Robust human detection through fusion of color and infrared video

Juan Serrano-Cuerda

Instituto de Investigación en Informática de Albacete, 02071 - Albacete, Spain
Advisors: Antonio Fernández-Caballero and María T. López
Date and location of PhD thesis defense: 16 December 2013, University of Castilla-La Mancha

---

1 Abstract

Nowadays, robust *human detection* is still a key challenge in the field of Computer Vision. Many people detection systems are based on the use of color cameras. Yet, these cameras have problems when the scene is poorly illuminated or when there are sudden lighting changes. This is why, the use of thermal-infrared cameras seems to be an interesting alternative. Indeed, these cameras show a good performance in cold environments. But they offer many troubles in warm scenarios. Under these adverse conditions, human temperature is similar to the thermal readings of the remaining scene elements. This fact makes it hard to distinguish humans from the environment. This PhD thesis develops and implements a robust multisensor [1] human detection system based on fusing the information provided after segmenting infrared [2] and color videos. The final system has been developed based on the INT^3-*Horus* framework [3] recently created in our *n&aIS* research team [4].

The framework starts with an *acquisition level* which grabs frames from both cameras and synchronizes their output. The features of the captured images are used to assign a *confidence degree* to each spectrum. The confidence degree is the core element for the purpose of information fusion between infrared and visible spectra. The mean illumination value of a color image is the base for establishing confidence in the color spectrum. On the other hand, the mean illumination and the standard deviation of the image provide the necessary information about its contrast in the infrared spectrum. Now, the actual confidence is updated in each couple of acquired frames. So, if the scene changes its temperature, illumination, and so on, the fusion algorithm will adapt itself to the new conditions. A scheme of how the confidence levels are assigned is shown in Figure 1.

The next level is *segmentation* of infrared and visible videos [5, 6]. A series of different human detection algorithms have been implemented at this level, both for infrared and visible (color) spectra. The different algorithms are compared among each other with the objective of choosing the most suitable on for human detection at each spectrum for a given scenario and condition. The results of the selected color and infrared algorithm are then used at *fusion level*. Fusion is based on a rule-based system which uses the confidence degree assigned to each spectrum. The location of the humans detected by each segmentation approach is also used. In accordance with the rules, the decision taken can be (1) to add humans to the final result, (2) to better adjust the dimensions and the amount of detected humans by analyzing the results gotten by the other spectrum, or (3) to ignore the current detection if the confidence degree assigned to the spectrum is not above a certain value.
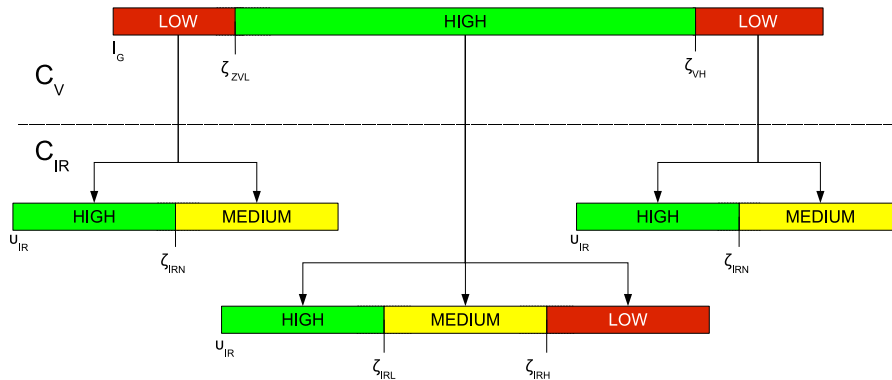

Figure 1: Confidence levels for color and infrared video fusion

Finally, a *tracking level* decides if the humans previously detected remain in the scene, although they have not been detected in the current frame.

An outdoor environment was chosen for *evaluation of the system*. Twelve different sequences were recorded under diverse atmospherical and illumination conditions. The sequences offer a varying complexity where a different number of humans are present in diverse situations. These situations are that simple as a single human walking on the scene, and that complex as multiple people walking together. The tests show a significant improvement between the results achieved by human detection algorithms focused on a single spectrum and those offered after information fusion. This improvement does not only occur in adverse conditions for each spectrum, but also in situations where both spectra collaborate through reinforcing their results. It is confirmed that the developed system shows an increased performance in terms of *precision* and *F-score*.

Acknowledgments

References

[1] J.M. Gascueña, J.C. Castillo, E. Navarro, A. Fernández-Caballero, "Engineering the development of systems for multisensory monitoring and activity interpretation", *International Journal of Systems Science* 45(4):728-740, 2014.

[2] A. Fernández-Caballero, M.T. López, J. Serrano-Cuerda, "Thermal-infrared pedestrian ROI extraction through thermal and motion information fusion", *Sensors* 14(4):6666-6676, 2014.

[3] A. Fernández-Caballero, J.C. Castillo, M.T. López, J. Serrano-Cuerda, M.V. Sokolova, "INT3-Horus framework for multispectrum activity interpretation in intelligent environments", *Expert Systems with Applications* 40(17):6715-6727, 2013.

[4] natural & artificial Interaction Systems, http://www.i3a.uclm.es/louise/nais/index_EN.htm, 2014.

[5] A. Fernández-Caballero, M.V. Sokolova, J. Serrano-Cuerda, "Lateral inhibition in accumulative computation and fuzzy sets for human fall pattern recognition in colour and infrared imagery", *The Scientific World Journal* article ID 935026, 2013.

[6] A. Fernández-Caballero, J.C. Castillo, J. Serrano-Cuerda, S. Maldonado-Bascón, "Real-time human segmentation in infrared videos", *Expert Systems with Applications* 38(3):2577-2584, 2011.