Open access • Proceedings Article • DOI:10.1109/ICPR.2016.7899716

Dynamic collaboration of far-infrared and visible spectrum for human detection
— Source link ↗

Paul Blondel, Alex Potelle, Claude Pegard, Rogelio Lozano ...+1 more authors

Related papers:

- Illumination independent Change Detection for Real World Image Sequences

- A hand posture recognition system utilizing frequency difference of infrared light

- SparsePPG: Towards Driver Monitoring Using Camera-Based Vital Signs Estimation in Near-Infrared

- The ship-borne infrared searching and tracking system based on the inertial platform (EI CONFERENCE)

- A New Method of High Temporal Resolution Remote Sensing Imaging: Moving Target Detection with Bionics Compound Eye

Dynamic collaboration of far-infrared and visible spectrum for human detection

Paul Blondel, Alex Potelle, Claude Pegard, David Lara, Rogelio Lozano

HAL Id: hal-01711411

https://hal.archives-ouvertes.fr/hal-01711411

Submitted on 23 Feb 2018

Dynamic collaboration of far-infrared and visible spectrum for human detection

Paul Blondel, Alex Potelle and Claude Pégard
Université Picardie Jules-Verne
80000 Amiens, France
firstname.lastname@u-picardie.fr

David Lara
University of Tamaulipas
Reynosa Tamaulipas, Mexico
dlara@docentes.uat.edu.mx

Rogelio Lozano
Université Technologique de Compiègne
60200 Compiègne, France
rogelio.lozano@hds.utc.fr

Abstract—This paper is about the collaborative use of a far-infrared spectrum human detector and a visible spectrum human detector; the idea is to make collaborate these two detectors of different nature to automatically adapt the human detection whatever the luminosity changes and whatever the infrared emission changes of the scene. Our collaborative approach of detection handles: 1) gradual luminosity changes due, for instance, to the passage from night to day (and vice-versa), 2) sudden luminosity changes due, for instance, to navigation in a forest (when going through a glade in a forest), 3) infrared emission saturation when the global temperature of the scene is very high and does not permit to distinguish human people in infrared. Our approach of detection permits to detect people 24 hours a day and regardless the weather conditions. Furthermore, the proposed approach is relatively fast: it is practically as fast as using one detector alone whereas two are used in the same time.

I. INTRODUCTION

Human detection is still a challenging problem despite the recent advances in the field. There is still room for improvement to make human detection faster and more robust. The classic approaches to enhance the detection often consist of designing better features and/or choosing better classification methods. Furthermore, as is frequently the case, the detection relies on one camera only. In this paper, we propose to extend the robustness of human detection using a far-infrared human detector and a visible human detector in concert, in a flexible and fast manner.

A. Human detection in the visible spectrum

Almost all the actual state-of-the-art human detection algorithms rely on classifiers: a classifier is first trained and, at runtime, the classifier is called to take decisions. A very popular algorithm based on this approach is the Histogram of Oriented Gradients (HOG) detector of Dalal and Triggs [1]. Since then, researchers have continued to improve detectors' speed and robustness. Amongst the top-performing ones, many state-of-the-art algorithms are based on the Integral Channel Features (ICF) detector of Dollár et al. [2]. It uses integral images and a soft-cascade classification to speed up the detection. Amongst the most well-known ICF-based detectors: the Fastest Detector in the West (FPDW) [3] approximates the features in the image pyramid, whilst the very recent and fast Aggregate Channel Features (ACF) detector performs pixel look-ups in aggregated image channels [3].
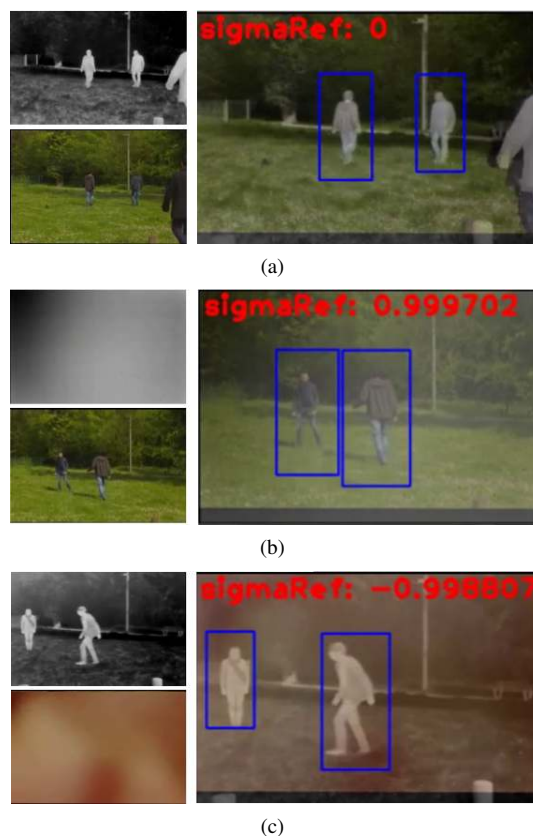

