Intelligent Fusion of Infrared and Visible Spectrum for Video Surveillance Application

Rania Rebai Boukhriss[1(✉)], Emna Fendri[2], and Mohamed Hammami[2]

[1] Miracl-ISIMS, Sfax University, Sakiet Ezzit, Sfax, Tunisia
[2] Miracl-FS, Sfax University, Road Sokra, Sfax, Tunisia
rania.rebai@hotmail.fr, fendri.msf@gnet.tn, Mohamed.Hammami@fss.rnu.tn

Abstract. In video surveillance, we can rely on either a visible spectrum or an infrared one. In order to profit from both of them, several fusion methods were proposed in literature: low-level fusion, middle-level fusion and high-level fusion. The first one is the most used for moving objects' detection. It consists in merging information from visible image and infrared one into a new synthetic image to detect objects. However, the fusion process may not preserve all relevant information. In addition, perfect correlation between the two spectrums is needed. In This paper, we propose an intelligent fusion method for moving object detection. The proposed method relies on one of the two given spectrum at once according to weather conditions (darkness, sunny days, fog, snow, etc.). Thus, we first extract a set of low-level features (visibility, local contrast, sharpness, hue, saturation and value), then a prediction model is generated by supervised learning techniques. The classification results on 15 sequences with different weather conditions indicate the effectiveness of the extracted features, by using C4.5 as classifier.

Keywords: Image fusion · Classification · Weather conditions · Moving object detection

1 Introduction

Moving object detection in complex scenes is an active research topic in computer vision. The related research area includes intelligent video analysis, which can be applied to monitor outdoors areas such as airports, streets, highways, subways and parking lots etc. The research diversity is justified by the complexity of the problem and the variability of its challenges, still incompletely resolved, like detection in night-time, in total occlusion and in presence of non-stationary background objects. Therefore, we relied on two categories of cameras: visible ones and infrared ones which provide as respectively visible (VIS) spectrum and infrared (IR) spectrum.

In the literature, several methods[1],[2],[3],[4],[5],[6] have been proposed for moving object detection in VIS spectrum. These methods are based either on background modeling [1],[2],[3], on optical flow [4],[5],[6], or on inter-frame difference [7],[8],[9]. However, these methods suffer from many limitations such as

L. Rutkowski et al. (Eds.): ICAISC 2015, Part I, LNAI 9119, pp. 768–777, 2015.
DOI: 10.1007/978-3-319-19324-3_69

failure to face camouflage, night or poor visibility conditions (fog, snow, rain, etc.).

In order to overcome these limitations, many works [10],[11],[12] propose to use IR sensor. We can distinguish two approaches to detect moving object from IR sensor: pixel/region-based approaches [12],[13],[14] and model-based approaches [13],[14],[15],[16]. These methods achieve a good performance especially in night and/or poor visibility conditions, but fail in presence of some climatic conditions like a hot sunny day or when the object has low contrast (not warmer than the background) [17],[18].

Face to limitations of the use, at once, of visible or infrared spectrums, recent researches [14],[17],[19],[20] propose to use both Visible and Infrared sensors. A fusion between the information provided by VIS and IR cameras for moving object detection would offer complementary solutions: relying on visible images in sunny days, we achieve a good detection and can extract a rich content; while the use of infrared sensor seems to give better results for moving object detection in presence of darkness, limited levels of luminosity, shadows, light reflections, or some weather variations. Visible and infrared fusion aims to perform correct moving object detection all over the day (morning, afternoon and night) for particular hot objects such as persons and vehicles [21].

In the literature, moving object detection using both visible and infrared spectrums suppose merging information in different levels: low-level, medium-level or high-level. Experimentations prove that, in some cases, merging information may reduce the quality of moving object detection. Thus, in this paper our main contribution is to propose an intelligent fusion of VIS-IR spectrums based on weather conditions' classification. The rest of this paper is organized as follows: Section 2 provides an overview of literature related to fusion techniques. Section 3 describes the proposed method for VIS-IR spectrums fusion. Section 4 outlines the results of a quantitative and a qualitative evaluation. Finally, Section 5 recapitulates the presented method and outlines future work.

