different visual sensors, it is commonly concluded that the advantages and disadvantages are complementary in using visual-light and thermal infrared spectra. On the one hand, although the information obtained from an infrared camera is useful for detecting humans in nocturnal environments, it presents severe problems in other environments [5]. This is the case for hot or thermally homogeneous environments. Moreover, visible-light yields good results when conducting human detection in well-lit environments, but it is problematic in dark environments or in areas of the scene that present shadows or have low visibility in general. In order to enhance the performance of people monitoring, some researchers are performing image fusion by using visible and infrared images together [6]. However, the fusion of thermal infrared and visible images is not trivial [7]. This is why there has been a growing interest in visual surveillance using multimodal sensors, such as thermal and visible cameras, in both civilian and military applications [8].

This paper introduces a proposal based on confidence measures in a bimodal visual sensor setup for automatically selecting visible-light or thermal infrared in response to natural environmental changes. The purpose of the setup is to robustly detect people in dynamic outdoor ambiences. The rest of the article is organized as follows. Section 2 describes the work related to robust people detection in visible-light and thermal infrared video. Then, Section 3 introduces our proposal based on confidence measures to automatically select between visible-light and thermal infrared sensor. After that, two efficient segmentation algorithms, one for the visible-light spectrum and another one for the thermal infrared spectrum, are introduced in Section 4. In Section 5, the segmentation algorithms are applied to five different video sequences recorded under very different environmental conditions in a bimodal sensor setup. The results of the segmentation in both spectra allow one to establish the most suited confidence thresholds and values. Finally, some conclusions are provided in Section 6.

2. Related Work

To date, a widespread approach for detecting pedestrians is the single use of grey scale [9] and color information [10–12]. In the approach in [13], the scene background is dynamically adapted by filtering elements, such as shadows, specular reflections, *etc.*, that appear after usual background subtraction algorithms. Another proposal uses histograms of oriented gradients to perform an initial step in human detection [14]. These detections are classified by means of support vector machines (SVMs). The most interesting contribution is the description of opponent color space as a biologically-inspired alternative to detect humans. The authors demonstrate that their method achieves superior results compared to the equivalent segmentation using the RGB color space. However, using visible-light information is usually problematic when facing changes in lighting in a scene or when there are illumination problems in a scene, such as fog or zones covered by darkness.

Images in the thermal infrared spectrum show a set of differentiating features compared with visible spectrum camera frames [15–17]. The grey level value of objects in infrared stems from their temperature and radiated heat and does not depend on the illumination. When adopting a people detection algorithm in the thermal infrared spectrum, heat is taken into account, as people appear warmer than other elements in the scenario. Nonetheless, this is not always true [5]. This is usually faithfully fulfilled in winter and at night. In addition, due to the technological limitations of infrared cameras, many infrared images have low spatial resolution and lower sensitivity than images of the visible spectrum. The limitations usually result in a large number of image noise and low image quality.

Many approaches combine the properties of appearance and shape in this spectrum. Indeed, humans are initially segmented from their appearance (they generally look brighter than other objects in the scene) and filtered and sorted on a shape basis. For instance, in [16], human candidates are initially detected by a thresholding procedure, using the previously described idea. The candidates found are decomposed into a series of layers using wavelets. Their features are then extracted through the use of high-frequency bands. Finally, human regions are classified by an SVM. Even so, there remains some problems to solve at the time of performing an accurate segmentation of humans, such as, for instance, the appearance of halos that decrease accuracy when defining the outlines of the silhouettes or problems that arise in situations where the temperature of objects and persons present in the scene are quite homogeneous. In [18], these problems are addressed using a histogram-based segmentation, where vertical and horizontal adjustments are realized in human candidate regions. Two different alternatives are established for winter and summer conditions. While grey value levels are used in the former to delimit humans, intensity changes in the human limits are searched for in the latter.

In order to take advantage of the strengths of both visible-light and thermal infrared sensors, some image fusion techniques are arising. In this sense, a background-subtraction technique that fuses contours from thermal and visible imagery for persistent object detection in urban settings is presented [19]. Statistical background subtraction in the thermal domain is used to identify the initial regions-of-interest. Color and intensity information are used within these areas to obtain the corresponding regions-of-interest in the visible domain. The objective of another work is to authenticate individuals based on the appearance of their faces [20]. The main novelty is an algorithm for the decision-level fusion of two types of imagery: one acquired in the visual and one acquired in the infrared electromagnetic spectrum. Very recently, a new method based on a nonsubsampled contourlet transform, has been proposed to fuse the infrared image and the visible light image [21]. Furthermore, an adaptively weighted infrared and visual image fusion algorithm has been developed based on the multiscale top-hat selection transform [22].

Some other proposals do not use image fusion, but use multimodal camera setups. The problem turns now into a camera selection process, as recently described in [23]. The advantages of jointly using a thermal camera and a visible camera without fusion have been studied and discussed extensively in a few works, such as [24]. The two main benefits of the joint use of thermal and visible sensors are first the complementary nature of different modalities that provides the thermal and visible-light information of the scene and, second, the redundancy of information captured by the different modalities, which increases the reliability and robustness of a surveillance system. Moreover, the joint use of multiple imaging modalities is one means of improving some of the quality measures of the input data, in hopes of ultimately improving overall system performance [24]. Furthermore, a selection method based on a partially observable Markov decision process model has been introduced [8]. Additionally, an innovative evaluation function identifies the most informative of several multi-view video streams by extracting and scoring features related to global motion, the attributes of moving objects and special events, such as the appearance of new objects. In another paper, also a multi-camera video surveillance system with automatic camera selection is presented [25]. A confidence measure, quality-of-view, is defined to automatically evaluate the camera's view performance for each time instant.

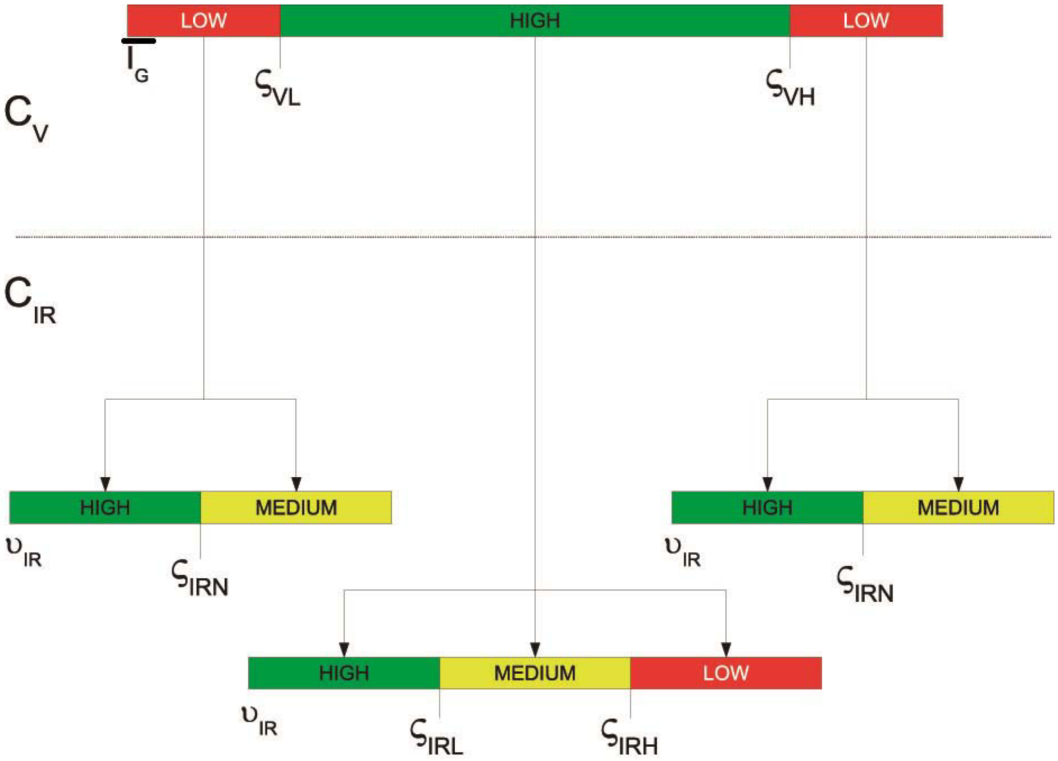
By leveraging the relative strengths of reflective and emissive modalities, bimodal sensor systems are capable of operating throughout any known environmental conditions.

