Combination of Colour and Thermal Sensors for Enhanced Object Detection

Louis St-Laurent, Xavier Maldague
Department of Electrical and Computer Engineering
Université Laval
Quebec city, Canada
louis.st-laurent@ino.ca, maldagx@gel.ulaval.ca

Donald Prévost
Advanced Imaging Systems
National Optics Institute
Quebec city, Canada
donald.prevost@ino.ca

Abstract − *In uncontrolled environments, with dynamic background and lighting changes, performing efficient and real-time foreground − background segmentation is very challenging. This work is based on the hypothesis that the combination of long wave infrared (LWIR) (8-12 µm) and colour cameras can significantly improve the robustness of moving objects extraction. Pros and cons of colour and thermal imagers in outdoor video monitoring applications are discussed. In order to fuse information from both sensors, we favoured an approach based on "analytical fusion" rather than "representative fusion". Starting from a state-of-the-art algorithm for moving objects extraction in colour video (non-parametric codebook model [1]), we first adapted the method for processing of "Red-Green-Blue-Thermal" video format. A preliminary objective performance evaluation of detection accuracy is presented. Original image sequences grabbed with co-aligned thermal and visible fields of view was used. Finally, some improvements to the original codebook model are proposed.*

Keywords: Video monitoring, Object detection, LWIR-colour image fusion, Codebook model.

1 Introduction

Over the last decade, we saw the appearance of automated video monitoring in numerous applications. To make apart foreground and background pixels is the first step of most automated video monitoring systems aiming at object tracking, event detection or scene interpretation. Still today, this basic task is challenging when the monitoring takes place in uncontrolled environments.

Most solutions proposed for moving object extraction are based on the visible spectrum. In brightly illuminated scenes, standard colour cameras provide the best information for object segmentation. However, in outdoor applications, darkness and other environmental conditions such as fog, rain and smoke strongly decrease the efficiency of standard cameras. In many applications, achievement of zero miss detection rate is a critical requirement and investment in more powerful imaging systems is justified. This opens the way to video systems combining thermal and colour cameras. This work is based on the hypothesis that the addition of LWIR cameras (8-12 µm) can significantly improve the robustness of foreground − background segmentation in uncontrolled environments.

Only recently, thermal imagery has been exploited in video monitoring applications, and especially for pedestrians detection [2][3][4][5]. Usually, the fact that human skin appears brighter than the background in long-wave thermal imagery is used as the main cue for the detection. However, this constraint is not always satisfied, especially in outdoor scenes. Resorting to others features, like edges or blobs characteristics, is often performed.

As widely known in the field of image fusion, the combination of thermal and visible images is not trivial. Many research groups previously addressed the fusion of thermal and colour images in order to generate a new image more informative or intuitive for a human observer. As proposed in [6], this kind of processing can be classified as *representative* image fusion. It is important to understand that the generation of such an image is not required for automated video monitoring applications. For this reason, *analytical* fusion is more suitable for automated video monitoring. It consists in combining available information from both sensors for a more robust analysis and interpretation of video content.

A few recent works exploit analytical fusion. In [7], objects of interest are first extracted from infrared images along the hypothesis that pedestrians are warmer than background. Selected regions of interest are then used in both spectrums for contour extraction and fusion. In [8], moving objects are detected and tracked independently in each spectrum. An analysis of object's temporal persistence is used at every frame to select the more reliable sensor. Similarly, the fusion scheme proposed in [9] is based on sensor reliability. Detection and targets tracking tasks are performed independently for visible and thermal images and a confidence measure is employed to weight sensor data at every time instant.

The work presented in this paper has the distinctive characteristic of combining information from both sensors at pixel level. More specifically, every pixel is classified as foreground or background along its similarity to thermal-colour background model. To provide the

required image registration, we used an acquisition unit with both fields of view aligned at hardware level. As a starting point for background modelling, we selected a state-of-the-art technique developed for colour videos [1]. Our first contribution lies in the adaptation of their algorithm for the processing of augmented *"Red-Green-Blue-Thermal"* (*RGBT*) video format. Secondly, we propose some improvements to their background model in order to improve processing time and detection accuracy.

