Background subtraction with multispectral video sequences

Yannick Benezeth, Désiré Sidibé, Jean-Baptiste Thomas

Background subtraction with multispectral video sequences

Yannick Benezeth, Désiré Sidibé and Jean Baptiste Thomas
Université de Bourgogne - LE2I UMR CNRS 6306
France
Email: {firstname.lastname} @u-bourgogne.fr

***Abstract*—Motion analysis of moving targets is an important issue in several applications such as video surveillance or robotics. Background subtraction is one of the simplest and widely used techniques for moving target detection in video sequences. In this paper, we investigate the advantages of using a multispectral video acquisition system of more than three bands for background subtraction over the use of trichromatic or monochromatic video sequences. To this end, we have established a dataset of multispectral videos with a manual annotation of moving objects. To the best of our knowledge, this is the first publicly available dataset of multispectral video sequences. Experimental results indicate that using more than three spectral sub-bands provide better discrimination between foreground and background pixels. In particular, the use of the near infra-red (NIR) spectral band together with visible spectra provides the best results.**

I. Introduction

Multispectral imaging has been an area of active research the last decades with great success in several applications. One could locate multispectral imaging between monochromatic imaging and hyperspectral imaging. Multispectral imaging having a better spectral resolution than the former one, and an indirect relation with radiance compared with the later. Recent advances in technology offer us now the possibility to record multispectral of more than three spectral sub-band in the visible and near infra-red (NIR) part of the spectrum. These systems are mostly based on silicon sensors, which limits the reasonable spectral bandwidth from below 400 nm up to slightly over 900 nm. Several acquisition systems have been proposed for multispectral imaging either at the expense of sensitivity (multi-CCD dichroic filter based cameras), at the expense of spatial resolution (MSFA/CFA based cameras [1], [2]) or at the expense of temporal resolution (sequential filtering [4]). Some systems show also a compromise between these drawbacks (*e.g.* with light-field cameras, combination of cameras).

Although the design of multispectral imaging system is still a running problem, different systems are commercially available for still image acquisition. However, it was very difficult until now to consider multispectral video. New products have been made commercially available very recently, which permit video capture. This opens new capabilities for video processing and require studies to investigate how beneficial this could be.

Until now, mostly mono or trichromatic cameras within the visible spectrum (Graylevel and RGB) or Near Infrared (NIR) part were used in video analytics. Obviously, the use of RGB over monochromatic cameras has the advantage of a better segmentation between homo-luminance areas where objects are only separable by their chromatic information. NIR made possible to better separate shadows from the object, as well as a higher spatial resolution in the presence of fog or haze, and more homogenous object reflectance [3].

To the best of our knowledge, there is no published work that uses real multispectral video to solve common issues of video analytics. Usually, existing work are dealing with still recombined images. For example, Conaire *et al.* [6] propose an object segmentation method for surveillance video that fuses RGB standard video with thermal Infrared (IR) videos. They use Transferal Belief Model to fuse the two modalities and segment objects.

In this work, we propose to evaluate the advantages of using multispectral video of more than three spectral sub-bands to solve common issues of video analytics. A very important element in computer vision algorithms is to compute similarity or dissimilarity between image pixels or patches. This core capability represents the basis for most applications. Improvements in the robustness of similarity measurement between pixels or patches using higher spectral resolution would therefore have significant impact on all these applications. The use of measure between spectra instead of RGB triplets will discern subtle differences between colors impossible with regular RGB images. This capability is critical for example to detect moving objects in a video. That is why we demonstrate the advantages of using multispectral videos rather than color or grayscale ones for background subtraction problems. Three ways of using multispectral videos to perform background subtraction are described and evaluated in this paper. More precisely, we investigate the use of the well-known Mahalanobis distance [9], the combination of background subtraction results obtained on each spectral band of the multispectral image sequence and the use of a spectral similarity measurement. In order to perform a quantitative analysis, we established a dataset of multispectral videos including outdoor and indoor scenes with different challenges. The images are manually annoted to provide ground truth data at pixel resolution. To the best of our knowledge, this is the first dataset of multispectral video sequences and it has been made publicly available to the research community[1].