Fig. 1. Our collaborative detection approach adapts itself if one detector is suddenly failing or one camera is becoming inefficient: (1) the visible and the far-infrared spectrum are available, people are detected, (2) the infrared spectrum is no longer available, the detection is still continuing and (3) the visible spectrum is no longer available, the detection is still continuing.

B. Human detection in the infrared spectrum

Most of the works concerning human detection in the infrared spectrum are adaptations of approaches used in the visible spectrum. Zhang et al. compared different configurations of detectors and suggested that there is no need to invent radically different methods in this modality to perform well [4]. Olmeda et al. proposed a discrete approach for human detection based on a robust infrared spectrum descriptor: the histogram of phase congruency orientation [5]. The phase congruency is less sensitive to changes of contrast and, thus, performs well on infrared images [6]. However, computing the phase congruency is computationally intensive when compared

to other simpler and yet efficient descriptors. Recently, Brehar et al. proposed to combine a search space reduction with a slightly modified ACF detector: resulting in 30Hz fast human detection [7].

C. Combining the visible and the infrared spectrum

The information available through the visible spectrum is very rich. Indeed, it is possible to obtain: gradient information, contrast information, color information, etc. This information allows a very detailed analysis of the scene. However, this richness of information is only available if there is enough luminosity in the scene. Indeed, a lack of luminosity impoverishes the information: the contrast decreases and the color range shrinks. The performance of lots of computer vision approaches are impacted by this phenomenon.

Infrared emission is more stable in time than visible light emission. Because infrared emission sources are less impacted by changes of luminosity: they might continue to emit in total darkness depending on the emission source, such as the human body. However, infrared camera sensors have some limitations: they can saturate when the infrared emission is too strong. Besides, when two infrared emission sources overlaps each other, it is hard to distinguish from one another. Note that, people's clothing may also absorb a lot of infrared emission.

When it comes to detect people, the visible and the infrared spectrums may be seen as complementary. Human beings are natural infrared emission sources, they may be spotted just by finding abnormally high temperature regions of the scene having a general human shape. In the visible spectrum, human beings are sets of very specific visual features: complex mixtures of contours and colors. There are two majors fusion approaches to benefit from the combination of the visible and the infrared spectrum: 1) the low-level fusion and 2) the high-level fusion. 1) The low-level fusion is a fusion of the visible and infrared images that can be used, for instance, to highlight particular properties of the scene before the detection step [8][9][10]. 2) The high-level fusion is the fusion of the detections obtained from different detectors. Xu et al recently proposed a fusion framework to add as many detectors as wanted and increase the detection performance each time [11].

In this paper, we propose another fusion approach: a mid-level fusion approach based on detection scores which is fast and flexible.

D. Content of the paper

In this work, a far-infrared human detector and a visible human detector work together in a collaborative manner to adapt the detection whatever the moment of the day and whatever the weather conditions. With our approach, 24 hours a day adaptive detection is made possible. Our mid-level fusion approach is more flexible than low-level fusion approaches and we showed that it is faster than high-level fusion approaches.