In the next section, we first discuss the pros and cons of colour and thermal imagers in outdoor video monitoring applications. In section 3, we present the codebook background subtraction algorithm in its original form as proposed by [1]. Then, we describe how we integrated thermal information in the process. For performance evaluation, preliminary quantitative assessments were achieved in order to measure the accuracy improvement obtained from the combination of LWIR and visible spectrums. This is discussed in section 4. Finally, improvements to the codebook background model are proposed in section 5.

2   Imaging in uncontrolled environment

In this work, the term "uncontrolled environment" is employed to refer to outdoor scenes where illumination and temperature changes occur frequently, and where various atmospheric conditions can be observed. Specific environmental condition will not bear an equivalent impact on thermal and visible imaging because the properties of radiation propagation in the atmosphere vary greatly with wavelength range.

It is worth to mention that the LWIR spectral range (8-12 μm) is more suitable than the MWIR range (3-5 μm) for video monitoring of human activity. The main reason is that emitted radiation from objects at ambient temperature (300 K) peaks in this long-wavelength range [10]. More precisely, a human body, at 37°C, have a dominant wavelength of 9.3 μm.

2.1   Luminosity and temperature changes

Obviously, visible cameras are more sensitive than LWIR ones to illumination changes. However, since the sunlight warms up exposed surfaces, long-term illumination changes also affect thermal images. And sunlight does not equally warm up all objects of the scene. Those with high thermal inertia, like water, trees and people, will present little changes comparatively to cars or building, which have low thermal inertia. Moreover, dark surfaces with low reflective properties, as asphalt, will be more strongly affected by sun's rays than brighter objects. It is why a scene observed at 4PM and 4AM appears so differently. In the first case, asphalt and other inanimate objects have been heated all day long and appear warmer than people.

2.2   Atmospheric conditions

A first factor that has significant impact on contrast of thermal images is the wind, which accelerates warm objects cool down. Hence, the background will tend to have a more uniform temperature and less contrast during strong wind days. This is demonstrated by comparing images of top and middle rows of figure 1.

Fig. 1: Visible and LWIR images during slight overcast (top), strong wind and rain (middle) and snow (bottom).

Fog, rain and snow are others climate factors that affect thermal wavelength transmission, thus decreasing thermal image contrast. However, along some studies [11][12], and based on our own experience, absorption and scattering in LWIR range seem slightly inferior than in visible range. This can be noticed on last row images of figure 1, where a building can be distinguished in the background of the thermal image, but not in the colour one. Figure 2 also illustrates this fact. Airplanes, vehicles and ground installations can be seen in the LWIR image (right), but it is almost impossible to recognize any object in the visible image (left).

Fig. 2: Visible (left) and LWIR (right) images in thick fog (courtesy of Ottawa airport and Searidge Technologies, Ottawa, Canada).

3 Objects extraction

Since moving objects are usually the objects of interest in video monitoring applications, motion-based segmentation techniques are the most suitable for objects extraction. In this work, a still acquisition platform is assumed, thus no camera motion compensation is needed. For this situation, the most efficient objects extraction methods, both in term of accuracy and processing time, are those based on background model subtraction.

3.1 State-of-the-art object extraction methods in colour video

In the last decade, many efficient background modelling and maintenance approaches have been proposed for videos from the visible spectral range. Two of them have brought most attention and have been selected and improved by numerous research groups. The first is the Gaussian Mixture Model of [13] and the second is the non-parametric kernel-based density estimator of [14]. Their interest comes from the fact that they allow modelling of multi-modal background (moving background), like those met in outdoor scenes. They also support the presence of foreground objects in the scene during the background model initialization phase. Recently, a new non-parametric technique inspired from [14] but based on codebook model [1] has been presented. This one is particularly well suited for real-time applications. Indeed, they demonstrated in their performance analysis that this new method is about three times faster than [13] and [14], with similar segmentation accuracy. It is this algorithm that we selected as a starting point for our "*RGBT*" objects extraction technique.