3. Visible-Light and Thermal Infrared Confidence Measures

Our aim is to detect people robustly under dynamic circumstances through automatically choosing between a visible-light and a thermal infrared sensor in a bimodal camera setup. To accomplish this purpose, we introduce a series of so-called confidence measures. The confidence measures are set for each spectrum based on some relevant features of the frames acquired by each kind of camera. The frame features used are the mean illumination and the standard image deviation in the thermal infrared spectrum, while the average grey level value is the main cue in the visible-light spectrum.

A scheme for how the confidence measures are set up in both spectra is shown in Figure 1. The top of Figure 1 shows a division into three different zones (intervals) separated by two thresholds, ζ_{VL} and ζ_{VH}, which establish the confidence measures in the visible-light spectrum. C_V is the confidence value in the visible spectrum and takes the linguistic values $HIGH$ and LOW. Notice that a $MEDIUM$ confidence value is not used in the visible-light spectrum. $\overline{I_G}$ is the average grey level value of the input frame in the visible spectrum.

Figure 1. Set up of the confidence measures.



The bottom of Figure 1 offers three possibilities for assigning confidence intervals in the thermal infrared spectrum. Notice that the confidence value in the thermal infrared, C_{IR}, is directly affected by the confidence value established in the visible-light spectrum. Firstly, for $C_V = LOW$ (there is a low reliability on the results obtained by the visible camera), C_{IR} takes the two different values, $HIGH$ and $MEDIUM$. The reason is the following. When the confidence value of the visible-light spectrum is set to LOW, it is clear that the visible-light sensor is almost unable to distinguish humans. This is why the infrared spectrum is always forced with a confidence value above LOW. Indeed, under this circumstance, the infrared sensor always works better than the visible-light one. On the other hand, when $C_V = HIGH$, C_{IR} takes the three possible values, $HIGH$, $MEDIUM$ and LOW. This means that, when the detection in visible-light offers good results, the detection in thermal infrared can be very worse or worse, but sometimes equal or even better.

Next, the setup of the confidence measures is explained in more detail.

3.1. Visible-Light Confidence Measures

The average or mean grey level value of visible-light image I_V is the basis for establishing the confidence intervals in the visible-light spectrum. In order to improve the invariability from the camera settings, the color image I_V is transformed into a grey level image I_G.

Let us firstly consider the left confidence interval at the top of Figure 1. A low average of I_G means that the image is captured under poor lighting conditions, making any object (including humans) hard to distinguish from the rest of the scene, just as shown in Figure 2a. The thermal infrared equivalent of this frame is shown in Figure 3b. On the other hand, consider the right confidence interval at the top of Figure 1. A very high grey level mean value denotes that it is snowing or the environment is under fog conditions (as depicted in Figure 2c). Regardless, we can be in a situation where a lighting source directly pointing toward the camera is blinding it. Therefore, the visible camera is unable to distinguish anything in the scene. Lastly, an intermediate grey level mean indicates that the lighting conditions in the scene are adequate and that humans are easily distinguished. This is depicted as the central confidence interval at the top of Figure 1. An example of this situation is shown in Figure 2b.

Figure 2. Different confidence values for the visible spectrum. (**a**) Visible-light image with a low confidence value in night conditions; (**b**) visible-light image with a high confidence value; (**c**) visible-light image with a low confidence value in fog conditions.

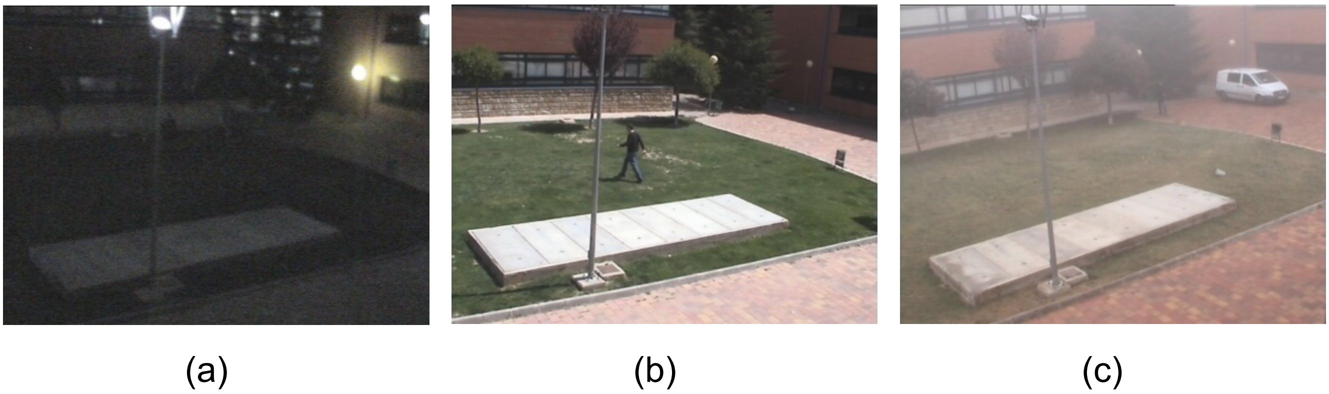
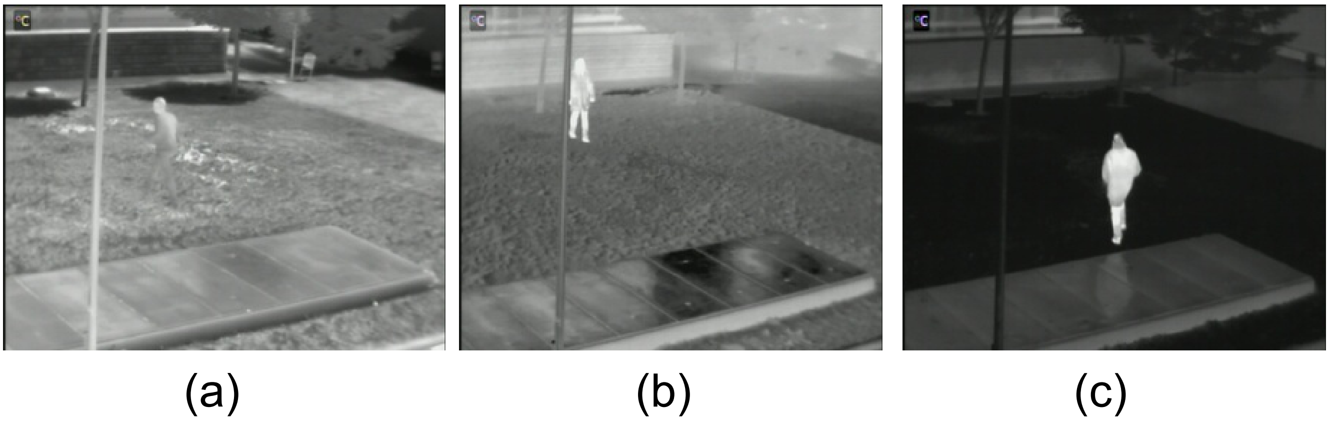


Figure 3. Different confidence values for the thermal infrared spectrum. (**a**) Thermal infrared image with a low confidence value; (**b**) thermal infrared image with a medium confidence value; (**c**) thermal infrared image with a high confidence value.