2 State of Art on Fusion Techniques

Moving objects detection' methods relying on Visible/Infrared fusion can be classified into three categories according to the level of processing [17],[18],[22]: Low-level fusion, Medium-level fusion and High-level fusion. In low-level fusion, also called signal, data or pixel-level fusion, fused images are generated by merging pixel information from both spectrums. Therefore, infrared and visible images must be synchronized so that all pixel positions of all the input images correspond to the same location. The most common pixel level fusion techniques [17],[23],[24] are: techniques based on Weighted Averaging, techniques based on Pyramid Transforms and techniques based on a Wavelet Transforms.

In medium-level fusion, also called feature-level fusion, they first extract features from both the infrared spectrum and the visible one. Then, they fuse the extracted features. This fusion could be achieved in two ways: among the two modules of features extraction and features selection or after both of them [19].

Since, one of the essential goals of fusion is to preserve the image features, feature level methods have the ability to yield subjectively better fused images than pixel based techniques [17],[25].

Finally, the last category of fusion methods concerns high-level fusion. In this latter, the fusion is applied either at decision level or at score level. In the fusion at decision level, the classifiers are applied independently to each sensor output. The given decisions are combined to make a final decision [17]. In the score level fusion, multiple classifiers produce a set of scores which represent the probabilities that one object belongs to different possible classes. These score can be combined by a weighted parameter in order to obtain a new score which is then used to make the final decision [19].

The choice of the fusion level depends on the nature of the handled application. In our context, we aim to improve the quality of moving object detection. Thus, we will be interested with low-level fusion. However, in this fusion, we must satisfy a set of constraints [26]: the fusion process should preserve all relevant information on the input imagery in the composite image; the fusion process should not introduce any artifacts or inconsistencies and the fusion process should be shift and rotational invariant. Moreover, low-level fusion techniques suppose that a perfect correlation between the images is performed before performing the fusion itself. When images are not well correlated, it could lead to errors in the image fusion process [19].

For this reasons, we propose an intelligent fusion of VIS-IR videos to profit from the quality of both of them, without having neither to correlate the spectrums nor to generate a fused spectrum that may be different from both of them. To make this manuscript clear to read and easy to grasp, these works will be detailed in next section.

3 Proposed Method

We propose an intelligent fusion method for moving object detection. The proposed method rely either on visible spectrum either on infrared spectrum according to weather conditions and timing of the video acquisition. Thus, the visible spectrum is used in sunny days under normal weather conditions, while the infrared spectrum will be used at night or in presence of fog, rain, snow, etc. Our method is composed of two steps: (i) offline step adopting a data-mining process in order to build the adequate prediction model for abnormal weather classification and (ii) an online step to classify VIS images into image in Normal conditions or in Abnormal conditions and to detect Moving objects in IR spectrum or in VIS spectrum. Fig. 1 shows the framework of the proposed method.

3.1 Offline Step

Our offline step is composed of two major steps: (1) Data preparation step, and (2) Data mining step which aims to build a generic prediction model by the use of several data-mining algorithms.

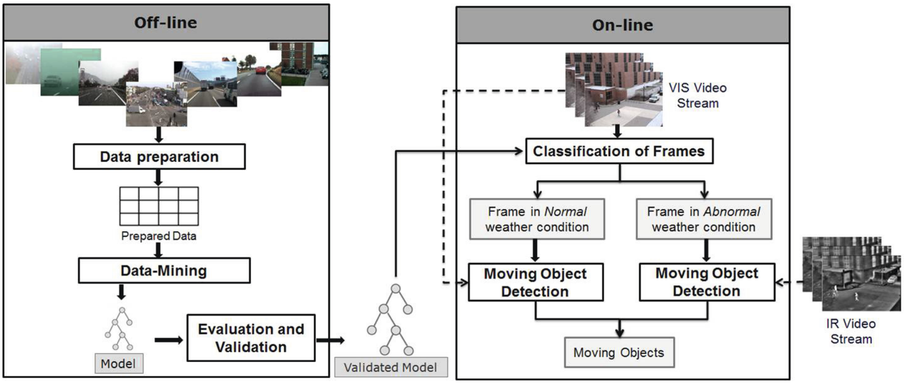

Fig. 1. Proposed method of intelligent fusion VIS-IR for moving object detection

Data Preparation. In this step, we identify efficient weather conditions features in order to build a two-dimensional table from our training corpus. This table is devoted thereafter to the learning step. In our case, the robustness of an image classification technique depends on reliable and strong environmental features. A thorough look on the features that are most commonly employed for describing visibility in the literature provided the grounds for the ultimate selection of eight features for consideration in our work : Visibility, Local Contrast, Sharpness, Hue, Saturation and Value [27] and two temporal features based on the autocorrelation of each pixel's intensities over time [28], detailed below.