In short, a quantization/clustering technique is used to generate a compressed form of background model. Every pixel is represented by L codewords defined by :

$$CW_{k=1...L} = \left\{ \bar{R}_k, \bar{G}_k, \bar{B}_k, I_k^{\max}, I_k^{\min}, f_k, p_k, q_k, MNRL_k \right\} \quad (1)$$

$\bar{R}_k, \bar{G}_k$, and $\bar{B}_k$ are red, green and blue values, $I_k^{\max}$ and $I_k^{\min}$ are maximum and minimum observed brightness of codeword k, and f, p and q are respectively the number of matches, the time stamp of the first match, and the time stamp of the last match. The *MNRL* (*Maximum Negative Run-Length*) is the key parameter. It stores the length of the period (in number of frames) during which a codeword has not been matched. Based on the hypothesis that background pixels are observed periodically, codewords belonging to background will have a small *MNRL* value while codewords generated by the temporary presence of an object of interest will have large *MNRL*. Thus, a threshold on *MNRL* parameter can be used to filter out codewords belonging to objects of interest

For every new frame, every pixel is associated to the first sufficiently similar codeword in regard to its colour and brightness. If no codeword can be matched, a new one is created and added in a cache codebook. Codewords from the cache are promoted to permanent background codebook when they are repetitively matched, and codewords from permanent background codebook not matched since a long period of time are deleted. Typically, mono-modal scene areas will be modelled by only one codeword, while multi-modal areas, like swaying trees, will need more codewords. In our implementation, we limited the permanent background codebook to a maximum of 10 codewords per pixel.

The background model maintenance is achieved by updating red, green and blue values of the matched codeword via a weighted average. Here is how the red value of pixel x is determined for frame $t+1$:

$$\bar{R}_{k,t+1}(x) = \frac{\alpha * \bar{R}_{k,t}(x) + R_t(x)}{\alpha + 1} \quad (2)$$

The α coefficient defines the learning rate. A smaller α will bring a faster codeword adaptation, thus leading to a lower rate of false detection when quick illumination changes occur. For more details on specific aspects of the algorithm, as how colour thresholds are defined, please see the original description in [1].

3.2 Integration of thermal information

Thanks to the alignment of thermal and colour images provided by our acquisition platform, we can directly combine information at pixel level. More specifically, we added a parameter $\bar{T}_k$, for thermal intensity, in the codeword representation of every pixel (Eq. (1)). Like for colour parameters, thermal intensity of matched codewords is updated by the weighted average of Eq. (2).

In order to associate a *RGBT* pixel to codeword k, a condition on its thermal intensity must also be satisfied :

$$\left| T(x) - \bar{T}_k \right| \leq 5 * \sigma_T \quad (3)$$

where σ_T is the global standard deviation observed on 10 consecutive thermal frames. This standard deviation is refreshed periodically (every 2 minutes in our implementation). Contrarily to colour images, we determine gain and offset adjustment for thermal images at the beginning of each sequence. Thus, a higher gain is applied to thermal images of videos grabbed from scenes with poor thermal contrast, leading to higher noise. It is why we decided to relate the thermal intensity threshold to standard deviation.

With our combination of thermal and colour conditions, a pixel is classified as foreground if either its colour **OR** thermal value differs from background codewords. The detection mask obtained leads to a minimum miss detection rate, which is generally a very important requirement for video monitoring applications.

4 Performance analysis

4.1 Accuracy analysis

In this preliminary assessment of our combination approach, we compared the detection accuracy with the one achieved if only visible or thermal information is exploited. For this purpose, we used the *Detection Rate* (*DR)* and *False Alarm Rate* (*FAR*) metrics proposed in [15] :

$$DR = \frac{TP}{TP + FN} \qquad (4)$$

$$FAR = \frac{FP}{TP + FP} \qquad (5)$$

where *TP*, *FN* and *FP* hold for True Positive, False Negative and False Positive detections. For example, a pixel wrongly classified as foreground (object of interest) is a False Positive detection.

Obviously, these metrics require the availability of references images, commonly referred as *Ground Truth*, in which the correct foreground – background classification is defined. Manual generation of these reference images is a long and laborious task. Thus, the accuracy assessment as been performed on a limited number of four video sequences, and on a sample of 10 regularly spaced frames in each of these videos.

Test image sequences have been grabbed with our co-aligned LWIR / colour acquisition platform. We use a glass beamsplitter with ITO coating (Indium-Tin-Oxide). This component reflects 93% of 8-12 μm emitted energy and transmits about 85% of visible waves. The LWIR camera used is the ThermoVision A10, which has a VO2-based microbolometer detector array of 164 x 128 pixels and a sensitivity of about 80 mK. The Marlin F33C CCD camera (640 x 480 pixels with Bayer filter) was used as colour sensor.