Equation (1) shows how the reliability is established in the visible spectrum (denoted as C_V), with thresholds ζ_{VL} and ζ_{VH} fixed experimentally, since the grey level values of the elements in the scene determine the conditions where the visual spectrum is trustworthy and also where it is not reliable.

$$C_V = \begin{cases} \text{HIGH}, & \text{if } \zeta_{VL} < \overline{I_G} < \zeta_{VH} \\ \\ \text{LOW}, & \text{otherwise} \end{cases} \tag{1}$$

3.2. Thermal Infrared Confidence Measures

The illumination mean and the standard deviation provide very useful information about the contrast of an image in the thermal infrared spectrum. Since infrared sensors are designed to distinguish humans from the background, supposing that they usually appear warmer, this information is key to establishing the sensors' reliability in environmental conditions that satisfy this condition, *i.e.*, in sequences where the human's temperature is higher than the environment's.

Let us define the contrast υ_{IR} as the coefficient between the average grey level value $\overline{I_{IR}}$ of infrared image I_{IR} and its standard deviation $\sigma_{I_{IR}}$, just as shown in Equation (2).

$$\upsilon_{IR} = \frac{\overline{I_{IR}}}{\sigma_{I_{IR}}} \tag{2}$$

An image with a high grey level mean and a low standard deviation denotes that a large number of pixels have similar values, making humans hard to distinguish from the background. An example of a frame with these features is shown in Figure 3a (with the equivalent visible-light frame shown in Figure 2b). On the other hand, an image with a great standard deviation and a low mean value has a small number of pixels with high grey level values and the rest of them with low values. The high value pixels usually correspond to humans. An example of a frame with this situation is shown in Figure 3c. The intermediate case where humans are distinguished from the background not as clearly as in the previous situation can be appreciated in Figure 3b.

As mentioned before, we are in front of two possibilities when considering the confidence intervals in the thermal infrared spectrum. Firstly, we consider the case when the confidence in the visible-light spectrum is $HIGH$. Here, the equation for establishing the reliability of the frames in the infrared spectrum (denoted as C_{IR}) is shown in Equation (3), where thresholds ζ_{IRH} and ζ_{IRL} are experimentally established, since they are dependent on the particular heat distribution of the test scenario.

$$C_{IR} = \begin{cases} \text{HIGH}, & \text{if } (C_V = HIGH \text{ AND } \upsilon_{IR} < \zeta_{IRL}) \\ \\ \text{MEDIUM}, & \text{if } (C_V = HIGH \text{ AND } \zeta_{IRL} < \upsilon_{IR} < \zeta_{IRH}) \\ \\ \text{LOW}, & \text{if } (C_V = HIGH \text{ AND } \upsilon_{IR} > \zeta_{IRH}) \end{cases} \tag{3}$$

As previously mentioned, the thermal infrared confidence measures take on great importance when the confidence value in the visible-light spectrum is LOW. An example of this situation is seen in Figure 2a for the visible-light camera and in Figure 3b for the infrared camera, where it can be appreciated that the human is still hard to distinguish, but more easily than in the visible spectrum. Thus, the thermal infrared confidence values at night (and also under bad atmospheric conditions) are restricted to $MEDIUM$ and $HIGH$, with a new threshold ζ_{IRN} used to separate both values, as shown in Equation (4).

$$C_{IR} = \begin{cases} \text{HIGH}, & \text{if } (C_V = LOW \text{ AND } \upsilon_{IR} < \zeta_{IRN}) \\ \\ \text{MEDIUM}, & \text{if } (C_V = LOW \text{ AND } \upsilon_{IR} > \zeta_{IRN}) \end{cases} \tag{4}$$

4. People Segmentation

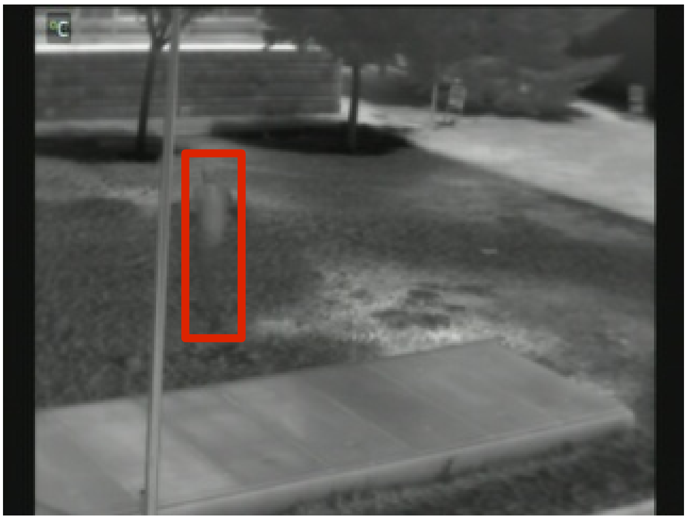
The main features of each spectrum, exploiting their properties, are used to develop robust human segmentation algorithms. Thus, the thermal difference between the humans and their environment is a cue used in the thermal infrared spectrum, and the information provided by the color in the scene is exploited in the visible-light spectrum. Next, two segmentation algorithms are described. The first one, designed for thermal infrared video, is based on frame subtraction, whereas the second one, to be used in the visible-light spectrum, is based on background subtraction.

4.1. People Detection in Thermal Infrared Based on Frame Subtraction

We implemented a single-frame-based human detection system similar to what was described in [26]. First, a set of human candidates are extracted from the scene, using the thermal information contained in the frame. The size and location of these initial candidates is then refined. Lastly, potential false positives that may have appeared in the algorithm are eliminated.

However, there are some environmental conditions that may adversely affect the visual contrast in the infrared spectrum. For instance, people are very difficult to detect in hot environments, where the ambient temperature is similar to the human one. An illustration of this fact is shown in Figure 4. The human being that is hard to see stands out manually. However, if the motion information is used from the video, humans are easily detected, since they do not remain still over long periods of time. Therefore, an extension for the human detection based on a single frame is developed using the motion information in the scene.

Figure 4. An example of a human being hard to detect in the thermal infrared spectrum.



While the list L_{SF} of humans obtained from detection based on a single frame is used, information from two new stages is added to the previous list. A new phase, called frame subtraction analysis, is introduced in this extension in order to take advantage of the motion information in the scene. The results L_{MOV} from this new stage are later refined into a new list L_S, which will be joined with the list L_{SF} in order to reduce the number of false negatives in the scene.