In this paper, the two detectors are: the top-performing ACF detector of Dollár et al [3] and the adapted ACF detector of Brehar et al [7]. For the purpose of this work, we built an heterogeneous steroscopic system composed of a far-infrared and a visible cameras.
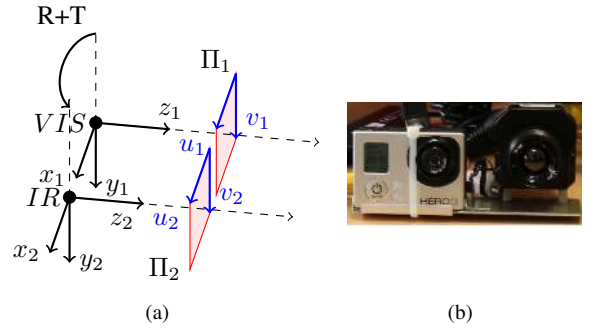

Fig. 2. a) the layout of our stereo acquisition system. VIS and IR are respectively the visible and the infrared camera, Π_1 and Π_2 are the image planes of respectively camera VIS and IR. R: rotation matrix from camera VIS to camera IR, T: translation matrix from camera VIS to camera IR. b) picture of our stereoscopic system.

Section II describes the design of our stereoscopic system and the methods we used to synchronize the frames, Section III describes in details the ACF detector of Dollár et al, Section IV describes our fast and adaptive collaborative approach of detection and Section V is about the tests and results of our approach.

II. HARDWARE

In order to align visible and far-infrared images, we have built an heterogeneous stereoscopic system composed of a GoPro camera and a Flir tau 2 camera. The both cameras are arranged as illustrated in Fig.2.

In order to capture human patterns in the same time and at the same places in both images it is necessary to temporally and spatially synchronize the frames of the two cameras.

1) Temporal synchronisation: During the grabbing, there might be a slight temporal shift between the visible and the far-infrared images. This shift might be due to a difference of acquisition frequencies between the two cameras. In order to fix this problem, two approaches are generally considered: 1) driving one camera's acquisition using the Trigger output of the other camera or 2) using a reference top to fix the shift afterwards, regardless of what caused the shift.

In our case, we have obtained an almost perfect temporal synchronisation by using two identical acquisition pipeline: the two video streams are converted from analogical signals to digital signals using two grabbers with the same specifications. With this approach, our maximum temporal shift was one frame.

2) Spatial synchronisation: According to Krotosky et al, there are four different approaches of spatial synchronisation approaches of parallel-axis stereoscopic systems [12]. These four approaches are: i) global image registration, ii) stereo geometric registration, iii) partial image ROI registration and iv) infinite homography registration. (i) With the global image registration approach, the synchronization of the objects of the scene can only be done at a certain depth of the scene. (ii) With stereo geometric registration, a third information: the depth, is used to geometrically register objects of the scene depending on their depth in the scene. (iii) With partial image ROI resgitration, specific regions of interest of the scene are matched

and local geometric transformations are applied to synchronize the objects. (iv) With infinite homography registration, objets located at the infinite of the stereoscopic system are registered using the assumption that objects perfectly overlap each other at infinity.

Fig. 3. Examples of synchronized pairs of visible and far-infrared images using the infinite homography registration. People close to the stereoscopic system are not well synchronized, people far from it are well synchronized.

Our choice fell on the 4th approach, that is: the infinite homography registration (Fig.3). We chose this approach for three reasons: in people detection applications we cannot make assumptions about the position of the persons in the scene, we did not want to use any extra material to get the depth cue and we did not want to choose a computationally intensive approach.

The infinite homography transformation requires to make the assumption that people are located at infinite of the stereoscopic system. To ensure that, the baseline between the two cameras has to be much smaller than the distance from the stereoscopic system to the people of the scene [12].

Concretely, several parameters are required to compute the infinite homography: the rotation matrix R from camera VIS to camera IR and the matrices of intrinsic parameters K_1 and K_2 of the both cameras.