To compensate for the distinct cameras resolutions and for the slightly different sensors form factors, the colour image is decimated and the thermal image is interpolated to produce resulting RGBT images of 328 x 254 pixels. The bilinear interpolation of the thermal image also acts as high frequency filter, thus removing part of the image noise. Figure 3 illustrates the achieved quality of the registration of thermal and colour images. For visualization purposes, the red channel of the colour image as been replaced by the scaled thermal image.

It is worth to mention that such a hardware registration of thermal and colour images provides, for some applications, a major advantage over systems with parallel or convergent optical axis. The benefit comes from the fact that image registration is valid for a wide depth of field.
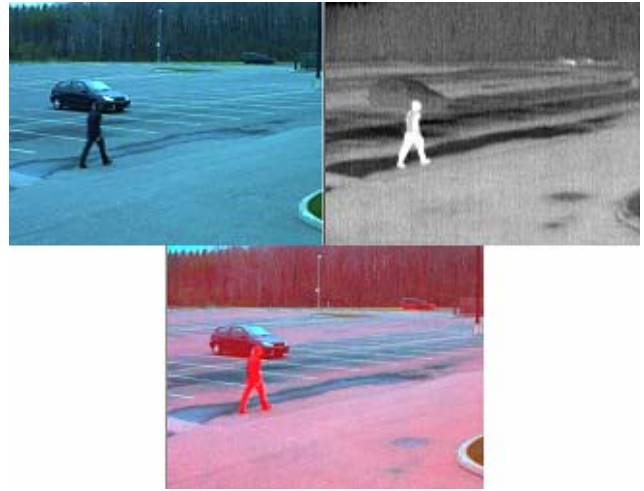

Fig. 3: Visualization of image registration.

Test sequences have been selected in order to represent different observation conditions. They are illustrated in figure 4, with examples of *Ground Truth* and segmentation results in figure 5.

Table 1 presents cumulated *Detection Rate* and *False Alarm Rate*. For all test sequences, we empirically fixed the learning rate α to 7. It is important to understand that since only the matched codewords are updated, a fast learning rate can be used without causing early integration of motionless objects in the background model. We also fixed the training period to 200 frames, which represent about 25 seconds with our acquisition rate of 7.5 fps.

Table 1. Comparison of segmentation accuracy.

Video format	Video *A*		Video *B*		Video *C*		Video *D*	
	DR	*FAR*	*DR*	*FAR*	*DR*	*FAR*	*DR*	*FAR*
RGB	0.82	0.07	0.88	0.78	0.66	0.26	0.94	0.38
Thermal	0.56	0.15	0.89	0.43	0.84	0.19	0.78	0.48
RGBT	0.91	0.17	0.97	0.83	0.94	0.27	0.99	0.45

From results of table 1, we first notice that, for all videos, *Detection Rate* obtained with the combination of thermal and colour information is higher than with any sensor used alone. About the comparison of detection accuracy between thermal and visible videos, we can note that a better *DR* was achieved with colour images for videos *A* and *D.* In sequence *B*, the only moving objects are pedestrians, which are generally characterized by a good thermal contrast. It is why a high thermal *DR* of 89% was measured with video *B*. For sequence *C*, grabbed at sunset, the thermal sensor outperforms the visible one as predicted because of the low illumination.