4.1.1. Frame Subtraction Analysis

The previous image $I_{IR}(x, y, t - \Delta t)$ and the current one $I_{IR}(x, y, t)$ are used. Image subtraction and thresholding are performed on these frames, as shown in Equation (6), where θ_{sub} is experimentally fixed as 16% of the maximum value of a 256 grey level image. This binarized image is combined with the image I_c (obtained during the stage based on a single frame after an initial thresholding and morphological filter of I_{IR}) by an "AND" operation, obtaining binary image I_{sc}. This way, false positives due to small illumination changes are discarded, by ensuring that the zones with motion have also warm heat concentrations similar to humans. Regions-of-interest (ROIs) with an area superior to A_{min} (calculated as shown in Equation (5)) and with a percentage of pixels set to MAX greater than a rate threshold ψ (experimentally fixed at a 5% of the area of the ROI) are extracted from I_{sc} in the list of blobs L_{MOV}. Notice that A_{I_c} is the area of image I_c.

$$A_{min} = \frac{A_{I_c}}{400} \tag{5}$$

$$I_s(x, y) = \begin{cases} \max, & \text{if } |I_{IR}(x, y, t) - I_{IR}(x, y, t - \Delta t)| > \theta_{sub} \\ \min, & \text{otherwise} \end{cases} \tag{6}$$

4.1.2. Refinement of Human Candidate Blobs

In the next stage, ROIs obtained from the blobs in L_{MOV} are vertically and horizontally delimited in the same way as the ROIs are refined in the human detection based on a single frame. Human candidates are then enlisted in a list L_s. This list is finally checked along with L_c (obtained from human detection based on a single frame) to remove redundancies encountered in both lists. This way, humans that can only be found through motion information are added to the initial algorithm. These humans are enlisted into the final list L_{SUB}.

4.2. People Detection in Visible-Light Based on Background Subtraction

A human detection system using a classic background subtraction approach is also implemented and tested, adapted from [27]. After performing an adaptive Gaussian background subtraction, the resulting image is binarized and filtered with a series of morphological operations. Finally, blobs with human shapes are extracted from the binarized image.

4.2.1. Gaussian Background Subtraction

An adaptive Gaussian background subtraction is performed on input image I_V obtained from the visible-light camera, as shown in Figure 5a. The subtraction is based on a well-known algorithm [28]. The algorithm builds an adaptive model of the scene background based on the probabilities of a pixel having a given color level. The authors begin their estimation of the Gaussian mixture model of the background by expected sufficient statistics update equations with a learning rate ϱ, then switch to L-recent window version when the first L samples are processed. The expected sufficient statistics update equations provide a good estimate at the beginning before all N samples can be collected. This initial estimate improves the accuracy of the estimate and also the performance of the tracker, allowing fast convergence on a stable background model. The N-recent window update equations give priority over recent data; therefore, the tracker can adapt to changes in the environment.

Shadow removal is performed by the comparison of a non-background pixel against the current background components. If the difference in both chromatic and brightness components are within some thresholds, the pixel is considered as a shadow. With this objective, an effective computational color model similar to the one proposed by [29] is used. An example of the background model is shown in Figure 5b. A shadow detection algorithm, based on the computational color space used in the background model, is also used. After the background segmentation is performed, an initial background segmentation image (I_B is obtained, as shown in Figure 5c).

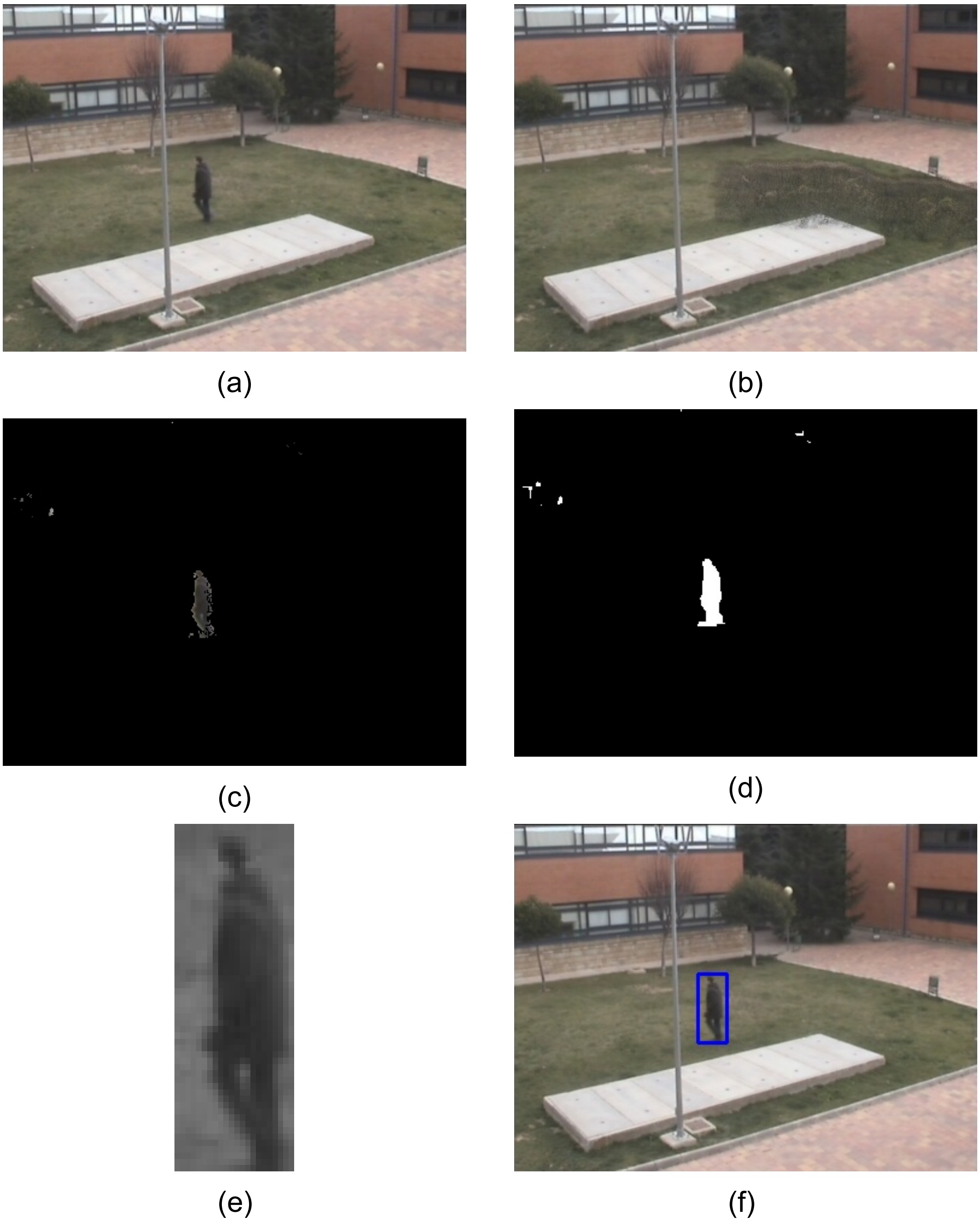
4.2.2. Removal of Image Noise

However, the resulting image contains some noise, which must be eliminated. Thus, an initial threshold θ_0 is applied, as shown in Equation (7), where min is fixed to zero (since we are obtaining binary images) and max is the maximum grey level value that a pixel can have in I_B (e.g., 255 for an eight-bit image). The value of this threshold will be experimentally fixed according to the features of the image. The result is shown in Figure 5d.

$$I_{Th}(x,y) = \begin{cases} \text{min}, & \text{if } I_B(x,y) \leq \theta_0 \\ \text{max}, & \text{otherwise} \end{cases} \tag{7}$$

After this operation, two morphological operations, namely opening and closing, are performed to eliminate the remaining noise of the image, obtaining I_S.