The rest of the paper is organised as follows. First, we give an overview of background subtraction methods in Section II. The extension to multispectral videos is described in Section III. The multispectral video dataset used for experiments and evaluation metrics are presented in Section IV.

[1] http://ilt.u-bourgogne.fr/benezeth/projects/ICRA2014/

Section V shows the results, and concluding remarks are given in Section VI.

II. Overview of background subtraction methods

Background Subtraction (BS) techniques presented in literature differentiate in the way they build the background model, the way the model is updated over time, and how foreground pixels are detected. Simple BS methods label "in motion" every pixel whose color is significantly different from the ones in the background. Detecting motion through BS is not always as easy as it may first appear. Some videos with poor SNRs caused by a low-quality camera or compression artefacts can generate numerous false positives. Actually, false positives can be induced by several reasons. Recent reviews and evaluations of BS methods (*e.g.* [7], [8], [9]) have identified several challenges a BS algorithm needs to address. Among other difficulties, a BS technique should be robust to illumination changes (sudden or gradual), background movements such as trees shaken by the wind, shadows, camouflage effects (color similarity of object and background) and intermittent object motion that may cause "ghosting" artefacts. Even if BS is one of the most commonly encountered low-level task in computer vision, no widely accepted BS technique has solved all the problems mentioned above. This can be gauged by the large number of papers that are published regularly on this topic. The objective of this section is not to present a comprehensive survey of BS techniques, but to present the most important categories with reference methods and recent improvements. More detailed surveys and reviews can be found in [9], [10], [11].

One common approach is to model each pixel with a probability density function (PDF) learned over a series of training frames. In this case, the BS problem becomes a PDF-thresholding issue for which a pixel with low probability is likely to correspond to a foreground moving pixel. These methods usually model each pixel in a video frame with a Gaussian distribution. A very popular technique of this category is [12]. By using more than one Gaussian distribution, it is possible to manage backgrounds made of animated textures such as waves or trees shaken by the wind. One of the most widely used background subtraction method models each pixel as a mixture of weighted Gaussian distributions [13]. This model still raises a lot of interest as authors continue to propose improvements and extensions to the Gaussian Mixture Models (GMM), *e.g.* in [14], [15].

To avoid the difficult question of finding appropriate parameters for the PDF, non-parametric methods to model background distributions such as Kernel Density Estimation (KDE) have been proposed. Since the cost to compute the kernel density estimate at each pixel is very high, several pre-calculated lookup tables are used to reduce the burden of computation of the algorithm. As the GMM-based methods, the KDE approach initially proposed by Elgammal *et al.* [16] has often been modified, *e.g.* in [18], [19], [20].

Instead of using parametric or non-parametric background models, it is also possible to collect samples for the background model. Recently, Wang and Suter [21] have proposed to memorize the last N most recent images using a first-in first-out strategy. Then, in order to classify a pixel value as foreground or background, a consensus is sought simply by counting the number of times previous samples agree with the current sample. Another approach of this category is the $ViBe$ system [17], [22]. A set of values is also used to model each location but the background values are updated by a random scheme and updated pixels can diffuse their current value to neighbouring pixels. The codebook algorithm [23] is another approach close to the ones of this category. Based on a training sequence, the method assigns to each background pixel a series of key color values (called codewords) stored in a codebook. Those codewords describe which color a pixel is likely to take over a certain period of time.

Most techniques described in the literature correspond to one of the three categories described above. However, it is interesting to cite other approaches which use principal component analysis [24], [25], independent component analysis [26], artificial neural networks [27] or a Bayesian framework [28] to name a few.

In the following section, three ways of using mulispectral videos to perform background subtraction are described.

III. Background subtraction on multispectral videos

A. Straightforward extension of color-based BS algorithms

Most BS techniques make the assumption that the observed video sequence I is made of a static background B in front of which moving objects are observed. Numerous BS methods can be summarized by the following formula:

$$\mathcal{X}_t(s) = \begin{cases} 1 & \text{if } d(\boldsymbol{I}_{s,t}, \boldsymbol{B}_{s,t}) > \tau \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where τ is a threshold, $\mathcal{X}_t$ is the motion label field at time t (also called motion mask), d is the distance between $\boldsymbol{I}_{s,t}$, the pixel value at time t and location s, and $\boldsymbol{B}_{s,t}$, the background model. If I is a regular color-based video sequence, $\boldsymbol{I}_{s,t}$ is a vector defined by $\boldsymbol{I}_{s,t} = [I^R_{s,t}, I^G_{s,t}, I^B_{s,t}]$, where R, G and B stand for the red, $green$ and $blue$ channels respectively. If I is a multispectral video sequence, $\boldsymbol{I}_{s,t}$ is also a vector defined by $\boldsymbol{I}_{s,t} = [I^1_{s,t}, I^2_{s,t}, \ldots, I^k_{s,t}]$, where k stands for the number of spectral bands composing each multispectral frame (with $k > 3$). The easiest way to model the background B is through a single image free of moving objects. With this simple background model, pixels corresponding to foreground moving objects can be detected by thresholding a distance function, *e.g.* the Euclidian distance:

$$d_2 = \left(\sum_{i=1}^{k} \left(I^i_{s,t} - B^i_{s,t} \right)^2 \right)^{\frac{1}{2}}. \tag{2}$$

Modelling B with a single image requires a rigorously fixed background void of noise and artefacts. Since this requirement cannot be satisfied in every real-life scenario, it is possible to model each background pixel with a probability density function (PDF) learned over a series of training frames. In this case, the BS problem becomes a PDF-thresholding issue for which a pixel with low probability is likely to correspond to

a foreground moving object. For instance, in order to account for noise, it is possible to model every background pixel with a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_{s,t}, \boldsymbol{\Sigma}_{s,t})$ where $\boldsymbol{\mu}_{s,t}$ and $\boldsymbol{\Sigma}_{s,t}$ stand for the average background multispectral vector and covariance matrix at pixel s and time t. In this context, the distance metric can be the Mahalanobis distance:

$$d_M = |\boldsymbol{I}_{s,t} - \boldsymbol{\mu}_{s,t}|\boldsymbol{\Sigma}_{s,t}^{-1}|\boldsymbol{I}_{s,t} - \boldsymbol{\mu}_{s,t}|^T. \quad (3)$$

Since the covariance matrix contains large values in noisy areas and low values in more stable areas, $\boldsymbol{\Sigma}$ makes the threshold locally dependent on the amount of noise. In other words, the noisier a pixel is, the larger the temporal gradient $|\boldsymbol{I}_{s,t} - \boldsymbol{\mu}_{s,t}|$ has to be to get the pixel labeled in motion. This makes the method significantly more flexible than the basic motion detection one.

Since the illumination often changes in time, the mean and covariance of each pixel can also be iteratively updated following this procedure:

$$\boldsymbol{\mu}_{s,t+1} = (1-\alpha).\boldsymbol{\mu}_{s,t} + \alpha.\boldsymbol{I}_{s,t} \quad (4)$$

$$\boldsymbol{\Sigma}_{s,t+1} = (1-\alpha).\boldsymbol{\Sigma}_{s,t} + \alpha.(\boldsymbol{I}_{s,t} - \boldsymbol{\mu}_{s,t})(\boldsymbol{I}_{s,t} - \boldsymbol{\mu}_{s,t})^T. \quad (5)$$

Note that even if $\boldsymbol{\Sigma}$ is by definition a $k \times k$ matrix, it can be assumed to be diagonal to reduce memory and processing costs. This first background subtraction method for multispectral video is a straightforward extension of a color-based BS algorithm however we will demonstrate its good performance in Section V.

B. Pooling

A second approach to perform BS on multispectral videos is to combine BS results from several spectral bands with a pooling of results. A typical and simple definition would be:

$$\mathcal{X}_t(s) = \begin{cases} 1 & \text{if } \sum_i \mathcal{X}_t^i(s) > \rho \\ 0 & \text{otherwise} \end{cases}, \quad (6)$$

assuming that $\mathcal{X}_t^i(s)$ is the motion mask obtained with BS on the spectral band i, at the spatial location s and time t. ρ is the definition of the majority, spanning from 1 (equivalent to a logical OR) to k, the number of bands (equivalent to a logical AND).

C. Spectral distance measurement

Instead of using a straightforward extension of the color-based Mahalanobis distance, it is possible to use a dedicated spectral distance measure. Several similarity or dissimilarity measures have been designed for the comparison of spectral vectors. A popular one is the spectral angle d_θ that calculates the angle between two spectra [30] . Spectra are considered as vectors in a k-dimensional space. Small angles indicate similar vectors.

