Robust Human Detection and Tracking in Intelligent Environments by Information Fusion of Color and Infrared Video

Juan Serrano-Cuerda, María Teresa López and Antonio Fernández-Caballero

Abstract—This paper is related to ambient intelligence systems capable of locating and tracking humans. These are the first steps of a human-centered ambient intelligent system, ranging from data acquisition to robust tracking, for the purpose of interpreting human behaviors in monitored environments. The first objective is to improve human detection through the fusion of thermal-infrared and color video segmentation. On the level following to segmentation, the traditional tracking problems (e.g. occlusions, crossings, etc.) are faced. Finally, the use of several classifiers such as support-vector machines and artificial neural networks are proposed to enhance the tracking level. The work proposes a combination of both color and thermal-infrared spectra in human tracking.

I. INTRODUCTION

Ambient Intelligence (AmI) is a paradigm that supports the design of intelligent systems which adapt the environment to the user’s need [19]. With this objective, AmI groups a large number of disciplines (e.g. software design, computer vision, information fusion, speech recognition, and so on). In AmI information is usually extracted from a heterogeneous network of distributed sensors. This proposal focuses on extracting and interpreting the information captured from both a thermal infrared and a color camera. Then the information is fused to interpret the obtained data for detecting the humans present in the scene. Also, the humans’ trajectories are evaluated as a starting step in an AmI system.

Let us remind that human detection as such is still a challenging issue in computer vision. Nowadays the use of color information is widely spread in closed-circuit television (CCTV) systems. In current surveillance domains, mostly an automatic or semi-automatic approach to human detection in color video is investigated. Unfortunately, color-based segmentation methods have low resistance to lightning changes and environmental phenomena (e.g. fog or smoke). In order to minimize the influence of these external factors, the use of cameras in the thermal-infrared spectrum has also been proposed. However, thermal cameras carry their own problems, including the presence of halos which decrease the algorithms’ precision when delimiting the human silhouettes.

II. RELATED WORK

Nowadays, human detection and tracking is a main research field in computer vision. The number of approaches to human detection and tracking is enormous. Thus, only a few approaches in human detection [1], [2] and human tracking [18], [12] in infrared and color spectra are described next. Also, different proposals using information fusion of both spectra [13], [15] are detailed.
A. Segmentation and tracking in the infrared spectrum

The infrared spectrum has been widely used to detect humans. In [5], a background subtraction is initially performed to identify local foreground objects. Then gradient information from the objects and the background is combined to generate a contour map representing the confidence for each pixel to belong to a person’s boundary. Finally, a path search is performed to complete broken segments. Also, an approach based on histograms is introduced in [7]. In this work a vertical and a horizontal resize are realized on the pedestrian candidate regions. Two different cases for summer and winter are established, using the gray level values to delimit pedestrians in the winter case, and the intensity changes in the human boundaries to delimit pedestrians in summer sequences. Recently, a pedestrian detection method in thermal infrared imagery using the wavelet transform has been proposed [16]. The proposal uses region features obtained from a conversion to the frequency domain as input to SVM.

These approaches usually assume that there is enough thermal contrast to distinguish the humans from the background. Obviously, color information should complement the thermal information when there is low contrast in the input image.

B. Segmentation and tracking in the color spectrum

Many color spaces are used in computer vision, offering different information about the color in the image, regarding intensity or composition. The following color-based approaches use different approximations to detect and track the humans present in a video. The approach introduced in [2] represents the input frame by clustering the pixels in a common feature space of color-texture-location with an automatic segmentation algorithm. First, an appropriate scale is selected for each pixel extracting color, texture and location features in the selected scale. Then, the pixels are clustered through modeling the pixels feature distribution. An automatic segmentation algorithm is proposed in [6]. First, the image colors are quantified to represent classes which are used to characterize regions in the image. In second place, the image pixel colors are replaced by their new label in an image classes map. Next, a criterium is applied to local windows with the map, obtaining an image where the high and low values correspond to possible region boundaries and their centers, respectively. Finally, a region growing method is used to segment the image. More recently, a segmentation method based in background subtraction method is introduced [1]. The background is dynamically adapted by filtering elements that usually appear in background subtraction such as shadows, reflections, and so on.