Figure 5. Stages of the background segmentation algorithm. (**a**) Original input frame; (**b**) background model calculated; (**c**) foreground image calculated; (**d**) foreground image after binarization; (**e**) ROI extracted from the original image; (**f**) final result.



4.2.3. Detection of Human Candidate Blobs

Now, human candidates must be extracted from I_S. Blobs with an area A_R (see Equation (10)) lower than an area A_{minBS} are discarded, while a series of restrictions similar to the human detection based on a single frame in the thermal infrared spectrum are imposed on the remaining ones, establishing an ROI for each blob detected. Let us remind ourselves that an ROI is defined by its coordinates (x_{start}, y_{start}) and (x_{end}, y_{end}). Since background subtraction usually extracts the humans in their entirety (like the human detection based on a single frame), similar restrictions are used, using height/width ratios, as shown in

Equation (11). An ROI that satisfies the criteria is shown in Figure 5e, while the final result is shown in Figure 5f.

$$h_R = x_{end} - x_{start} \tag{8}$$

$$w_R = y_{end} - y_{start} \tag{9}$$

$$A_R = h_R \times w_R \tag{10}$$

$$hwR = \frac{h_R}{w_R} \tag{11}$$

5. Data and Results

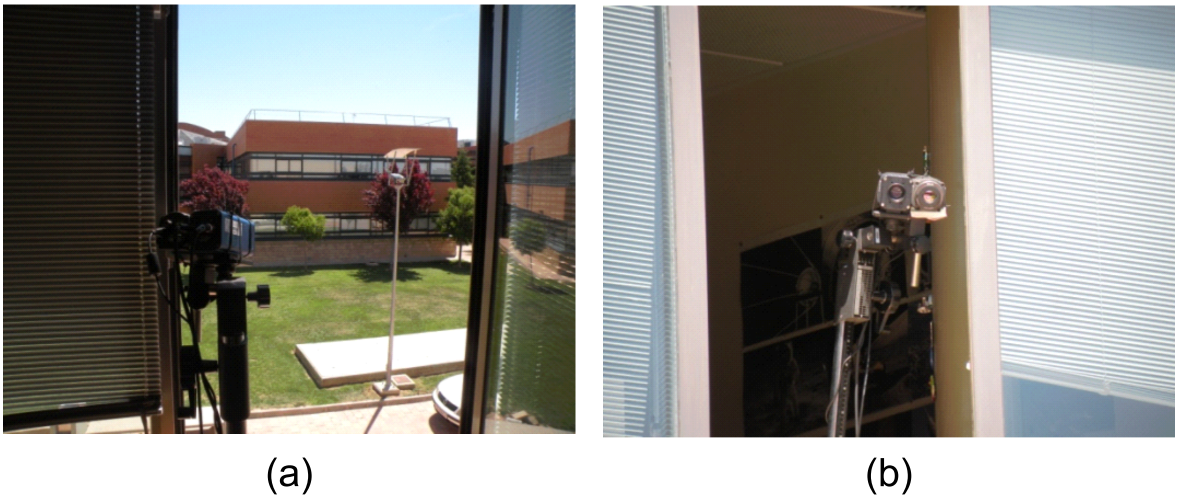
A test environment has been selected, and a bimodal sensor setup has been put into practice. Five different video sequences have been recorded under very different environmental conditions. The two segmentation algorithms described previously have been performed on the five video sequences in the visible-light and the thermal infrared spectra. In accordance with the evaluation criteria used during the experimentation, the confidence intervals have been tuned to select the best sensor (visible-light *vs.* thermal infrared) segmentation output.

The complete process is explained next in full detail.

5.1. Test Environment and Multimodal Sensor Setup

Both sensors (visible-light and thermal infrared camera) are placed in parallel and focused to a common point of the same scenario, since the objective is to obtain two similar views of the same scene. Back and front views of our installation can be observed in Figure 6a,b.

Figure 6. Installation for simultaneous acquisition on the thermal infrared and visible-light spectra. (**a**) Back view; (**b**) front view.



Simultaneous and synchronized acquisition from both cameras is possible thanks to a video encoder capable of grabbing frames from two cameras in the same instant. The frames acquired by the cameras connected to the encoder are separated into channels (one channel for each video input) and added to a buffer, which streams the video.

The chosen test environment is an outdoor environment. The two cameras are placed in a window of a building, 6 m in height, looking down at an angle of about 45 degrees. It was decided to work in an outdoor environment, since such an environment shows a high variety of temperatures (over a year) and lighting conditions (over one day, for example). Instead, an indoor environment does not usually admit such a large range of changes. The scene has no defined entrance area for humans. In the lower part of the scene is a concrete platform, which easily stores the environmental heat. This same property exists in the buildings surrounding the scene. The detection of humans in the thermal infrared spectrum will be problematic in the building in the background of the scene, since the infrared thermal camera automatically performs a thermal attenuation, resulting in a lack of precision in obtaining the temperature of humans near the building. Specifically, attenuation makes humans merge with the building.

5.2. Test Sequences

We have worked in this scenario with five different sequences in order to validate the proposal. In sequence *−2°Fog*, we find a single human being performing different activities in the environment. The human is walking almost all of the time, but also runs, crouches and sits in the central concrete platform. The sequence was recorded when fog partially covered the scene. In this sequence, it is difficult to distinguish the human in the visible light spectrum. However, the pedestrian can be recognized with relative ease in the visible-light spectrum. However, it is not simple to detect the human when he/she is close to the building. The *8°Night* sequence was recorded in order to evaluate the performance of our proposal at night (in the darkness). Visibility is almost zero in the color spectrum, and the infrared spectrum also presents problems. We see in this sequence that buildings remain hot, due to the heat that they have accumulated during daylight hours. Therefore, buildings are sometimes confused with humans walking around them. In this second sequence, there are two people walking in the scene. Both pedestrians occasionally cross each other's path.

In sequence *3°Sunny*, initially, we have a human being who is walking in the environment. Sometimes, he performs different actions, like crouching. Later, a second man appears, walking on different paths. Finally, the two humans cross their paths and meet in the concrete deck. The following sequence, *15°Cloudy*, presents a number of more complex actions. These actions are performed by a single human being. Let us highlight the action of sitting on the concrete deck core. The increase in temperature causes the appearance of human shadows on the platform. There is no doubt that in this way, complexity increases for human segmentation in the infrared spectrum. The last sequence is called *28°Sunny* and is the most complex for detecting people in the thermal infrared spectrum. In this sequence, we have three pedestrians walking in the scene and carrying out actions, such as sitting, crossing their paths and meeting. The high temperature makes it very difficult to distinguish humans in the infrared spectrum, especially in the concrete deck.