The angle between the multispectral vector $\boldsymbol{I}_{s,t}$ of the current image and $\boldsymbol{\mu}_{s,t}$ the background model is defined by:

$$d_\theta(\boldsymbol{I}_{s,t}, \boldsymbol{\mu}_{s,t}) = \cos^{-1}\left(\frac{<\boldsymbol{I}_{s,t}, \boldsymbol{\mu}_{s,t}>}{\|\boldsymbol{I}_{s,t}\| \left\|\boldsymbol{\mu}_{s,t}\right\|}\right), \quad (7)$$

where $<,>$ is the dot product between two vectors. One of the key advantage of this metric is that it is intensity invariant because the angle between two vectors is independent of the vector length. This property is very interesting for BS problems to handle shadows and illumination variations.

Another popular way of measuring distance between spectra is the spectral information divergence d_{SID} [29]. d_{SID} derived from the information theory and is based on the Kullback-Leibler divergence. Here, $\boldsymbol{I}_{s,t}$ is considered as a random variable. Probability measures $P_i(\boldsymbol{I}_{s,t})$ and $P_i(\boldsymbol{\mu}_{s,t})$ are defined with:

$$P_i(\boldsymbol{I}_{s,t}) = \frac{I_{s,t}^i}{\sum_{j=1}^k I_{s,t}^j} \quad \text{and} \quad P_i(\boldsymbol{\mu}_{s,t}) = \frac{\mu_{s,t}^i}{\sum_{j=1}^k \mu_{s,t}^j}, \quad (8)$$

with $1 \leq i \leq k$ and k the number of bands of the multispectral image. The spectral information divergence $d_{SID}(\boldsymbol{I}_{s,t}, \boldsymbol{\mu}_{s,t})$ between the current spectral vector $\boldsymbol{I}_{s,t}$ and the background model $\boldsymbol{\mu}_{s,t}$ can now be defined with:

$$d_{SID}(\boldsymbol{I}_{s,t}, \boldsymbol{\mu}_{s,t}) = \sum_{i=1}^k P_i(\boldsymbol{I}_{s,t}) \log \frac{P_i(\boldsymbol{I}_{s,t})}{P_i(\boldsymbol{\mu}_{s,t})} + \sum_{i=1}^k P_i(\boldsymbol{\mu}_{s,t}) \log \frac{P_i(\boldsymbol{\mu}_{s,t})}{P_i(\boldsymbol{I}_{s,t})}. \quad (9)$$

Unlike d_θ which extracts geometric features between two spectra, d_{SID} views each pixel spectrum as a random variable and measures the discrepancy of probabilistic behaviors between two spectra. This measure is relatively recent and is expected to be more effective than d_θ in preserving spectral properties [29].

IV. Experimental protocol

A. Video dataset

In order to highlight the advantages of using multispectral video to perform background subtraction and to compare the various strategies presented in Section III, we have established a dataset of multispectral videos. Illustrations of the dataset are presented in Fig. 1. Our dataset is composed of 5 video sequences containing between 250 and 2300 frames of size 658×491. The video dataset represents 1 indoor video sequence and 4 outdoor scenes with different challenges such as gradual illumination changes, shadows, camouflage effects (color similarity of object and background) and intermittent object motion. The objective of this work is not to evaluate the robustness of BS algorithms on multimodal backgrounds, we focus on the ability to better discriminate subtle color differences and robustness to illumination variation using multispectral information. Therefore, we have based our study on videos with quasi-static backgrounds.

Fig. 1. Snapshots of each camera viewpoint of the video dataset.

For each scene, we have acquired a sequence of multispectral images whose framerate depends on the overall scene illumination, *i.e.* from 5 frames per second for dark scenes (*e.g.* video 1) to 15 frames per second for bright ones (*e.g.* video 3). The acquisitions are performed with a commercial camera from *FluxData, Inc.* (the FD-1665-MS). This camera is based on a 3-CCD system that provides simultaneous data for each spectral band. It can acquire 7 spectral narrow bands, six in the visible spectrum and one in the near infrared. An example of a multispectral image is presented in Fig. 2. This camera is light and compact and is adapted for videosurveillance applications.