The infinite homography matrix is defined as follows:

$$H_\infty = K_2 \times R \times K_1^{-1} \tag{1}$$

And every pixel of camera VIS is projected onto the image plane of camera IR using the following mathematical relation:

$$\begin{pmatrix} u_2 \\ v_2 \\ w_2 \end{pmatrix} = H_\infty \times \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} \tag{2}$$

We obtainted the matrices of intrinsic parameters K_1 and K_2 and the rotation matrix R using the Bouguet's Matlab toolbox [13].

III. THE BASIS DETECTOR : THE AGGREGATE CHANNEL FEATURES DETECTOR

The original ACF detector of Dollár et al [3] extracts ten different channels from a pre-filtered image: the three LUV color channels, the normalized gradient magnitude channel and 6 channels corresponding to the bins of the histogram of oriented gradients. Once the channels are extracted, they are divided into 4×4 adjacent pixel blocks. Each pixel block is aggregated to form aggregated versions of the channels. And to finish, the aggregated channels are filtered [3]. At training and during the detection, visual features are simply pixel lookups in the 10 filtered aggregated channels of the image.

Brehar et al proposed an infrared version of the ACF detector [7]. Instead of using 10 channels they use only 8 channels by simply substituting the three LUV channels by the grayscale channel. In the rest of the article, we will name this detector the IR-ACF detector. Note that in our work we do not use the search space reduction approach also proposed by the same authors to work jointly with the IR-ACF detector [7]

Two boosted classifiers are trained for our collaborative detection: the first one for the ACF detector of Dollár et al (for the visible camera) and the second for the IR-ACF detector of Brehar et al (for the infrared camera). A serial combination of 2048 depth 2 decision tree weak-classifiers is trained for each derecrtor. This is done by looking for the pixels in the filtered aggregated channels that best separate positive samples (persons) and negative samples (the rest) samples in the training data [3]. At detection, the weak-classifiers are evaluated in a soft-cascade manner: a detection occurs if all the weak-classifiers are "passed", otherwise no detection occurs. A weak-classifier is passed if its boosting result (added to previous weak-classifier boosting results) is superior to a trace, called the rejection trace [2].

IV. COLLABORATIVE DETECTION

In this paper, we consider the problem of people detection in both the visible and the far-infrared spectrum as a bi-objective problem. This will allow our detection approach to adapt itself with respect to the environmental conditions. In order to solve such a bi-objective problem we have to find the best fused infrared/visible scores of the search space. This can be done using the so-called Pareto dominance critera. Fast exploration of bi-objective solution spaces is possible using the Multiple Objective Particle Swarm Optimization algorithm (MOPSO) of Coello et al [14]. However it does not adapt its solutions according to the environmental conditions, that is: it does not take into account that one detector might be failing when searching for the best fused infrared/visible scores. In this paper we propose an approach to solve this and thus permit dynamic adaptation.

The rest of the section is divided as follows: the first sub-section is about the optimization of mono-objective problems using PSO, the second section concerns the optimization of bi-objective problems using MOPSO and the last subsection is about our approach of detection: the Multiple Modalities (M2D) approach.

A. Mono-objective problem

Mono-objective problems that cannot be optimized analytically can be optimized using meta-heuristic optimization algorithms; searching for people in visible spectrum images is one of these problems. Amongst these optimization algorithms the Particle Swarm Optimization algorithm is a very popular and efficient meta-heuristic algorithm where the optimization is performed by using a swarm of particles "flying" together in the search space [15]. Each particle $\vec{p_k} \in P$ (P is the particle swarm) is moved in the search space according to both a global and a local behaviour at each iteration. The global and local behaviours are, respectively, controlled using the social parameter s and the cognitive parameter c. Other parameters are r which is a random value between 0 and 1, ω as the inertia parameter, $\vec{b_k}$ as the personal best position of the particle k, $\vec{v_k}$ as the velocity of the particle k, and $\vec{g}$ as the global best position (Equ.3 and Equ.4). A 64x128 detection window is centered on each particle $\vec{p_k} = (u, v, scale) \subset S$ (S is the search space).