5.3. Evaluation Criteria

Recall, precision and F-score (F) were considered to evaluate the performance of the previously described segmentation algorithms. These are some measures widely used by the computer vision community. The definitions of the previous measures are provided in Equations (12)–(14), respectively.

$$recall = \frac{TP}{TP + FN} \tag{12}$$

$$precision = \frac{TP}{TP + FP} \tag{13}$$

$$F = \frac{2 \times precision \times recall}{precision + recall} \tag{14}$$

where *TP* (true positives) are the correct detections in the sequence, *FP* (false positives) are the mistaken detections and *FN* (false negatives) are the humans that are not detected, but are really present in the scene.

Precision is the ratio of true positives with respect to the total number of detections, *i.e.*, the ratio of detections that really correspond to a human. Moreover, recall is the ratio of a human on the scene to be really detected or not. Lastly, the F-score considers precision and recall, in this way providing an overall vision of the system performance. The F-score is a weighted average; an ideal system will show an F-score of one.

5.4. Confidence Threshold Setup

The results from the segmentation of the different sequences are analyzed in order to experimentally set the confidence interval thresholds in both spectra. Let us highlight that a setting up of all of the system parameters is necessary for each different environment and sensor setup. Although this paper does not describe a learning phase, this is obviously performed.

5.4.1. Confidence Thresholds in the Visible-Light Spectrum

The evaluation measures after segmenting the five video sequences from the visible-light sensor are shown in Table 1. For each sequence, in the first place, the table shows its average grey level value $\overline{I_G}$ and its segmentation statistics (recall, precision and F-score). Based on these input and output data, the last column of the table offers the confidence value that was experimentally assigned.

Table 1. Confidence values C_V for each sequence in the visible-light spectrum.

Sequence	$\overline{I_G}$	**Recall**	**Precision**	***F***	C_V
8°Night	49	0.08	1.00	0.15	*LOW*
3°Sunny	133	0.93	0.96	0.95	*HIGH*
15°Cloudy	125	0.97	0.98	0.97	
28°Sunny	116	0.81	0.97	0.88	
−2°Fog	147	0.52	1.00	0.68	*LOW*

When examining the results depicted in Table 1, a significant difference is observed between the sequences recorded under adverse lighting conditions (such as fog or darkness) and those recorded with more suitable conditions for the visible-light camera. These results confirm our initial assumption that a low or high average grey level value of the image would produce bad segmentation results, whereas an intermediate mean value would offer good or excellent results.

Indeed, in the sequences where the average grey level value $\overline{I_G}$ of the frames is between 95 and 135, the results are usually excellent, with F-score values always close to or greater than 90%. Their sensitivities are also usually very close to 100% of detected humans in the scene. On the other hand, the evaluation values are always remarkably lower in sequences with frames with an average grey level value outside of the established intermediate range. Thus, confidence threshold ζ_{VL} is set to the average grey level value of 95, while confidence threshold ζ_{VH} is assigned a value of 135.

5.4.2. Confidence Thresholds for Human Detection in the Thermal Infrared Spectrum

Since the confidence value for human detection in the thermal infrared spectrum C_{IR} for a sequence is based on confidence value C_V for that sequence, experimental results for the thermal infrared spectrum have been divided into two different tables, regarding the value of C_V. Results for values *HIGH* and *LOW* of C_V are shown in Tables 2 and 3, respectively. Both tables are organized in a similar manner as Table 1. The difference is that the new tables include the standard deviation of the average illumination of the sequence, as well as the the contrast υ_{IR}.

Table 2. Confidence values C_{IR} for each sequence in the visible spectrum with C_V set to $HIGH$.

Sequence	$\overline{I_{IR}}$	$\sigma_{I_{IR}}$	υ_{IR}	Recall	Precision	*F*	C_{IR}
3°Sunny	62	52	1.19	0.98	0.91	0.94	$HIGH$
15°Cloudy	86	44	1.96	0.91	0.97	0.94	$MEDIUM$
28°Sunny	113	46	2.46	0.39	0.96	0.55	LOW

Table 3. Confidence values C_{IR} for each sequence in the visible spectrum with C_V set to LOW.

Sequence	$\overline{I_{IR}}$	$\sigma_{I_{IR}}$	υ_{IR}	Recall	Precision	*F*	C_{IR}
−2°Fog	50	26	1.93	0.95	0.95	0.95	$HIGH$
8°Night	93	42	2.22	0.81	0.86	0.83	$MEDIUM$

Let us examine the results depicted in Table 3. Remember that this case corresponds to conditions where the visible-light results are poor. At first glance, a remarkable difference is appreciated between both sequences, where the night sequence has an F-score lower than 90%. The standard deviation of the grey level value $\sigma_{I_{IR}}$ of the frames of this sequence has a low value with respect to their average grey level value $\overline{I_{IR}}$, determining that humans are difficult to distinguish from the background in many cases. Therefore, confidence threshold ζ_{IRN} has been experimentally set to 2.0 based on the tested sequences.

Let us now consider the case where the visible-light spectrum offers good or excellent results. Under these conditions, less significant differences are observed in Table 2, except for the last sequence, where the environment temperature is greater than 20° and the results are drastically worse. Hence, the threshold ζ_{IRH} is experimentally set at a value of 2.2. It has been proven, as well, that precision suffers an abrupt decrease with environmental temperatures greater than 10°, when the parametrization is not set to more restrictive values than at lower temperatures. Since coefficient υ_{IR} is always above 1.9 for high temperatures, 1.9 is the value assigned to confidence threshold ζ_{IRL}.

5.5. Final Results

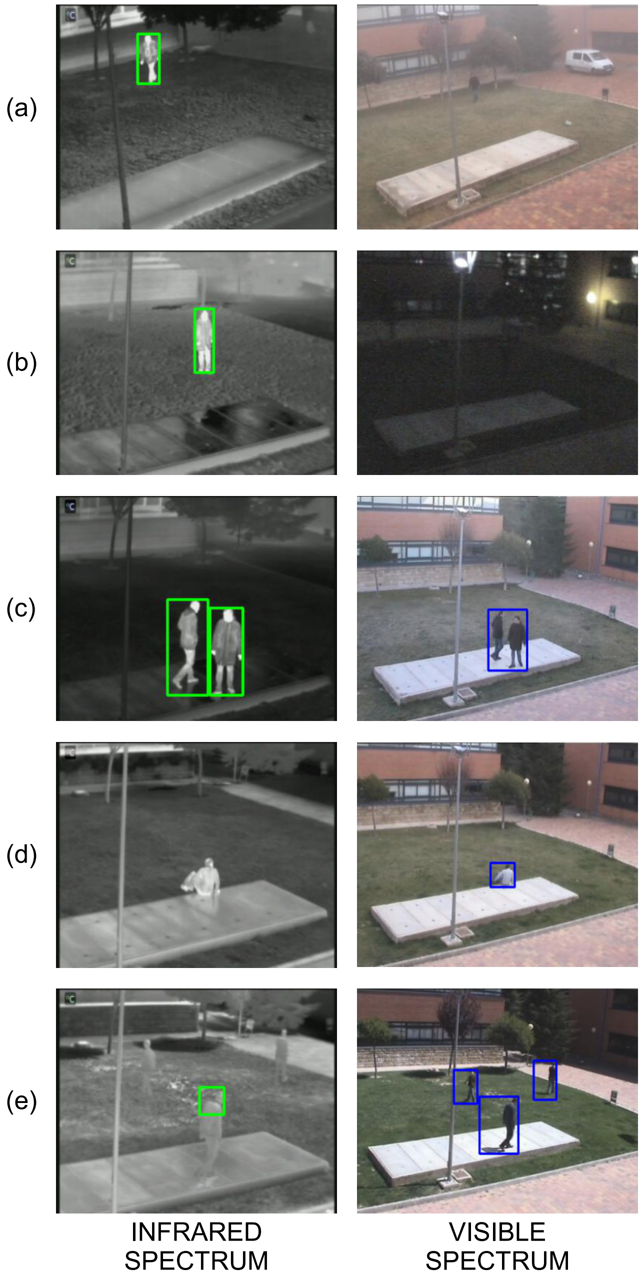
Once the confidence intervals for both visible-light and thermal infrared spectra have been fixed through the experimental assignment of the confidence threshold values (see Figure 1), we summarize the segmentation results in Table 4 with the objective of highlighting which spectrum is suited better for the environmental conditions of the recorded sequences. The *−2°Fog* and *8°Night* sequences show better performance in the thermal infrared spectrum, whilst the *15°Cloudy* and *28°Sunny* sequences work better in the visible-light spectrum. Lastly, sequence *3°Sunny* shows similar results for both spectra.