Visibility metric. The visibility metric (equation 1) calculates the ratio between contrast and noise of Image estimated by a Gaussian kernel.

$$Visibility = \frac{\sum_L \sum_C \sqrt{IM_{noise^2}}}{L * C} . \tag{1}$$

Where L and C represent the number of row and column of the image, respectively, IM_{noise} is the image noise filtered by a Gaussian filter.

Local Contrast. This feature calculates the contrast between a pixel and its neighbors (equation 2).

$$I_{Cont}\left(i\right) = \frac{I\left(i\right) - M_w}{S_w} . \tag{2}$$

Where M_w and S_w is respectively the mean of the pixels' gray values and the standard deviation of the neighbors window.

Sharpness. Seeing that visible images in normal conditions have sharp edges with large contrast differences, the sharpness was considered as a meaningful feature to classify images. Roser and al. [27] proposed a measure of sharpness (equation 3), based on the average of the Sobel gradient magnitude.

$$T_{adv} = \frac{\sum_i \delta_i \, \rho\left(i\right) \sqrt{S_X^2\left(i\right) + S_Y^2\left(i\right)}}{\sum_i \delta_i} . \tag{3}$$

Where, $S_X(i)$ and $S_Y(i)$ are the Sobel filter values for each pixel i, δ_i informs if the pixel is an edge one (= 1, 0 otherwise), and $\rho(i)$ is a weighting factor that is assumed to be inversely proportional to the local contrast.

Features based on AutoCorrelation Function (ACF). Based on the autocorrelation function (ACF) of the intensities of each pixel in the time, two temporal characteristics (C and S) are used for the classification of weather conditions [28] as are represented in the equations 4 and 5. The feature C indicates whether the spatial average of the current frame of the sequence has weaker time-average autocorrelation of intensity change or not, while the feature S shows whether the spatial average of the current frame of a sequence in weather condition under classification is only in short-time autocorrelation of intensity change or not. In fact, the intensity change at the pixel is proportional to the illumination variation speed. For instance, for fast illumination variation, we notify a strong short-time autocorrelation; however, in presence of rain streaks or snowflakes (two brightness states at a fixed pixel), it leads to weaker time-average ACF value than that of gradual illumination variation.

$$C = \underset{y \epsilon \Omega}{mean}\left(\sum_{k=1}^{T-1}\frac{\hat{\rho_y}(k)}{T}\right). \tag{4}$$

$$S = \underset{y \epsilon \Omega}{mean}\left(\hat{f}_y(k)\right). \tag{5}$$

Where $\hat{\rho_y}(k)$ is equivalent to ACF value at location y in the k^{th} frame interval. Ω is the current frame and T is the limited time length. $\hat{f}(k)$ represents the quadratic fit of $\hat{\rho}(k)$.

Color Features. Weather conditions variation can be detected in case of color variation. We choose to consider the Hue, Saturation and Value of images in HSV color space. This space is known as being the closest one to human perception.

Data Mining. Our goal is to build a predictive model to classify the VIS image in Normal weather conditions or abnormal ones. This prediction model is obtained by supervised learning technique. In supervised learning, the efficiency and genericity of the generated classifier increases when the size of the training set and the number of relevant features increase. Another pertinent setting to consider is the choice of the appropriate learning technique.

Evaluation and Validation. The objective of this step is to evaluate the previous one. It consists in comparison of different prediction models learned in order to determine the best prediction model for images' classification according to weather conditions. Therefore, as for the majority of recent works, we construct the confusion matrix to evaluate the quality of a prediction model. From this confusion matrix we can calculate the Total Correct Classification TCC detailed in the equation 6.

$$TCC = \frac{n_{AA} + n_{BB}}{n_{AA} + n_{AB} + n_{BB} + n_{BA}}. \tag{6}$$

- n_{AA}: represents the number of frames in Abnormal weather conditions correctly classified
- n_{BB}: represents the number of frames in Normal weather conditions correctly classified.
- n_{AB}: represents the number of frames in Abnormal weather conditions classified as in Normal weather conditions.
- n_{BA}: represents the number of frames in Normal weather conditions classified as in Abnormal weather conditions.