As aforesaid, color-based methods for segmentation and tracking in the color spectrum are limited by image illumination. But, this color information can be mixed with the thermal-infrared information in order to improve the results of the segmentation.

C. Segmentation and tracking using information fusion

Many approaches to information fusion can be found in the literature, but this work focuses infrared and color spectra data fusion. Information fusion approaches usually use classifiers to assign weights or to match the information extracted from each spectrum (e.g. SVM, Kalman filters [14], and so on). In [12], it is established that the performance of visible cameras is independent of the ambient temperature, but thus depends on the fact of having some minimal level of ambient light. Also, performance of infrared cameras strongly depends on the scene temperature, making them less reliable for summer scenarios. While efficient segmentation can be achieved in summer using background subtraction, classification based on shape or on human body feature analysis is more difficult due to the lower amount of object information. A paper [18] presents a system able to detect and track multiple people using a stereo camera placed at a frontal position below the people heads. People are tracked by combining color and position information using a Kalman filter. In another approach [13], background subtraction is performed in both spectra, extracting the silhouettes from each spectrum. Then, a genetic algorithm is employed to find correspondences so that the preliminary silhouettes from the color and thermal images are well matched. Finally, probabilistic strategies are used to obtain better body silhouette extraction results.

III. ROBUST HUMAN DETECTION AND TRACKING IN COLOR AND INFRARED VIDEO

As already mentioned previously, our proposal starts from a robust human detection in infrared and color video. The results obtained are then fused in order to improve the detection using the information from both spectra. Finally, a human tracking approach that uses the information supplied by the fusion of the color and infrared segmentation is proposed. These steps are shown in Fig. 2.

![Fig. 2. Our proposal of robust human detection and tracking in fused color and infrared video](image-url)
A. Robust human detection in infrared video

Our research team has already proposed several algorithms for human detection in infrared image sequences. These are briefly described next.

1) Human detection based on a single frame: The first approach to human detection in infrared does not use motion information and is based on a single frame captured from a thermal infrared camera [8]. The main idea is based on the notion that humans usually appear warmer than the background in a thermal infrared image. Using this information, a single frame is used as an input image. The image is binarized and the obtained candidate-to-human blobs are resized in accordance with valid human heights and widths.

2) Human detection based on optical flow: Using the human detection algorithm described above, an improvement is proposed to work with frames captured from a moving camera [3]. The camera is mounted on a mobile robot. This approach includes an additional stage for dynamically analyzing the motion present in the frame through optical flow calculation. The information provided in the human detection based on a single frame is combined with the optical flow result. Notice that the information obtained by the human detection in a single frame is never discarded. Indeed, optical flow reinforces the belief of the presence of moving humans in a single image. Therefore, the humans remaining static are still detected, although more restrictions are applied to static than to moving people.

3) Human detection based on image subtraction: As aforementioned, sometimes humans have low contrast with the background in frame sequences acquired with a static thermal-infrared camera. However, humans are detected through an image subtraction operation if they are not stationary. The idea is similar to the one used in the optical flow (the moving blobs are treated in a similar way in both approaches). Motion information is used to increase the number of human detected. We are able to find the humans present in the scene without discarding the information obtained from the human detection based on a single frame. The main difference with the optical flow is that this detection is developed to work with static cameras. Also notice that this option is computationally less expensive. A comparison between the performance of human detection based on optical flow and image subtraction has been offered [10].