Similarly to *DR*, *False Alarm Rate* obtained with *RGBT* video format is higher than the *FAR* measured for each sensor alone. This error accumulation comes from the use of a simple logical OR for the combination of thermal and visible detections.
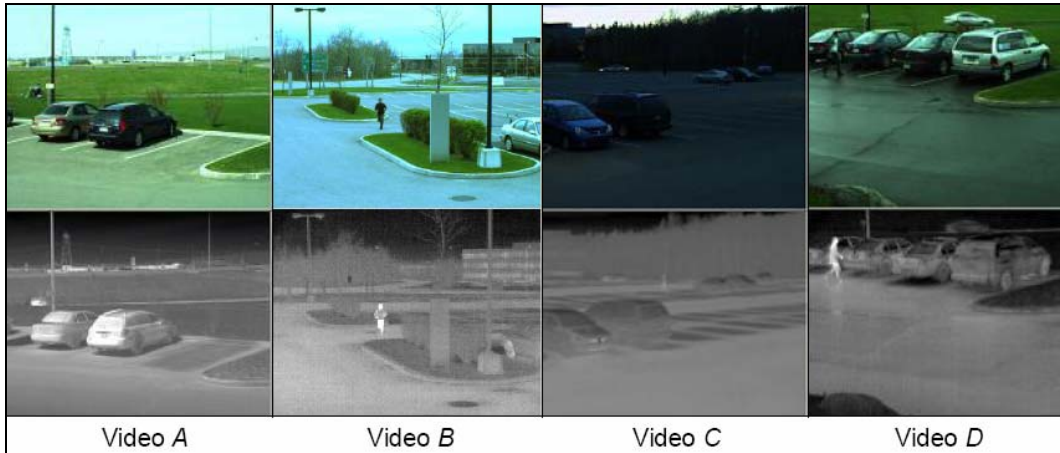
Fig. 4: Test sequences. Video A : sunny afternoon; B : sunrise; C : sunset; D : cloudy and rainy.
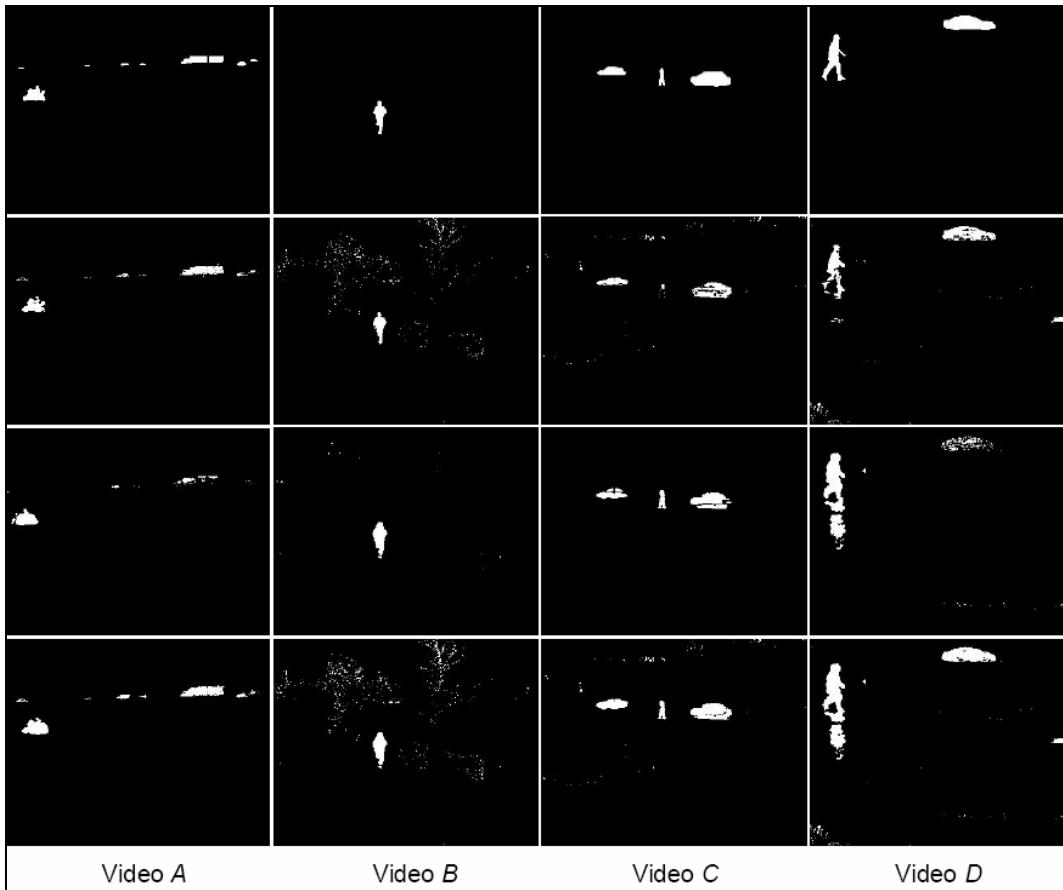

Fig. 5: Examples of *Ground Truth* (first row) and detection results from colour (second row),
thermal (third row), and from RGBT images (fourth row).