3.2 Online Step

In the proposed approach, after the offline step, an online step is carried out. This latter starts with a step of VIS images classification by the extracted prediction model. Then, according to the decision, we perform either moving object detection on VIS spectrum or on IR spectrum.

Classification of Frames. The objective of this step is to classify the images of VIS spectrum into image in Normal weather conditions or in Abnormal weather conditions. This classification is based on the prediction model extracted from the offline step.

Moving Object Detection. As soon as images are classified as in Normal or Abnormal weather conditions, the visible (respectively infrared) spectrum is considered to perform a moving object detection technique. In this work, we have adopted a method based on background modeling with dynamic matrix and spatio-temporal analyses of scenes [29]. This method has shown high performances and robustness in foreground segmentation under various complex scenes conditions such as sudden and gradual illumination changes, ghost and foreground speed.

4 Experimental Results

In order to evaluate our proposed method, we carried out a series of experiments. We performed experiments on a large and representative corpus shown in table 1 and table 2 (15 famous outdoor sequences recorded in typical conditions). This corpus consists of 7 sequences in Normal weather conditions (5440 images) and 8 sequences in Abnormal weather conditions (3525 images). The Abnormal sequences present several challenges such as fog, rain, snow etc. We randomly selected sequences from the database to build up our fixed training (70%) and testing data sets (30%). We ensure that no sequence is used for both training and testing at the same time for each class.

We then have build learning data: an $N * 9$ matrix of extracted features from our training corpus (N is the number of pixels from our training corpus). Once learning data were defined, we proceeded to selecting the appropriate learning technique. In fact, in literature, we find several techniques of supervised learning, each with its own advantages and drawbacks. Therefore, the learning is

Table 1. Sequences of image in Abnormal State

Sequences	Number of Frames	Resolution (Pixels)	Used for
Brouillard1[30]	700	720∗576	Learning
Set10_RainyDay[1]	1049	640∗360	Learning
Set10_SunStrokes[1]	347	640∗360	Learning
Set10_Tunnel[1]	216	640∗360	Learning
Dtneu_nebel[2]	349	768∗576	Learning
Dtneu_winter[2]	299	768∗576	Test
Dtneu_schnee[2]	298	768∗576	Test
Brouillard2[30]	267	720∗500	Test

Table 2. Sequences of image in Normal State

Sequences	Number of Frames	Resolution (Pixels)	Used for
Set3_SuburbanFollow[1]	1171	1280∗1024	Learning
Set10_Daylight[1]	902	640∗360	Learning
Balcony5_Vis[3]	564	640∗480	Learning
qmul_junction[4]	867	360∗288	Learning
Set3_SuburbanBridge[1]	851	1280∗1024	Learning
Set3_TrailerFollow[1]	800	1280∗1024	Test
Set10_Snowy[1]	285	640∗360	Test

performed by 3 different learning algorithms: the decision tree C4.5, SVM with Radial Basis Function Kernel and Multilayer Perceptron Neural Network (MLP). Each method is considered as reference in its category. This data mining algorithms were compared according to Total Correct Classification (equation 6). The experimental results are shown in figure 2. We obtained a best classification rate on learning and Test set by C4.5 (81.16%).

Fig. 3 shows some promising results of our method. In fact, the images (a) and (b) are extracted from two sequences which have two bad weather conditions respectively snowy day and fog. These two images are classified in Abnormal conditions. In the other hand, the images (b) and (c) present two scenes in favorable weather conditions that are classified as in the Normal conditions. Note that in the scene of (b) there's snow in the boards of the road but it does not snow when it is recorded. This is due to the characteristics based on the autocorrelation of pixel-wise intensities over time that allow to distinguish the motion blur caused by rain streaks or snowflakes.