$$\vec{v_k} = \omega \vec{v_k} + cr(\vec{b_k} - \vec{p_k}) + sr(\vec{g} - \vec{p_k}) \quad (3)$$

$$\vec{p_k} = \vec{p_k} + \vec{v_k} \quad (4)$$

B. Bi-objective problem

Let's first define f_1 as the visible detector and f_2 as the infrared detector. $\forall d \in \{1, 2\}$, $f_d(\vec{p_k})$ is the response of the detector f_d for the detection window centered on $\vec{p_k}$. The simultaneous optimization of the responses of two detectors is a multi-objective optimization problem. It is the optimization of:

$$\vec{f}(\vec{p_k}) = \{f_1(\vec{p_k}), f_2(\vec{p_k})\} \quad (5)$$

Here, contrary to the mono-objective optimization problem, a different optimization concept has to be considered: the goal is to find trade-off solutions rather than a unique best solution. Indeed, improving the response of one detector may come at the expense of the response of the other detector. In this context the Pareto dominance is extensively used to find the trade-off solutions [14]: $\vec{p_1}$ is said to be Pareto dominated by $\vec{p_2}$ (also written as $\vec{p_2} \succeq \vec{p_1}$) if $\forall d \in \{1, 2\}$ $f_d(\vec{p_1})$ is worse than, or equal to $f_d(\vec{p_2})$ and if $\exists d' \in \{1, 2\}$ such that $f_{d'}(\vec{p_1})$ is worse than $f_{d'}(\vec{p_2})$. All the trade-off solutions are part of the Pareto optimal set P^* (Equ.6). The Pareto front (PF^*, Equ.7) is the evaluation of the Pareto optimal set in the bi-dimensional objective space (blue lines in Fig.4). Note that, with this approach, the problem can be extented to more than two detectors. In the rest of the article the trade-off solutions will be analyzed to find good detection candidates.

$$P^* = \{\vec{p_k} \in P / \nexists \vec{p_{k'}} \in P, \vec{p_{k'}} \succeq \vec{p_k}\} \quad (6)$$

$$PF^* = \{\vec{f}(\vec{p_k}) / \vec{p_k} \in P^*\} \quad (7)$$

C. Our approach: the M2D

The pseudocode of the M2D approach is showed in Alg.1 and 2 with the Pareto-dominance being used in both the algorithms (at line 3 for Alg.2). The M2D uses a specific convergence stopping criteria: the number of survivals which are combined to a local contraction (ctr_k, lines 4 and 6 in Alg.2). A non-dominated particle is said to have survived once if it is still present in P^* after one iteration. A non-dominated particle is said to have committed a local contraction of the swarm if another non-dominated particle appears in its close neigborhood after one iteration (at first, the neigborhood is the whole search space). The general idea is to find non-dominated particles having both survived and aided in contracting the swarm a sufficient number of times (more than $max\_contractions$, line 6 in Alg.2). These "candidate detections" are checked: the percentages of passed weak-classifiers for f_1 and f_2 are compared to the minimum required percentages of passed weak-classifiers according to σ_{ref} (line 7, Alg.2). A candidate detection $\vec{p_k}$ is considered as a detection $\in D$ if both the percentages of passed weak-classifiers are above the minimums given by the curves in Fig.5.

D. Dynamic adaptation (σ_{ref})

The dynamic adaptation of the detection is made possible thanks to the σ_{ref} value (Equ.10) which is used as a reference for each image. It is updated at each iteration of the M2D main algorithm (line 11 in Alg.1) and quickly converges towards a stable value after few iterations (σ_{ref} is inspired by the work of Mostaghim et al. [16]). When both the detectors have equitable responses, M_1 and M_2 are close to one another (Equ.8 and Equ.9) : $m_1(\vec{p_k})$ and $m_2(\vec{p_k})$ are the numbers of passed weak-classifiers on $\vec{p_k}$ for respectively f_1 and f_2. M_1 and M_2 are the biggest numbers of passed weak-classifiers found for respectively f_1 and f_2. Therefore, in this case, σ_{ref} is close to zero (Fig.4.a). When one detector is not working as good as the other (M_1 and M_2 very different), most of the particles are gathered near the axis of the better working
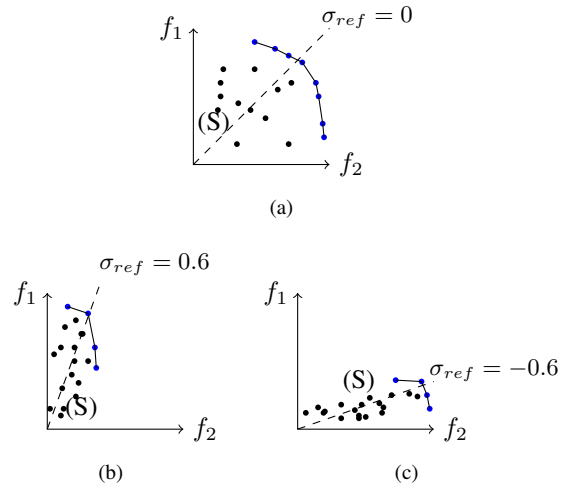