A very high *FAR* of 83% was measured for video B. The main reason behind this is that many bushes and trees are present in the background and a quite strong wind was blowing during the acquisition. Moreover, the small size of the pedestrian in video *B* also contributes to this high *FAR*. When objects of interest are very small, the number of *True Positive* detections will be also very small (*TP* of Eq. (5)), thus leading to high *FAR*.

We can also remark that *FAR* measured for video *D*, grabbed during rain, are significantly larger than for video *A*. The visible and thermal reflections on wet surfaces are

responsible of these numerous false detections, which can be clearly observed in detection masks of figure 5.

Another aspect that must be considered in the interpretation of *FAR* is the fact that no image enhancement has been performed on detection masks for results presented in table 1. For comparison purpose, table 2 indicates *DR* and *FAR* measured after morphological closure and elimination of blobs smaller than 30 pixels.

Table 2. Segmentation accuracy after mask enhancement.

Video format	Video *A*		Video *B*		Video *C*		Video *D*	
	DR	*FAR*	*DR*	*FAR*	*DR*	*FAR*	*DR*	*FAR*
RGB	0.81	0.05	0.88	0.47	0.77	0.14	0.91	0.38
Thermal	0.61	0.14	0.92	0.32	0.92	0.19	0.81	0.47
RGBT	0.91	0.16	0.97	0.69	0.97	0.20	0.94	0.45

4.2 Processing time analysis

As mentioned in the introduction, an important requirement for automated video monitoring applications is real-time performance, and the foreground – background segmentation module is generally the processing time bottleneck of that kind of system. In this work, a particular attention has been addressed to this aspect and it is precisely for this reason that the codebook background model has been selected as a starting point for the foreground – background segmentation algorithm instead of a Gaussian Mixture Model or other non parametric approach.

Table 3 shows the mean processing time measured on each test video sequence. The processor used was a Pentium IV 3 GHz, with 1 GB of RAM. Image resolution is 328 x 254 pixels. Results of table 3 do not include detection mask enhancement. Up to 2 ms per frame can be added for the morphological closure and small blobs deletion.

Table 3. Comparison of processing time, in ms per frame.

Video format	Video sequence			
	A	*B*	*C*	*D*
RGB	10.1	12.4	10.2	10.0
Thermal	2.1	2.3	2.0	2.0
RGBT	11.1	13.4	11.0	10.7

We can notice that the processing time of every video is quite similar, except for sequence *B*, which is slightly longer. The high background instability caused by swaying bushes and trees is responsible of this observation. Table 3 also illustrates the interesting fact that the processing time for the *RGBT* format is smaller than the summation of the processing time required for *RGB* and *Thermal* formats individually.

Extending the results to larger images of 640 x 480 pixels would lead to a processing time of about 45 ms per frame. This corresponds to 22 fps, which is still acceptable for real-time video monitoring applications.

5 Improvements to codebook model

In order to improve object detection accuracy and processing time, we propose two significant modifications to the background subtraction approach proposed in [1].

5.1 Selective update of potential codewords

Instead of adding every new observed value in a *cache* codebook [1], we only add pixel values that do not belong to a target. To achieve this, the update of the codebooks (permanent and potential) is performed after the enhancement of the detection mask consisting in morphological closure and elimination of small blobs. In this way, only pixels considered as false detection (pixels removed from detection mask by enhancement operations) are added to potential background codebook.

Moreover, we propose to promote potential codewords re-appearing more than a certain amount of time (T_{add}) directly to permanent codebook rather than in a non-permanent layer. We also periodically refresh (every 2 minutes instead of every frame as in [1]) permanent and potential codebooks by removing codewords having too high *MNRL*.

5.2 *YCbCr* colour space

In order to decrease the required processing time, we perform background modelling and object detection in the {luma – blue chroma – red chroma} colour space:

$$Y = 0.299R + 0.587G + 0.114B$$
$$Cb = 0.56433 \cdot (B - Y) \qquad (6)$$
$$Cr = 0.71327 \cdot (R - Y)$$