Color image sequences are easily obtained with a linear integration of the original multispectral images weighted by three different spectral envelopes [31]. Let r_i, g_i and b_i be the weights on the i^{th} spectral band, defined according to the characteristics of the multispectral camera. Color pixel values $[I^R_{s,t}, I^G_{s,t}, I^B_{s,t}]$ are obtained with:

$$I^R_{s,t} = \sum_{i=1}^{k} r_i I^i_{s,t}, \quad I^G_{s,t} = \sum_{i=1}^{k} g_i I^i_{s,t} \text{ and } I^B_{s,t} = \sum_{i=1}^{k} b_i I^i_{s,t}. \tag{10}$$

In the remainder of the paper, the color video sequence is called RGB. The multispectral video sequence with 7 spectral bands per frame is called $7B$ while the multispectral video sequence with the 6 spectral bands in the visible is called $6B$.

In order to propose fair quantitative results, it is required to annotate the video dataset at pixel resolution. However, because such a precise annotation is very time consuming, most datasets in the litterature annotate moving objects with bounding boxes. The *changedetection* dataset [7] has been the first comprehensive dataset that contains pixel precision ground truth labels. In this work, we propose the first dataset that is composed of several registered multispectral and color image sequences. These videos have been manually annotated at pixel resolution. We have chosen to use the same labels than those used in the *changedetection* dataset, *i.e.*:

- static pixels are assigned grayscale value of 0,
- moving pixels are assigned grayscale value of 255,
- pixels outside the Region Of Interest (ROI) are assigned grayscale value of 85,
- unknown pixels are assigned grayscale value of 170.

In total, more than 7400 frames have been manually annotated with these labels. To the best of our knowledge, this is the first dataset of multispectral video sequences and it is available to the research community with other non annotated videos at the project webpage.

The goal of this study is to measure how well a background subtraction algorithm can differenciate subtle differences of color and handle illumination variation. Consequently, animated backgrounds are simply not considered in this study and are labelled outside of the ROI. We present in Fig. 3 two examples of annotation.

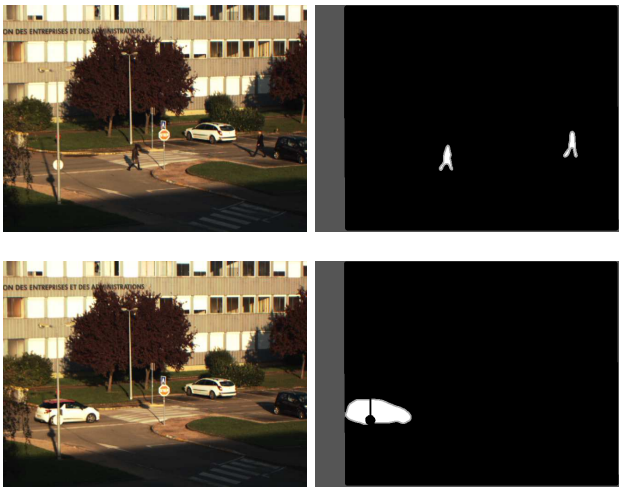

Fig. 3. Example of annotation with cars and pedestrians. The left part of the video contains animated texture (tree shaken by the wind) and has been labelled outside of the ROI.

B. Evaluation metrics

The performance of BS is evaluated at a pixel-level. Given the ground truth data, the correctness of foreground segmentation is measured in terms of *precision*, *recall*, and their harmonic mean *F-measure*:

$$\text{precision} = \frac{\#\text{correctly classified foreground pixels}}{\#\text{pixels classified as foreground}}$$

$$\text{recall} = \frac{\#\text{correctly classified foreground pixels}}{\#\text{ground truth foreground pixels}}$$

$$F_{measure} = 2\frac{\text{recall} * \text{precision}}{\text{recall} + \text{precision}}$$

By definition, a good BS algorithm is one producing simultaneously a small number of false positives and false

Fig. 2. Five specral bands extracted from a multispectral image.

negatives, *i.e.* both a high precision and recall value. The *F-measure* combines precision and recall and make the analysis of results easier.