Table 4. Final results achieved for each sequence in both spectra.

Sequence	Spectrum	Recall	Precision	*F*	Confidence Value
−2°Fog	Visible	0.52	1.00	0.68	*LOW*
	Infrared	0.95	0.95	0.95	**HIGH**
8°Night	Visible	0.08	1.00	0.15	*LOW*
	Infrared	0.81	0.86	0.83	**MEDIUM**
3°Sunny	Visible	0.93	0.96	0.95	**HIGH**
	Infrared	0.98	0.91	0.94	**HIGH**
15°Cloudy	Visible	0.97	0.98	0.97	**HIGH**
	Infrared	0.91	0.97	0.94	*MEDIUM*
28°Sunny	Visible	0.81	0.97	0.88	**HIGH**
	Infrared	0.39	0.96	0.55	*LOW*

Now, we are going to explain the results obtained for each sequence by providing some qualitative examples shown in Figure 7. Results for the sequence *−2°Fog* show problems in the visible spectrum when humans are close to the building in the scene background. There is a greater fog concentration in that zone, as shown in Figure 7a. At the same time, humans close to the camera are not easy to detect, because scene colors appear more attenuated, due to the fog, and humans are harder to detect. These problems were predictable due to the illumination features of the frames composing this sequence. On the other hand, human detection in the infrared spectrum barely has any problems, and false negatives only appear on a few occasions. In Figure 7a, it can be appreciated that the human is detected without any problem in the infrared spectrum. The results confirm the confidence value $HIGH$ in the thermal infrared spectrum.

Figure 7. Results obtained on both spectra for the analyzed sequences. (**a**) Sequence *−2°Fog*; (**b**) sequence *8°Night*; (**c**) sequence *3°Sunny*; (**d**) sequence *15°Cloudy*; (**e**) sequence *23°Sunny*.



Sequence *8°Night* is especially difficult for the different algorithms, because it was recorded in the early night hours. At that time, buildings have not yet cooled, and their thermal readings are still high compared to the environment, causing them to appear at almost the same temperature as the humans in the scene. However, human detection in the thermal infrared spectrum works well when people are far from the background, as shown in Figure 7b. Though, problems can appear when people approach the buildings and the human body temperature is similar to the thermal readings of the building in the background. These factors cause the confidence on the infrared spectrum to be set as $MEDIUM$. On the other hand, the visible-light camera is blinded, since the scene was recorded in night conditions. An example of this situation in the visible spectrum is shown in Figure 7b. Humans can only be detected when they are near the lamp post lighting the scene in one of the scenario sides. Because of this, the confidence of the visible spectrum is set as LOW.

Results for both spectra are great in the sequence *3°Sunny*. Due to the contrast and lighting conditions, both confidence values are set as $HIGH$, and both sensors usually detect humans in the scene. The thermal infrared spectrum only shows problems when the human is close to the building's wall, while the human detection in the visible spectrum shows false negatives when humans are close to each other (as shown in Figure 7c) and when a human is covered in shadows far away from the camera.

In the sequence *15°Cloudy*, temperatures of the humans are similar to the thermal readings of the elements in the scene and only motion information prevents the results of human detection in the thermal infrared spectrum from being lower. For example, Figure 7d shows how the human could not be detected in thermal infrared, since the thermal readings of his/her clothes are similar to the temperature of the concrete platform where he/she is sitting. Because of this fact, the contrast value υ_{IR} of the frames composing this sequence is usually intermediate, making the confidence C_{IR} on the infrared spectrum to be set as $MEDIUM$. However, the performance of the human detection in the visible spectrum is excellent, due to the illumination conditions of the scene, as shown in Figure 7d. These conditions confirm the confidence value on the visible spectrum to be assigned as $HIGH$.

Finally, sequence *28°Sunny* shows a poor performance of the human detection in the thermal infrared spectrum, since the thermal readings of the human are very similar to the temperature of the remaining elements in the scene. An example of this situation is shown in Figure 7e. A hint to this problem is shown by the high value of contrast υ_{IR} in the frames composing this sequence, which results in the confidence value C_{IR} of the thermal infrared spectrum being set to LOW. This confirms the difficulty for distinguishing humans with the thermal infrared camera under these conditions. On the other hand, a lot of groups appear in this video. This circumstance causes the performance of the human detection in the visible spectrum to worsen. However, when individuals walk alone, they are detected without any problem, as shown in Figure 7e. The lighting conditions of the scene confirm these results, and the confidence value C_V is set to $HIGH$.

6. Conclusions

This paper has introduced a confidence measure scheme in a bimodal camera setup for automatically selecting visible-light or thermal infrared in response to natural environmental changes. The purpose of the setup is to robustly detect people in dynamic outdoor scenarios under very different conditions.

Visual-light and thermal infrared sensors possess advantages and disadvantages that are complementary. On the one hand, thermal infrared cameras are useful for detecting humans under adverse illumination conditions. On the other side, visible-light cameras yield good results when conducting human detection in well lit environments, but it is problematic in dark environments or in areas of the scene that present shadows or have low visibility in general. In order to take advantage of both spectra, some researchers have decided to perform image fusion by using visible and infrared images together. However, as the fusion of thermal infrared and visible images is difficult, we have opted for using a selection-based approach that switches from one sensor to another depending on the environmental conditions of the scene.

The approach is based on confidence measures that are set for each spectrum using the mean illumination, the standard image deviation in thermal infrared, and the average grey level value in visible-light. In the visible-light spectrum, there are three different confidence intervals separated by two confidence threshold values. The confidence value in the visible spectrum takes the linguistic values $HIGH$ and LOW, corresponding to favorable and adverse illumination conditions, respectively. A $MEDIUM$ confidence value is not used in the visible-light spectrum.