In [1], a colour distortion measure based on RGB colour representation is used:

$$\delta\left(\mathbf{X}_t, \mathbf{CW}_k\right) = \sqrt{\left(R^2 + G^2 + B^2\right) + \frac{\overline{R}_k R + \overline{G}_k G + \overline{B}_k B}{\overline{R}_k{}^2 + \overline{G}_k{}^2 + \overline{B}_k{}^2}} \qquad (7)$$

where $\mathbf{X_t}$ is a pixel at frame t and $\mathbf{CW}_k$ is the kth codeword. Thresholds on brightness are defined via coefficients α and β :

$$I_{k,low} = \alpha I_k^{\max} \qquad (8)$$

$$I_{k,hi} = \min\left(\beta I_k^{\max}, \frac{I_k^{\min}}{\alpha} \right) \qquad (9)$$

where I^{max} and I^{min} are maximum and minimum observed brightness values for the kth codeword. Note that Euclidean distance is used in [1] to define brightness:

$$I = \sqrt{R^2 + G^2 + B^2} \qquad (10)$$

In order to match an observed pixel to a background codeword, its colour distortion δ must be inferior to a predefined threshold and its brightness I must be inside the range defined by $I_{k,low}$ and $I_{k,hi}$. With $YCbCr$ colour space, colour and brightness distortions are directly obtained via chroma and luma components:

$$\delta_{chroma} = \max\left(\left|C_b - \overline{C}_{b,k}\right|, \left|C_r - \overline{C}_{r,k}\right|\right) \quad (11)$$

$$\delta_{luma} = Y - \overline{Y}_k \quad (12)$$

Similarly to Eq. (8), we defined brightness thresholds by two coefficients α and β:

$$\begin{aligned} Y_{k,low} &= \alpha Y_k \\ Y_{k,hi} &= Y_k + \beta \end{aligned} \quad (13)$$

As shown in table 4, the use of $YCbCr$ colour space significantly decreases the required processing time for both background initialization and motion detection. Values of table 4 include detection mask enhancement operations and have been measured for colour channel of test video D, which is 328 x 254 pixels. Values for proposed method also include the processing time required for the colour space conversion of every input image (Eq. (6)). Selective background update defined in sub-section 5.1 was used for both codebook models.

Table 4. Processing time, in ms per frame.

Codebook	Initialization	Object detection
Original	10.4	12.3
Proposed	3.9	4.8

Another benefit of our approach is that memory requirement for the codebook model is decreased because it avoids storing I_{low} and I_{hi} parameters for every codeword. Furthermore, the $YCbCr$ colour space has the advantage that MPEG compression works with this format. Since most video monitoring applications are now using the MPEG standard for video transfer, colour space conversion of input images might be avoided.

5.3   Validation of detection accuracy

To compare detection accuracy in $YCbCr$ and RGB, we submitted our modified codebook model to the same objective performance evaluation than in section 4. Only colour channels have been considered in this analysis. Table 5 presents detection rates and false alarm rates measured with every test video. Detection mask enhancement is included in the measure.

We can notice that obtained results are very similar to those of the original approach. Even if a significant improvement is observable for test video B (where a smaller false alarm rate is measured for a similar detection rate), we can not conclude that better detection accuracy is achieved in $YCbCr$ colour space. Indeed, this improvement may be mainly caused by different thresholds on chrominance and brightness distortions, which have been experimentally and independently selected in both colour spaces. A more objective way to fix thresholds would be required in order to perform a very accurate comparison.

Table 5. Comparison of segmentation accuracy with proposed improvements

Codebook	Video A		Video B		Video C		Video D	
	DR	FAR	DR	FAR	DR	FAR	DR	FAR
Original	0.81	0.05	0.88	0.47	0.77	0.14	0.91	0.38
Proposed	0.81	0.05	0.90	0.27	0.79	0.17	0.88	0.35

6   Conclusion

Despite the typically poorer resolution and higher noise level of uncooled thermal cameras comparatively to visible spectrum sensors, the addition of thermal information improves sensitivity of detection of foreground – background segmentation algorithms for outdoor video sequences. The proposed approach is based on a state-of-the-art background model [1], but adapted for $RGBT$ video sequences grabbed with our co-aligned LWIR / colour acquisition platform. We demonstrated that the processing time required for $RGBT$ format is smaller than the summation of the processing time required for RGB and $Thermal$ formats individually.

Two significant modifications to the original codebook model have been proposed. By using a selective update of potential codewords, only pixels considered as false detection are added to potential background codebook. Thus, a codebook with a slightly smaller number of codewords is generated via the background updating process. By using the $YCbCr$ colour space, we reduced by a factor three the processing time for initialization and object detection (background update is included in object detection). We also validated that detection accuracy obtained with our modified codebook model is similar to the original method. Thus, these improvements are suitable for real-time video monitoring applications.