V. Experimental results

Results are presented in Table I. For each video sequence, we present results for the color video sequence (RGB), the visible multispectral video ($6B$) and the visible plus near-infrared multispectral video sequence ($7B$). To simplify the analysis of results, only the maximum *F-measure* obtained for each video is presented.

BS with the Mahalanobis distance
The objective of this first experiment is to quantify the advantages of using multispectral video to perform background subtraction. The Mahalanobis distance d_M described in Section III-A has been used to perform background subtraction on the video dataset.

As one would expect, it is shown in Table I that the performance is improved using the multispectral videos $6B$ or $7B$. $7B$ obtains the best results on all videos. BS algorithms are significantly more efficient on Video 1 using the multispectral information since *F-measure* increases by approximately 0.15. This result is confirmed on the other sequences even if the difference is less important. Actually, the difference between $6B$ and $7B$ on the Video 1 suggests that it is the NIR spectral band that improves the most the results. It is important to highlight that Video 1 has been acquired indoor while the others have been acquired outdoor. This first result confirms the interest of using multispectral video to perform BS, however significant differences can be observed depending on the video sequence.

Combination of BS results from several spectral bands
The Pooling column of Table I presents results of this second experiment described in Section III-B. Although the performances are below the two other approaches, results suggest that only a few number of band is enough to reach the maximum *F-measure* with this method. This indicates that only a few spectral bands actually define the moving objects. This assertion must be verified and investigated deeper analyzing object reflectance of typical moving objects (mainly human and cars). In Table I, one can observe such an indicator with in parenthesis the value ρ that represents the definition of majority in Eq. 6.

BS with spectral distance metric
This third experiment evaluates the interest of using metrics that are designed for multispectral data to perform BS. In Table I, we present results with the spectral angle d_θ and the spectral information divergence d_{SID} described in Section III-C.

First, one can observe that whatever the similarity metric or the distance that is used, performance of BS algorithms on multispectral videos is better. For each video sequence and each metric, results obtained with RGB are lower than those with $6B$ or $7B$. $6B$ is rarely better than $7B$, only on Video 1 using the spectral measures d_θ and d_{SID}. This observation is consistent with the conclusions of the first experiment. Then, this third experiment shows that results with spectral metrics are surprisingly more heterogeneous. d_θ and d_{SID} are clearly more effective than d_M on videos 1, 2 and 3 but underperform d_M on videos 4 and 5. These differences are probably due the spectral characteristics of moving objects in these videos.

Even if these two spectral measurements are conceptually very different, results in Table I show that their behavior is quite similar. d_{SID} is a little more effective than d_θ but both d_θ and d_{SID} vary in the same way with respect to d_M. d_θ extracts geometric features between two spectra measuring the angle. Consequently, it is invariant to the norm of the vectors and it only considers their relative direction. Concerning d_{SID}, the normalisation described in Eq. 8 also implies invariance to the length of the vector. On the contrary, the Mahalanobis distance d_M measures the relative direction and length of the vectors.

VI. Conclusion

In this paper, we have presented the use of multispectral cameras for background subtraction. Until now, mostly mono or trichromatic cameras within the visible (Gray-level and RGB) or NIR part of the sectrum were used in video analytics. Recent advances in multispectral imaging technologies give the possibility to record multispectral videos. In this paper, we have quantitatively compared BS algorithms on regular color videos and on multispectral videos. We have shown that whatever the similarity metric or the distance that is used, performance of BS algorithms on multispectral videos is better on each video of the dataset. Three ways of using multispectral videos to perform BS are also evaluated. We have investigated the use of the well-known Mahalanobis distance, the combination of BS results from several spectral bands and two others spectral measurements. The results obtained with the spectral metrics strongly depend on the characteristics of the video. The first dataset of multispectral videos annotated at the pixel resolution has been established for this project.

We are convinced that the arrival on the market of these new multispectral cameras will change radically video surveillance and will enable new applications. We have shown that the extension of traditional methods on multispectral videos gives good results. However, the use of methods dedicated to these new data should allow to go even further.