Fig. 4. Pareto dominated particles are black, Pareto non-dominated particles are blue (PF^*). (a) the two detectors contribute equitably, $\sigma_{ref} = 0$. (b) f_1 contributes more to the detection process than f_2, $\sigma_{ref} > 0$. (c) f_2 contributes more to the detection process than f_1, $\sigma_{ref} < 0$.

Data: $\vec{f}$ and search space (S)
Result: Detections (D)

```
 1  iteration = 0;
 2  random initialization of the particles;
 3  compute P*;
 4  while iteration < max_iterations do
 5      for each p⃗_k ∈ P do
 6          g⃗ = closest non-dominated particle;
 7          update v⃗_k and p⃗_k (Equ.3 and Equ.4);
 8          compute f⃗(p⃗_k) in S (update M_1 and M_2); update b_k
            with Pareto-dominance;
 9      end
10      Algorithm 2 (if returns true, go to line 2);
11      update σ_ref;
12      iteration = iteration + 1;
13  end
```

Algorithm 1: Main M2D algorithm.

```
 1  update P*;
 2  for each p⃗_k ∈ P* do
 3      if p⃗_k survived and contracts then
 4          ctr_k = ctr_k + 1;
 5      end
 6      if ctr_k ≥ max_contractions then
 7          if min contributions ok then
 8              D = D ∪ {p⃗_k};
 9              S = S − {area(p⃗_k)};
10              iteration = 0;
11          end
12          return true;
13      end
14  end
15  return false;
```

Algorithm 2: Detection algorithm.

detector (Fig.4.b and Fig.4.c). Thus, the equitable separation line (S) whose slope depends on σ_{ref} is closer to the axis of the better working detector (Fig.4.b and Fig.4.c). The relevance of M_1 and M_2 for indicating the detectors' abilities is related to the initial randomness of the particles' locations in the search space. The σ_{ref} value is used to modulate the required number of passed weak-classifiers for each detector in order to obtain a detection (Fig.5). This permits 1) looking for complementary visible/infrared results and 2) dynamically managing the sudden failure of one detector or the sudden ineffectiveness of one sensor.