[1] http://ccv.wordpress.fos.auckland.ac.nz/eisats/
[2] http://i21www.ira.uka.de/image_sequences/
[3] http://www.eeng.dcu.ie/~oconaire/dataset/
[4] http://www.eecs.qmul.ac.uk/~ccloy/index.html

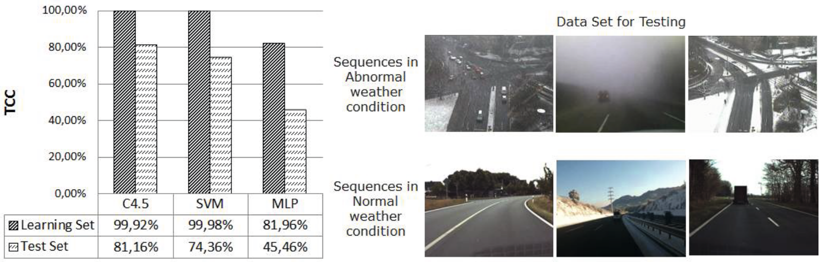

Fig. 2. Results of images' classification as normal/abnormal weather conditions

Fig. 3. Results of classification (a) and (b) are classified in Abnormal weather condition but (c) and (d) are classified in Normal weather condition

5 Conclusion

In this paper, we proposed a novel method of intelligent fusion for moving object detection. This method relies on a classification step of images according to the weather conditions. We consider visible or infrared spectrums according to weather conditions (darkness, sunny days, fog, snow, etc.). For thus, we first extract a set of low-level features (visibility, local contrast, sharpness, two features based on ACF which are C and S, hue, saturation and value), then we generate a prediction model by supervised learning techniques. Experimentations carried out on several sequences with different weather conditions prove the effectiveness of the generated prediction model. Compared to two other learning techniques (SVM and MLP), prediction model generated by C4.5 records the best classification rate with 81.16%. Our future orientations will examine the impact of our contribution on the accuracy of moving object detection in VIS and IR videos.

References

1. Brutzer, S., Höferlin, B., Heidemann, G.: Evaluation of background subtraction techniques for video surveillance. In: IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, pp. 1937–1944 (2011)
2. Biswas, S., Sil, J., Sengupta, N.: Background modeling and implementation using discrete wavelet transform, a review. International Journal on Graphics, Vision and Image Processing 11(1), 29–42 (2011)

3. Cristani, M., Farenzena, M., Bloisi, D., Murino, V.: Background subtraction for automated multisensor surveillance: A comprehensive review. EURASIP J. Adv. Sig. Proc. (2010)
4. Roth, S., Black, M.J.: On the spatial statistics of optical flow. In: 10th IEEE International Conference on Computer Vision, Beijing, China, pp. 42–49 (2005)
5. Pathirana, P.N., Lim, A.E.K., Carminati, J., Premaratne, M.: Simultaneous estimation of optical flow and object state: A modified approach to optical flow calculation. In: IEEE International Conference on Networking, Sensing and Control, London, pp. 634–638 (2007)
6. Lim, S., Apostolopoulos, J.G., Gamal, A.E.: Optical flow estimation using temporally over-sampled video. IEEE Transactions on Image Processing, 1074–1087 (2005)
7. Guo, J., Chng, E.S., Deepu, R.: Foreground motion detection by difference-based spatial temporal entropy image. In: IEEE Region 10 Conference, pp. 379–382 (2004)
8. Chang, M., Cheng, Y.: Motion Detection by Using Entropy Image and Adaptive State-Labeling Technique. In: IEEE International Symposium on Circuits and Systems, New Orleans, pp. 3667–3670 (2007)
9. Durus, M., Ercil, A.: Robust Vehicle Detection Algorithm. In: 15th IEEE International Conference on Signal Processing and Communications Applications, Siu, pp. 1–4 (2007)
10. Gilmore, E.T., Ugbome, C., Kim, C.: An IR-based Pedestrian Detection System Imple-mented with Matlab-Equipped Laptop and Low-Cost Microcontroller. International Journal of Computer Science & Information Technology 3(5), 79–87 (2011)
11. Xu, F., Liu, X., Fujimura, K.: Pedestrian Detection and Tracking with Night Vision. IEEE Transactions on Intelligent Transportation Systems 6(5), 63–71 (2005)
12. Fang, Y., Yamada, K., Ninomiya, Y., Horn, B.K.P., Masaki, I.: A shape-independent method for pedestrian detection with far-infrared images. IEEE Transactions on Vehicular Technology 53(6), 1679–1697 (2004)
13. Olmeda, D., Hilario, C., de la Escalera, A., Armingol, J.M.: Pedestrian Detection and Tracking Based on Far Infrared Visual Information. In: Blanc-Talon, J., Bourennane, S., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2008. LNCS, vol. 5259, pp. 958–969. Springer, Heidelberg (2008)
14. Bertozzi, M., Broggi, A., Rose, M.D., Felisa, M., Rakotomamonjy, A., Suard, F.: A Pedestrian Detector Using Histograms of Oriented Gradients and a Support Vector Machine Classifier. In: IEEE Intelligent Transportation Systems Conference, USA, pp. 143–148 (2007)
15. Nanda, H., Davis, L.: Probabilistic Template Based Pedestrian Detection in Infrared Videos. In: IEEE Intelligent Vehicles Symposium, Paris, France, pp. 15–20 (2002)
16. Dai, C., Zheng, Y., Li, X.: Pedestrian detection and tracking in infrared imagery using shape and appearance. Journal of Computer Vision and Image Understanding 106, 288–299 (2007)
17. Zin, T.T., Takahashi, H., Toriu, T., Hama, H.: Fusion of Infrared and Visible Images for Robust Person Detection. Image Fusion (2011) ISBN: 978-953-307-679-9
18. Goubet, E., Katz, J., Porikli, F.: Pedestrian tracking using thermal infrared imaging. In: Infrared Technology and Applications XXXII Proc. SPIE, vol. 6206 (2006), doi:10.1117/12.673132
19. Apătean, A.D.: Contributions to the Information Fusion, Application to Obstacle Recognition in Visible and Infrared Images. Doctoral thesis (2011)