Multispectral Sequence		Mahalanobis distance RGB	Mahalanobis distance 6B	Mahalanobis distance 7B	Pooling 7B (ρ)	Spectral angle RGB	Spectral angle 6B	Spectral angle 7B	SID similarity RGB	SID similarity 6B	SID similarity 7B
video 1	precision	0.6536	0.7075	0.7850	0.7475 (1)	0.9668	0.9598	0.8502	0.9342	0.9571	0.8537
	recall	0.6376	0.7684	0.8377	0.8568 (1)	0.8167	0.9688	0.9656	0.8919	0.9721	0.9566
	F-score	0.6536	0.7367	**0.8105**	0.7984 (1)	0.8854	**0.9642**	0.9042	0.9126	**0.9645**	0.9022
video 2	precision	0.8346	0.8435	0.8549	0.8639 (3)	0.9708	0.9260	0.9236	0.9763	0.9426	0.9748
	recall	0.9100	0.9138	0.9281	0.8997 (3)	0.8839	0.9816	0.9912	0.8653	0.9819	0.9626
	F-score	0.8707	0.8773	**0.8900**	0.8815 (3)	0.9253	0.9530	**0.9562**	0.9174	0.9619	**0.9686**
video 3	precision	0.7494	0.7400	0.7533	0.8809 (2)	0.8635	0.8373	0.8705	0.8683	0.8402	0.8672
	recall	0.5967	0.6288	0.6332	0.5134 (2)	0.8551	0.9375	0.9251	0.8361	0.9269	0.9263
	F-score	0.6644	0.6799	**0.6889**	0.6487 (2)	0.8593	0.8846	**0.8970**	0.8519	0.8814	**0.8958**
video 4	precision	0.8402	0.8462	0.8430	0.8146 (1)	0.6924	0.7195	0.6317	0.6017	0.6702	0.6454
	recall	0.7929	0.8100	0.8226	0.8654 (1)	0.4112	0.5343	0.7208	0.5142	0.5475	0.7361
	F-score	0.8158	0.8277	**0.8327**	0.8392 (1)	0.5747	0.6132	**0.6733**	0.5543	0.6027	**0.6878**
video 5	precision	0.7359	0.7299	0.7341	0.7373 (1)	0.6279	0.7320	0.6993	0.6139	0.7310	0.7139
	recall	0.7626	0.7938	0.8149	0.8066 (1)	0.6345	0.7153	0.7908	0.6204	0.7225	0.8204
	F-score	0.7490	0.7605	**0.7724**	0.7704 (1)	0.6312	0.7235	**0.7422**	0.6171	0.7267	**0.7574**

TABLE I. BACKGROUND SUBTRACTION RESULTS USING COLOR AND MULTISPECTRAL VIDEOS.

REFERENCES

[1] L. Miao, H. Qi, and W.E. Snyder, A generic method for generating multispectral filter arrays, *International Conference on Image Processing*, vol. 5, pp. 3343 3346, Oct. 2004.

[2] X. Wang, J-B. Thomas, J.Y. Hardeberg, and P. Gouton, Median filtering in multispectral filter array demosaicking, *Digital Photography IX*, vol. 8660 of Proc. SPIE, 2013.

[3] D. Rfenacht, C. Fredembach and S. Ssstrunk, "Automatic and Accurate Shadow Detection using Near-Infrared Information", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014.

[4] N. Gat, "Imaging spectroscopy using tunable filters: a review", *SPIE Wavelet Applications*, 2000.

[5] J. James, "Spectrograph Design Fundamentals", *Cambridge University Press*, 2007.

[6] C.O Conaire, N.E. O'Connor, E. Cooke and A.F. Smeaton, "Multispectral object segmentation and retrieval in surveillance video", *IEEE International Conference on Image Processing*, pp. 2381–2384, 2006.

[7] N. Goyette, P.M. Jodoin, F. Porikli, J. Konrad and P. Ishwar, "Changedetection.net: A new change detection benchmark dataset", *IEEE Workshop on Change Detection of the int. conf. on Computer Vision and Pattern Recognition*, 2012.

[8] Y. Dhome, N. Tronson, A. Vacavant, T. Chateau, C. Gabard, Y. Goyat and D. Gruyer, "A benchmark for Background Subtraction Algorithms in monocular vision: A comparative study", *int. conf. on Image Processing Theory Tools and Applications*, pp. 66–71, 2010.

