# Color Invariant Study for Background Subtraction

Lorena Guachi, Giuseppe Cocorullo, Pasquale Corsonello, Fabio Frustaci, Stefania Perri
Department of Electronics, Computer Sciences and Systems
DIMES - University of Calabria
Arcavacata di Rende, Italy
e-mail: loreanggeles@hotmail.com, g.cocorullo@unical.it, p.corsonello@unical.it, ffrustaci@deis.unical.it, perri@dimes.unical.it

*Abstract*—**Effectiveness detection to extract objects of interest is a fundamental step in many computer vision systems. In real solutions, the accurate Background Subtraction (BS) is a challenge due to diverse and complex background types. Being the color widely used as descriptor to improve accuracy in several BS algorithms, in this paper we analyze four Color Invariants (CIs) based on the Kubelka-Munk theory combined with Gray scale. The capability of several CIs combinations in segmenting foreground is evaluated referring to five video sequences. This experimental study provides a point-of-view to choose the best color combination considering accuracy and the channel numbers which can be applied for image segmentation. The results demonstrate that the combination of the color invariant H with Gray scale achieves higher performance for foreground segmentation for both indoor and outdoor video sequences. Furthermore, it uses the minimum number of color channels.**

***Keywords-image processing; background subtraction; color invariant.***

## I. INTRODUCTION

In the recent past, Background Subtraction (BS) has gained an extensive application as a fundamental pre-processing task of video systems especially to detect objects of interest (vehicles, people, animals and so on) for security, traffic monitoring, surveillance systems among others, which include people counting, intrusion detection and tracking [1][2].

The BS algorithms typically use five features as descriptor: color, edge, stereo, motion and texture features [2], each one having specific characteristics to handle different environments and critical situations as motion changes, viewing direction, structure background and illumination changes. In order to be more robust in presence of critical situations, some algorithms tend to combine different features. The multi-scale region BS algorithm [21] performs the Gaussian Mixture modeling in conjunction with color histograms, texture information, and consecutive division of image regions to detect efficiently edges of the moving objects. Also in [22], the use of color and edge information is applied to handle slow illumination changes and camera noise, being able to run on standard platform for real time applications.

However, the capability of segmenting moving objects from video sequences is even a challenge in vision systems, where many algorithms work for specific environments in very controlled situations.

The rest of this paper is organized as follows. Section II describes most relevant related works. The color invariant descriptors are introduced in Section III. Section IV describes the algorithm used for the evaluation. The experimental results are presented in Section V. Finally, in Section VI we discuss the conclusion.

## II. RELATED WORKS

Color features are widely used in many algorithms and the ability to be very discriminative is mainly related to the way of representing colors in the image. Different color spaces provide different accuracies [3], due to several limitations in the presence of shadows, illumination changes, and camouflage.

Several works have been proposed in order to determine which color space is best for shadow detection and BS. In [4], an experimental study is presented to show how the "RGB", "XYZ", "YCrCb", "HSV" and the normalized "rgb" formats differently affect the moving objects classification and the shadow detection. For identifying shadow edges, the color opponent space is exploited in sunlit scenes. Furthermore, it is shown that the different uniform color opponent space is the most suitable for indoor environments, and the LαB is an appropriate selection for both indoor and outdoor environments [5].

The influence of several color spaces on the shadow detection has been evaluated in [6]. That work demonstrated that CIE L*u*v color model allows moving objects to be extracted efficiently also in the presence of moving shadows. As demonstrated in [7] the color space also affects tracking methods. As an example, in tracking applications, YCbCr and HSV color models are more suitable than RGB and Grayscale color models [7].

With the main objective of improving the achieved performances and avoiding time consuming color transformation, in [9], YUV color model is exploited for shadow detection in video conference applications. On the other hand, the statistical BS algorithm proposed in [10], separates the brightness from the chromaticity component in a pixel to exploit a computational color space.

Photometric color invariants as normalized rgb, hue (H), saturation (S), l1l2l3 and c1c2c3 are functions that describe the color configuration discounting shadows, highlight and shading. These functions are invariant to surface orientation, viewing direction and illumination conditions [8]. C1c2c3 model is adopted in [11] to exploit the spectral and geometrical characteristics for automatically shadow detection for static images and video sequences. This

 ISBN: 978-1-61208-496-1 

approach firstly hypothesizes the presence of shadows, considering some initial evidence based on the fact that shadows darken the surface which they are cast upon.