In the thermal infrared spectrum, there are three possibilities for assigning the confidence intervals, as the confidence value in thermal infrared is directly related to the value in visible-light. When the confidence value of the visible-light spectrum is set to LOW, it is clear that the visible-light sensor is almost unable to distinguish humans. This is why the infrared spectrum is always forced with a confidence value above LOW (that is, $MEDIUM$ or $HIGH$). Indeed, under this circumstance, the infrared sensor always works better than the visible-light one. On the other hand, when the detection in visible-light offers good results ($HIGH$ confidence value), the confidence value in the thermal infrared spectrum can take values LOW, $MEDIUM$ or $HIGH$.

In order to evaluate the proposal, two efficient segmentation algorithms, one dedicated to the visible-light spectrum and another one to the thermal infrared spectrum, were implemented and five different video sequences were recorded under very different environmental conditions. The results of the segmentation in both spectra permitted setting up the best-suited confidence interval thresholds. Moreover, the results obtained confirmed the accuracy of the approach.

Acknowledgments

This work was partially supported by the Spanish Ministerio de Economí y Competitividad/FEDER under the TIN2013-47074-C2-1-R and TIN2010-20845-C03-01 grants.

Author Contributions

The three authors contributed to the different phases of the research and to the writing of this paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Fernández-Caballero, A.; Castillo, J.; Rodríguez-Sánchez, J. Human activity monitoring by local and global finite state machines. *Exp. Syst. Appl.* **2012**, *39*, 6982–6993.
2. Dollár, P.; Wojek, C.; Schiele, B.; Perona, P. Pedestrian detection: An evaluation of the state of the art. *IEEE Trans. Patt. Anal. Mach. Intell.* **2012**, *34*, 743–761.
3. Navarro, E.; Fernández-Caballero, A.; Marttínez-Tomás, R. Intelligent multisensory systems in support of information society. *Int. J. Syst. Sci.* **2014**, *45*, 711–713.
4. Fernández-Caballero, A.; López, M.; Serrano-Cuerda, J. Thermal-infrared pedestrian ROI extraction through thermal and motion information fusion. *Sensors* **2014**, *14*, 6666–6676.

5. Goubet, E.; Katz, J.; Porikli, F. Pedestrian tracking using thermal infrared imaging. *Proc. SPIE* **2006**, *6206*, doi:10.1117/12.673132.
6. Leykin, A.; Hammoud, R. Pedestrian tracking by fusion of thermal-visible surveillance videos. *Mach. Vis. Appl.* **2008**, *21*, 587–595.
7. Ding, M.; Wei, L.; Wang, B. Research on fusion method for infrared and visible images via compressive sensing. *Infrared Phys. Technol.* **2013**, *57*, 56–67.
8. Zhu, Z.; Huang, T. Multimodal surveillance: An introduction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 18–23 June 2007; pp. 1–6.
9. Enzweiler, M.; Gavrila, D. Monocular pedestrian detection: Survey and experiments. *IEEE Trans. Patt. Anal. Mach. Intell.* **2009**, *31*, 2179–2195.
10. Delgado, A.; Lo´pez, M.; Ferna´ndez-Caballero, A. Real-time motion detection by lateral inhibition in accumulative computation. *Eng. Appl. Artif. Intell.* **2010**, *23*, 129–139.
11. Schwartz, W.; Kembhavi, A.; Harwood, D.; Davis, L. Human detection using partial least squares analysis. In Proceedings of the IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 27 September–4 October 2009; pp. 24–31.
12. Rodriguez, M.; Shah, M. Detecting and segmenting humans in crowded scenes. In Proceedings of the 15th International Conference on Multimedia, Augsburg, Germany, 24–29 September 2007; pp. 353–356.
13. Carmona, E.; Martínez-Cantos, J.; Mira, J. A new video segmentation method of moving objects based on blob-level knowledge. *Patt. Recogn. Lett.* **2008**, *29*, 272–285.
14. Anwer, R.; Vázquez, D.; López, A. Opponent colors for human detection. In Proceedings of the 5th Iberian Conference on Pattern Recognition and Image Analysis, Las Palmas de Gran Canaria, Spain, 8–10 June 2011; pp. 363–370.
15. Olmeda, D.; de la Escalera, A.; Armingol, J. Far infrared pedestrian detection and tracking for night driving. *Robotica* **2011**, *29*, 495–505.
16. Li, J.; Gong, W.; Li, W.; Liu, X. Robust pedestrian detection in thermal infrared imagery using the wavelet transform. *Infrared Phys. Technol.* **2010**, *53*, 267–273.
17. Kumar, P.; Mittal, A.; Kumar, P. Fusion of thermal infrared and visible spectrum video for robust surveillance. In *Computer Vision, Graphics and Image Processing*; Springer: Heidelberg, Germany, 2006; pp. 528–539.
18. Fang, Y.; Yamada, K.; Ninomiya, Y.; Horn, B.; Masaki, I. A shape-independent method for pedestrian detection with far-infrared images. *IEEE Trans. Vehic. Technol.* **2004**, *53*, 1679–1697.
19. Davis, J.; Sharma, V. Background-subtraction using contour-based fusion of thermal and visible imagery. *Comput. Vis. Image Underst.* **2007**, *106*, 162–182.
20. Arandjelovic, O.; Hammoud, R.; Cipolla, R. Thermal and reflectance based personal identification methodology under variable illumination. *Patt. Recogn.* **2010**, *43*, 1801–1813.
21. Adu, J.; Gan, J.; Wang, Y.; Huang, J. Image fusion based on nonsubsampled contourlet transform for infrared and visible light image. *Infrared Phys. Technol.* **2013**, *61*, 94–100.

22. Bai, X.; Chen, X.; Zhou, F.; Liu, Z.; Xue, B. Multiscale top-hat selection transform based infrared and visual image fusion with emphasis on extracting regions of interest. *Infrared Phys. Technol.* **2013**, *60*, 81–93.
23. Li, Q.; Sun, Z.-X.; Chen, S.-L. A POMDP-based camera selection method. In Proceedings of the International Conference on Computer Vision Theory and Applications, Barcelona, Spain, 21–24 February 2013; Volume 1, pp. 746–751.
24. Socolinsky, D. Design and deployment of visible-thermal biometric surveillance systems. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 18–23 June 2007; pp. 1–2.
25. Shen, C.; Zhang, C.; Fels, S. A multi-camera surveillance system that estimates quality-of-view measurement. In Proceedings of the 14th IEEE International Conference on Image Processing, Atlanta, GA, USA, 8–11 October 2006; Volume 3, pp. 193–196.
26. Fernández-Caballero, A.; Castillo, J.; Serrano-Cuerda, J.; Maldonado-Bascón, S. Real-time human segmentation in infrared videos. *Exp. Syst. Appl.* **2011**, *38*, 2577–2584.
27. Serrano-Cuerda, J.; Castillo, J.; Sokolova, M.; Fernández-Caballero, A. Efficient people counting from indoor overhead video camera. In *Trends in Practical Applications of Agents and Multiagent Systems*; Springer: Heidelberg, Germany, 2013; Volume 221, pp. 129–137.
28. KaewTraKulPong, P.; Bowden, R. An improved adaptive background mixture model for real-time tracking with shadow detection. In *Video-Based Surveillance Systems*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2002; pp. 135–144.
29. Horprasert, T.; Harwood, D.; Davis, L. A statistical approach for real-time robust background subtraction and shadow detection. In Proceedings of the IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–25 September 1999; pp. 1–19.