B. Human detection in color video

Also, in this case, our research team has already proposed several algorithms for human detection in the visible spectrum. Therefore, this proposal follows the main stages of the accumulative computation approach. Up to date accumulated computation has been extensively applied to gray level video [11]. The approach proposed now is to adapt accumulative computation for color image sequences. The first idea is to process each one of the described stages for each different component of a chosen color space, such as red-green-blue (RGB) and hue-saturation-lightness (HLS) color spaces. Then, the spots obtained before the application stage are fused by means of an AND operation to obtain spots that present simultaneous motion in each different color component, thus reducing the amount of false positives. The goal consists in comparing several color spaces in order to establish the best performing space according to the conditions of the scene.

C. Fusion

Data fusion is a new topic of research on our team. But there are some relevant papers that fit our overall proposal. In [4] people are detected in the thermal infrared spectrum, using the information to segment the corresponding region in the color image. Firstly, humans are detected based on their thermal information, establishing a range of temperatures where humans are located. Then each region of interest (ROI) containing a person is divided into three subregions (head, torso and legs) whose thermal and color features act as input to a pattern recognition system. For each subregion, the mean variance and the average temperature as well as the hue, saturation and intensity statistical values (components of the HSV color space) are stored with a total amount of 24 features. We consider this approach very interesting because it uses the color features as well as the temperature which covers the cases where there is low lighting in the scene.

Another interesting trend found in literature involves the weighting of infrared and color segmentation. [15] combines fuzzy logic and Kalman filters to model the motion of the objects detected in each spectrum and fuzzy logic to assign weights to each spectrum according its reliability. This is related to the previously mentioned idea that each spectrum works well in different conditions which are evaluated in this approach.

Also, the initial ideas of this proposal include the use of SVM. Fusion has to include the processes of infrared and color segmentation. The parameters associated to segmentation can be used as elements of an input feature vector. In fact, SVM can be used to establish the weight of each segmentation parameter to reduce false positives and improving the belief that a detected object is really a human.

D. Tracking

In the recent past, an initial tracking scheme has been described in our research group [10]. This first scheme does not take into account color or infrared features. The way the tracking proposal works is shown in Fig. 3. The most important aspect of the approach is the use of a permanence charge value (related to the accumulative computation method) to control if a human has left the scene. The permanence is charged to the maximum value if a human is detected during several consecutive frames and is discharged if a previously detected human is no more detected. If the permanence reaches a minimum value and the human was close to the scene border when it stopped being detected, the blob is discarded. It is assumed that the human has left the scene. Otherwise, the human continues being labeled, as humans cannot simply disappear from the middle of a scene.
IV. CONCLUSIONS AND FUTURE WORK

The current proposal aims to study and implement different approaches to human detection in the thermal-infrared and the color spectrum, as well as several fusion methods in order to combine the information from both spectra. A tracking approach will also be developed according to the information supplied by the spectra fusion.

We have already developed different algorithms in the infrared spectrum, based on a single frame, on camera motion and on image subtraction. Our algorithms currently have a big amount of parameters that we plan to reduce through the use of classifiers. We have also a solid previous experience in color spaces. Our first approach studied the RGB and HLS color spaces, but we are planning to expand this approach with additional color spaces such as CIELAB, YUV, and so on. One part of the work is focused on the study of which color space works better in different situations, regarding the illumination conditions, the contrast of the scene, among others. The aim is to select the color space that better fits to the different real-world situations.

We are in the early stages of data fusion. We are considering the use of SVM classifiers to choose which features work better for each situation, as well as ANN to learn the parameters needed for a better tuning according to the scene conditions. We are also considering several approaches that include fuzzy logic as well as different approaches to match the information obtained from each spectrum.

An initial approach to robust tracking has also been developed. The information supplied by the fusion algorithms will be used as different features in the tracking task with the objective of smoothing the problems caused by occlusions and crossings when identifying humans in a complex scene.

V. ACKNOWLEDGMENTS

This work is partially supported by the Spanish Ministerio de Ciencia e Innovación under project TIN2010-20845-C03-01, the Spanish Ministerio de Industria, Turismo y Comercio under project TSI-020100-2010-261 and by the Spanish Junta de Comunidades de Castilla-La Mancha under projects PII2109-0069-0994 and PEII09-0054-9581.

REFERENCES