$$M_1 = \max_{\vec{p_k} \in P}(m_1(\vec{p_k})) \qquad (8)$$

$$M_2 = \max_{\vec{p_k} \in P}(m_2(\vec{p_k})) \qquad (9)$$

$$\sigma_{ref} = \frac{M_1^2 - M_2^2}{M_1^2 + M_2^2} \qquad (10)$$

V. TESTS AND RESULTS

The M2D approach has been tested on three scenarios (Fig.7): 1) with both cameras working, 2) with only the infrared camera working and 3) with only the visible camera working.
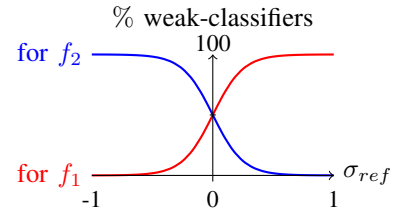
Fig. 5. In red and blue: sigmoidal curves representing the minimum required percentages of passed weak-classifiers with respect to σ_{ref}. If f_1 is ineffective ($\sigma_{ref} = -1$) then all the weak-classifiers of f_2 have to pass, if f_2 is ineffective ($\sigma_{ref} = 1$) all the weak-classifiers of f_1 have to pass.
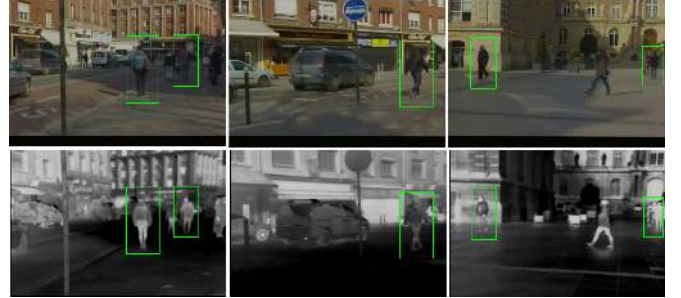
Fig. 6. Examples of detection results obtained on the synchronized pairs of far-infrared and visible images used during the tests.

The global performance in case (1) is a trade-off between the global performance of cases (2) and (3): the sensitivity is increased when compared to case (3) with the MR being reduced but the FPPI slightly increasing. This is due to the addition of the infrared sensor (green curve). This trade-off is made to the benefit of a fast collaborative (and dynamic) approach of detection (Tab.II). Note that, as the M2D is a non-deterministic algorithm we have run multiple tests for each case and took the mean results to draw the ROC curves of Fig.7.

Case	VIS ok IR ok	VIS dys IR ok	VIS ok IR dys
MR StdDev	0.014	0.012	0.004
FPPI StdDev	0.007	0.005	0.182

TABLE I. AVERAGE MISS RATE AND FFPI STANDARD DEVIATIONS FOR THE THREE CASES.

The average Miss Rate and FPPI standard deviations of the multiple tests show that our non-determinisic approach of detection is relatively stable (Tab.I). In the third case (VIS ok, IR dys), the average MR standard deviation is smaller and the average FPPI standard deviation is bigger: it is more sensitive when only the visible detector is considered.

Fig.8 shows how the detection performance of the M2D approach change with respect to the concentration of particles in the multimodal image pyramid. We can see that above 48×10^{-6} particles per square pixel the performance are similar: increasing the concentration slows down the detection and does not significantly improve the performance. Thus, having 48×10^{-6} particles per square pixel is ideal.

Detecting with the M2D is 1/0.58 times faster than detecting using directly the ACF and the IRACF detectors (ACF+IRACF approach). Besides, the ACF+IRACF approach
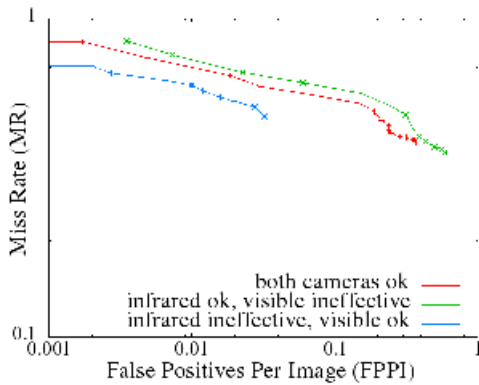
Fig. 7. M2D global performance for three scenarios: 1) both detectors working correctly, 2) visible detector ineffective and 3) infrared detector ineffective.
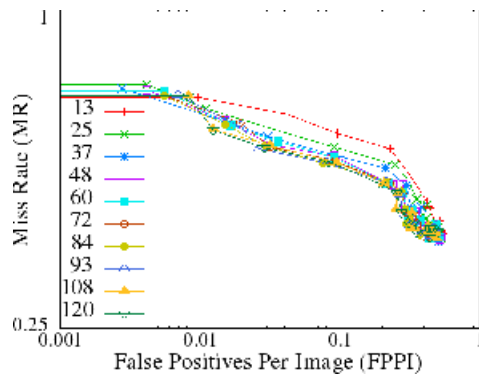

Fig. 8. Detection performance of the M2D approach with respect to the concentration of particles in the search space (in μ particles per square pixel).