Although many works presented in the literature have demonstrated how the color features interfere in the achieved accuracy, typical descriptors are based on specifically spectral information. On the contrary, the CIs are derived from a physical model and can take into account color spectral information and color spatial structure. Therefore, focused on CIs of the Kubelka-Munk theory, in [12] the CIs H, Wx, and Wy are introduced as descriptors in a novel BS algorithm to segment several video sequences in color similar situations. A first approach to combine the CIs with different color models is introduced in [13], where particularly CI H is used in conjunction with Gray scale to build robust descriptors with the aim of reducing post-processing task. Hence, this paper evaluates the possibility of combining the complete set of CIs (H, N, C and W) with Gray scale information. Several combinations are referenced to demonstrate that the efficiency in extraction of moving objects depends on the descriptors selected and combined through different logic operators. Overall results are presented for both indoor and outdoor experimental environments.

## III. COLOR INVARIANT DESCRIPTOR

This section presents the fundaments of CI and Gray color space. Since the color features are often very discriminative, many BS approaches use the color as descriptor, but in certain environments it has several limitations in the presence of camouflage, shadows and illumination changes. However, the combination with other features allows achieving more robust solutions for the BS [2].

Any method for describing CI model relies on assumptions about the physical variables involved and on photometric configuration [14]. Photometric CI are characterized as a function of surface reflectance, illumination spectrum and the sensing device, which consider the spatial configuration of color, and also the color spectral energy distribution coding color information [12]. Color spaces with properties independent of illumination intensity, reflectance property, viewing direction, and object surface orientation are defined as the color invariants [8]. These properties characterize the image color configuration discounting highlights, shadows, noise and shading. As an example, the Gaussian color model with spectral and spatial parameters is exploited in [12] to define a framework for the robust measurement of colored object reflectance. The CIs are derived from a physical reflectance model based on the Kubelka-Munk theory for colorant layers [14], where, illumination and geometrical invariant properties depend of the use of reflectance model.

The invariants are useful for materials as dyed paper and textiles, paint films, opaque plastics, dental silicate cements and up to enamel. A set of CIs derived from Kubelka-Munk theory is listed in TABLE 1. The latter shows how computing the CIs named H, N, C, and W, with E, $E\lambda$ and $E\lambda\lambda$ being

TABLE 1. SET OF COLOR INVARIANTS

| *CI* | *Definition* |
|---|---|
| H | $E\lambda$ / $E\lambda\lambda$ |
| N | $(E\lambda\chi \times E - E\lambda \times E\chi)$ / $(E \times E)$ |
| C | $E\lambda$ / E |
| W | $E\chi$ / E |

the spectral differential quotients based on the scale-space theory [15].

The above defined CIs can be combined incrementally to achieve an alternative to invariant features extraction [14]. The Gray color space model is based on the brightness information and uses the measurement of amount of light (intensity). It is applied for object tracking often on a blob or a specific region [7]. However, taking into account that the color furnishes more information on the objects in a scene, it would be expected that this model can be used in conjunction with other models to achieve more robust solutions and higher accuracy than the basic separated models. For this reason, the Gray color space is included in the study here presented with the additional advantage of using a color space that does not require complex color transformations.

## IV. BACKGROUND SUBTRACTION ALGORITHM

The main computational steps required to classify the foreground pixels by using CIs can be summarized as follows: 1) RGB input frames are processed to obtain the CIs; 2) the background model is initialized by collecting, as the historical frames, the CIs obtained for the first *Nf* frames and the current background is computed; 3) as soon as the (*Nf+1*)-th frame is acquired, the foreground detection initiates and it is executed pixel-by-pixel by comparing the CIs of the current pixel the CIs of historical frames; 4) the current background model is updated taking into account the obtained classification.

To study the performance of CIs defined in Section III, we reference the algorithm schematized in Figure 1, which uses only ten historical frames. Some evaluated combinations of the features selected include a channel with Gray scale information whereas others are compounded only by CIs. Each channel is analyzed separately computing the percentage variation between the current frame and the historical mean. To classify the pixels within the generic frame of a video sequence into the background and the foreground sets, a thresholding is performed for each adopted descriptor. In our study, we refer to H, W, N, C and Gray scale components with the threshold values Th=55, Tw=90, Tn=90, Tc=90, Tg=60 that have been set experimentally to the values for which the number of wrong classified pixels is minimized for typical benchmark video sequences [17-20]. Several tests have demonstrated that higher threshold values reduce the accuracy in detecting foreground pixels, whereas smaller values increase the noise