20. Wang, J., Liang, J., Hu, H., Li, Y., Feng, B.: Performance evaluation of infrared and visible image fusion algorithms for face recognition. In: International Conference on Intelligent Systems and Knowledge Engineering, Chengdu, China, pp. 1–8 (2007)
21. Ó Conaire, C., Cooke, E., O'Connor, N., Murphy, N., Smeaton, A.: Fusion of infrared and visible spectrum video for indoor surveillance. In: 6th International Workshop on Image Analysis for Multimedia Interactive Services, Montreux, Switzerland (2005)
22. Apatean, A., Rogozan, A., Bensrhair, A.: Information Fusion for Obstacle Recognition in Visible and Infrared Images. In: International Symposium on Signals, Circuits and Systems, pp. 1–4 (2009)
23. Blum, R.S., Xue, Z., Zhang, Z.: An Overview of Image Fusion. In: Multi-Sensor Image Fusion and Its Applications, pp. 1–36. Taylor & Francis, Boca Raton (2006)
24. Sadjadi, F.: Comparative Image Fusion Analysais. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, vol. 8(8), p. 25 (2005)
25. Samadzadegan, F.: Data Integration Related to Sensors, Data and Models. In: International Society for Photogrammetry and Remote Sensing (2004)
26. Yang, B., Zhong-liang, J., Hai-tao, Z.: Review of pixel-level image fusion. Journal of Shanghai Jiaotong University 15(1), 6–12 (2010)
27. Roser, M., Moosmann, F.: Classification of Weather Situations on Single Color Images. In: IEEE Intelligent Vehicles Symposium Eindhoven University of Technology, Eindhoven, The Netherlands, pp. 798–803 (2008)
28. Zhao, X., Liu, P., Liu, J., Tang, X.: Feature extraction for classification of different weather conditions. Frontiers of Electrical and Electronic Engineering 6(2), 339–346 (2011)
29. Hammami, M., Jarraya, S.K., Ben-Abdallah, H.: On line background modeling for moving object segmentation in dynamic scene. Multimedia Tools and Applications 63(3), 899–926 (2013)
30. Hammami, M., Ben Romdhane, N., Ben-Abdallah, H.: An Improved Lane Detection and Tracking Method for Lane Departure Warning Systems. International Journal of Computer Vision and Image Processing 3(3), 1–15 (2013)