The work presented in this paper is the first step through a real-time object tracking system efficient in outdoor environment all day and all year long. The next improvements related to foreground – background aspects will address minimization of false detections. As mentioned in section 4.1, the implemented algorithm leads to a very good detection rate, but also to a large false alarm rate that could be decreased by a higher level analysis of detection mask properties. Identification of $False\ Positive$ pixels caused by reflections on wet surfaces is an example.

Performance evaluation of a new foreground – background segmentation algorithm is essential. As mentioned in the previous section, manual generation of $Ground\ Truth$ images is a laborious task, and the

particular *RGBT* image format used in our work doesn't allow us to use publicly distributed videos with provided *Ground Truth*. Analysis of others evaluation methods not requiring the availability of references images is planned for our future development phases. This would also enable us to quantitatively evaluate the accuracy detection on a more representative sample volume.

.

References

[1] K. Kim, T.H. Chalidabhongse, D. Harwood and L. Davis, *Real-time foreground - background segmentation using codebook model*, Real-time Imaging, Vol. 11, No. 3, pp. 172-185, 2005.

[2] F. Xu and K. Fujimura, *Pedestrian detection and tracking with night vision*, In Proc. IEEE Intelligent Vehicles Symposium, Paris, France, 2002, Vol. 1, pp. 21-30.

[3] M. Bertozzi, E. Binelle, A. Broggi and M. Del Rose, *Stereo vision-based approaches for pedestrian detection*, IEEE Conf. on Computer Vision and Pattern Recognition - Workshops, San Diego, USA, 2005, Vol. 3.

[4] C. Dai, Y. Zheng and X. Li, *Layer representation for pedestrian detection and tracking in infrared imagery*, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition - Workshops, San Diego, USA, 2005, Vol. 3.

[5] Y. Fang, K. Yamada, Y. Ninomiya, B. Horn and I. Masaki, *Comparison between infrared-image-based and visible-image-based approaches for pedestrian detection*, In Proc. IEEE Intelligent Vehicles Symposium, Columbus, USA, 2003, pp. 505-510.

[6] C. Ó Conaire, *Thermal infrared and visible spectrum fusion for multi-modal video analysis*, Transfer Report, School of Electronic Engineering, Dublin City University, July 2005.

[7] J.W. Davis and V. Sharma, *Fusion-based background-subtraction using contour saliency*, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition - Workshops, San Diego, USA, 2005, Vol. 3.

[8] H. Toressan, B. Turgeon, C. Ibarra-Castanedo, P. Hébert and X. Maldague, *Advanced surveillance systems : Combining video and thermal imagery for pedestrian detection*, SPIE Vol. 5405, Thermosense XXVI, pp. 506-515, 2004.

[9] L. Snidaro, G.L. Foresti, R Niu and P.K. Varshney, *Sensor fusion for video surveillance*, In Proc. of Int. Conf. on Information Fusion, Stockholm, Sweden, 2004, Vol. 2, pp. 739-746.

[10] X. Maldague et al., *Infrared and Thermal Testing, Nondestructive Testing Handbook*, Vol. 3, 3rd edition, 2001.

[11] D.B. Rensch and R.K. Long, *Comparative studies of extinction and backscattering by aerosols, fog, and rain at 10.6μ and 0.63μ*, Applied Optics, Vol. 9, No.7, 1970.

[12] M.A. Naboulsi, H. Sizun and F. de Fornel, *Fog attenuation prediction for optical and infrared waves*, Opt. Eng., Vol. 43, No. 2, pp. 319-329, 2004.

[13] C. Stauffer and E. Grimson. *Adaptive background mixture models for real-time tracking*, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Fort Collins, USA, 1999, Vol. 8, pp. 252-258.

[14] A. Elgammal, R. Duraiswami, D. Harwood and L.S. Davis, *Background and foreground modeling using nonparametric kernel density estimation for visual surveillance*, Proceedings of the IEEE, Vol. 90, No. 7, pp. 1151-1163, 2002.

[15] G. Medioni, *Detecting and tracking moving objects for video surveillance*, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Fort Collins, USA, 1999, Vol. 2, pp. 319-325.