Time ratios	ACF	IRACF	ACF+IRACF	M2D
ACF	1	0.96	1.96	**1.14**
ACF+IRACF	0.511	0.49	1	**0.58**

TABLE II. SCANNING TIME RATIOS FOR FOUR DETECTION APPROACHES WITH RESPECT TO THE ACF DETECTOR AND THE ACF+IRACF APPROACH.

requires extra computation for fusioning the infrared and the visible detections at the end. The tests have been performed on the AVIS[1] dataset, 100 particles spread on 16 images levels were used, $max\_iterations$ has been set to 300 and $max\_contractions$ has been set to 5. Fig.6 shows some images of the AVIS dataset, which contains 316 pairs of synchronized far-infrared and visible images taken in streets. Note that our AVIS dataset contains more challenging syncronized cases than the well known OTCBVS dataset [17].

VI. CONCLUSION

In this paper, we proposed a collaborative detection approach using an infrared spectrum human detector and a visible spectrum human detector together to extend the detection capabilities 24 hours a day and whatever the weather conditions. We presented a detection approach using two detectors that dynamically adapts itself in case one detector is suddenly failing or one camera is suddenly becoming ineffective. Besides, we also showed that our approach is fast knowing that two detectors are used in the same time: it is almost as fast as using one detector alone.

REFERENCES

[1] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Conference on Computer Vision and Pattern Recognition*, 2005.

[2] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral Channel Features," in *Proceedings of the British Machine Vision Conference*, 2009, pp. 91.1–91.11.

[3] P. Dollár, R. Appel, S. Belongie, and P. Perona, "Fast Feature Pyramids for Object Detection," *Transactions on pattern analysis and machine intelligence (TPAMI)*, 2014.

[4] L. Zhang, B. Wu, R. Nevatia, I. Systems, and L. Angeles, "Pedestrian Detection in Infrared Images based on Local Shape Features," in *Computer Vision and Pattern Recognition (CVPR)*, 2007.

[5] D. Olmeda and J. M. Armingol, "Contrast Invariant Features for Human Detection in Far Infrared Images," in *Intelligent Vehicles Symposium (IV)*, 2012.

[6] D. Olmeda, "Pedestrian detection in far infrared images," Ph.D. dissertation, 2013.

[7] R. Brehar, C. Vancea, and S. Nedevschi, "Pedestrian Detection in Infrared Images Using Aggregated Channel Features," in *Intelligent Computer Communication and Processing (ICCP)*, 2014.

[8] L. Jiang, F. Tian, L. E. Shen, S. Wu, S. Yao, Z. Lu, and L. Xu, "Perceptual-based fusion of ir and visual images for human detection," in *Intelligent Multimedia, Video and Speech Processing*, 2004.

[9] E. T. Gilmore, P. D. Frazier, and M. Chouikha, "Improved human detection using image fusion," in *IEEE Conference on Robotics and Automation*, 2009.

[10] J. Serano-Cuerda, "Robust human detection through fusion of color and infrared video," Ph.D. dissertation, 2014.

[11] P. Xu, F. Davoine, and T. Denoeux, "Evidential combination of pedestrian detectors," in *British Machine Vision Conference (BMVC)*, 2014.

[12] S. J. Krotosky and M. M. Trivedi, "Mutual information based registration of multimodal stereo videos for person tracking," *Computer Vision and Image Understanding (CVIU)*, 2007.

[13] Bouguet, "Jean-yves bouguet's matlab toolbox for calibrating cameras," http://www.vision.caltech.edu/bouguetj/calib_doc/.

[14] C. Coello, G. Pulido, and M. Lechuga, "Handling multiple objectives with particle swarm optimization," *Transactions on Evolutionary Computation (EC)*, 2004.

[15] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *International Conference on Neural Networks (ICNN)*, 1995.

[16] S. Mostaghim and J. Teich, "Strategies for finding good local guides in multi-objective particle swarm optimization (MOPSO)," in *Proceedings of Swarm Intelligence Symposium (SIS)*, 2003.

[17] J. Davis and M. Keck, "A two-stage approach to person detection in thermal imagery," in *Proc. Workshop on Applications of Computer Vision*, 2005.

[1]http://mis.u-picardie.fr/∼p-blondel/papers/data