[9] Y. Benezeth, P.-M. Jodoin, B. Emile, H. Laurent and C. Rosenberger, "Review and evaluation of commonly-implemented background subtraction algorithms", *IEEE International Conference on Pattern Recognition*, 2008.

[10] T. Bouwmans, F. El Baf and B. Vachon, "Statistical Background Modeling for Foreground Detection: A Survey", *Handbook of Pattern Recognition and Computer Vision, World Scientific Publishing*, pp. 181–199, vol.4(2), 2010.

[11] S. Brutzer, B. Hoferlin and G. Heidemann, "Evaluation of Background Subtraction Techniques for Video Surveillance", *IEEE int. conf. on Computer Vision and Pattern Recognition*, pp. 1937–1944, 2011.

[12] C. Wren, A. Azarbayejani, T. Darrell and A.P. Pentland, "Pfinder: Real-time tracking of the human body", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 780–785, vol.19(7), 1997.

[13] C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking", *IEEE int. conf. on Computer Vision and Pattern Recognition*, 1999.

[14] Z. Zivkovic and F. V. D. Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction", *Pattern Recognition Letters*, vol.27, pp. 773–780, 2006.

[15] D. Lee, "Effective gaussian mixture learning for video background subtraction", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.27, pp. 827–832, 2005.

[16] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction", *European Conference on Computer Vision*, pp. 751–767, 2000.

[17] O. Barnich and M. V. Droogenbroeck, "Vibe: A universal background subtraction algorithm for video sequences", *IEEE Trans. on Image Processing*, vol.20(6), pp. 1709–1724, 2011.

[18] Y. Nonaka, , A. Shimada, H. Nagahara, and R. Taniguchi, "Evaluation report of integrated background modeling based on spatio-temporal features", *IEEE Workshop on Change Detection of the int. conf. on Computer Vision and Pattern Recognition*, pp. 9–14, 2012.

[19] A. Tavakkoli, M. Nicolescu, G. Bebis, and M. Nicolescu, "Non-parametric statistical background modeling for efficient foreground region detection", *Machine Vision and Applications*, vol.20, pp. 395–409, 2008.

[20] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation", *IEEE int. conf. on Computer Vision and Pattern Recognition*, pp. 302–309, 2004.

[21] H. Wang and D. Suter, "A consensus-based method for tracking: Modelling background scenario and foreground appearance", *Pattern Recognition*, vol.40(3), pp. 1091–1105, 2007.

[22] M. V. Droogenbroeck and O. Paquot, "Background subtraction: Experiments and improvements for vibe", *IEEE Workshop on Change Detection of the int. conf. on Computer Vision and Pattern Recognition*, 2012.

[23] M. Wu and X. Peng, "Spatio-temporal context for codebook-based dynamic background subtraction", *International Journal of Electronics and Communications*, vol.64, pp. 739–747, 2010.

[24] N. Oliver, B. Rosario, and A. Pentland, "A bayesian computer vision system for modeling human interactions", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.22(8), pp. 831–843, 2000.

[25] M. Seki, T. Wada, H. Fujiwara, and K. Sumi, "Background subtraction based on cooccurrence of image variations", *IEEE int. conf. on Computer Vision and Pattern Recognition*, 2003.

[26] D.-M. Tsai and S.-C. Lai, "Independent component analysis-based background subtraction for indoor surveillance", *IEEE Trans. on Image Processing*, vol.18, pp. 158–167, 2009.

[27] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications", *IEEE Trans. on Image Processing*, vol.17(7), pp. 1168–1177, 2008.

[28] L. Li, W. Huang, I. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection", *IEEE Trans. on Image Processing*, vol.13(11), pp. 1459–1472, 2004.

[29] C.-I. Chang, "An information theoretic-based approach to spectral variability, similarity and discriminability for hyperspectral image analysis", *IEEE Trans. on Information Theory*, vol. 46(5), pp. 1927–1932, 2000.

[30] R.A. Schowengerdt, "Remote Sensing: Models and Methods for Image Processing", *2nd ed. San Diego, CA: Academic*, 1997.

[31] N.P.Jacobson, M.R. Gupta, "Design goals and solutions for display of hyperspectral images", *IEEE Trans. on Geoscience and Remote Sensing*, vol.43(11), pp.2684–2692, 2005.