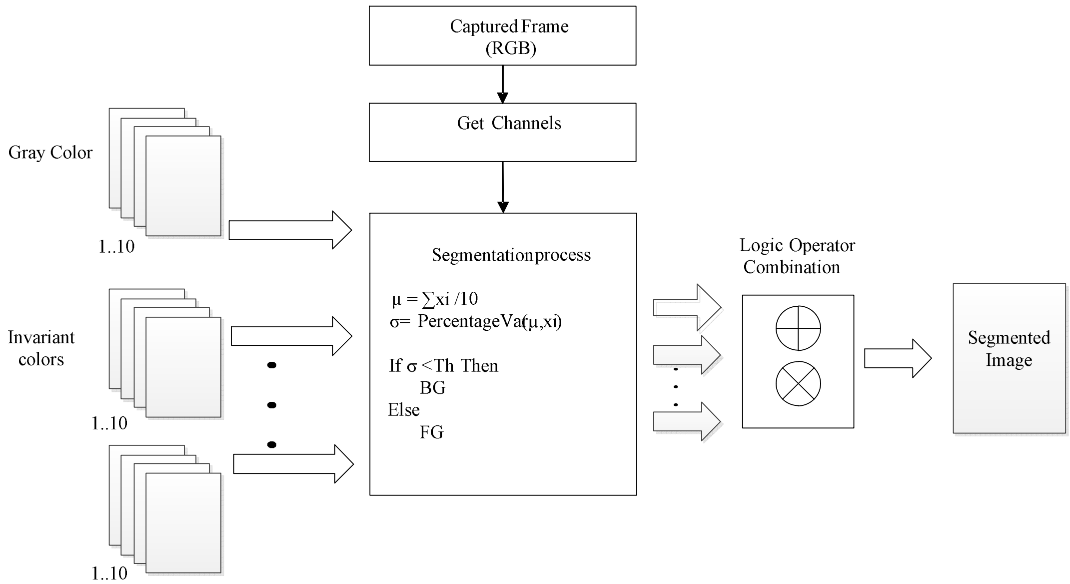


Figure 1. Background Subtraction diagram overview.

sensitivity. As suggested in [3], *Nf*=10 is initially used to model the background with a single Gaussian distribution. Several experiments demonstrated that increasing *Nf* does not significantly improve the accuracy but increases the memory requirement. Each component of the generic pixel of the current frame is compared to the mean value computed for the correspondent channel of historical frames. When the difference between the current examined channel and the historical mean overcomes the relative threshold, the current component is classified as belonging to a foreground pixel. Otherwise it is recognized as the component of a background pixel. Partial results obtained separately from the examined channels are then combined through appropriate logic operators to obtain the final segmented images. Background model is updated introducing a new frame at position zero, and discarding the oldest frame of position nine, all frames are sorted after each analysis.

## V. EXPERIMENTAL RESULTS

C++ software routines have been on purpose implemented to evaluate twenty three color combinations. Experimental tests have been done on different video sequences, related to both indoor and outdoor environments and the achieved performances are measured in terms of recall (Rec), specificity (Sp), precision (Pr), and percentage of correctly classified pixels (PCC). Rec measures the accuracy of the approach at the pixel level with a low False Negative Rate;

TABLE 2. COMPARISON RESULTS

| *Combination* | *Rec* | *Sp* | *Pr* | *PCC* |
|---|---|---|---|---|
| H AND GRAY | 11.13 | 99.77 | 81.82 | 93.88 |
| H OR GRAY | 52.65 | 89.87 | 27.61 | 87.50 |
| H AND N | 13.58 | 98.31 | 33.87 | 92.68 |
| H OR N | 54.60 | 82.18 | 18.08 | 81.74 |
| (H OR N) AND GRAY | 13.19 | 99.72 | 81.13 | 93.98 |
| (H OR N) OR GRAY | 59.34 | 82.10 | 19.27 | 80.68 |
| H AND C | 19.95 | 96.17 | 29.22 | 91.07 |
| H OR C | 50.09 | 85.77 | 26.24 | 83.40 |
| (H OR C) AND GRAY | 15.08 | 99.07 | 79.21 | 93.44 |
| (H OR C) OR GRAY | 56.93 | 89.11 | 27.61 | 87.13 |
| H AND W | 9.27 | 98.67 | 31.63 | 92.79 |
| H OR W | 59.79 | 76.14 | 15.44 | 76.41 |
| (H OR W) AND GRAY | 13.70 | 99.72 | 81.33 | 94.03 |
| (H OR W) OR GRAY | 64.02 | 76.06 | 16.26 | 75.30 |
| H OR N OR C | 64.08 | 75.50 | 15.97 | 74.77 |
| (H OR N OR C ) OR GRAY | 68.49 | 70.14 | 14.34 | 70.05 |
| (H OR N OR C) AND GRAY | 15.09 | 99.70 | 81.60 | 94.13 |
| H OR N OR W | 65.31 | 70.20 | 13.78 | 69.84 |
| (H OR N OR W) AND GRAY | 14.76 | 99.70 | 81.15 | 94.09 |
| (H OR N OR W) OR GRAY | 66.92 | 75.44 | 16.47 | 74.93 |
| H OR N OR C OR W | 68.76 | 69.63 | 14.19 | 69.59 |
| (H OR N OR C OR W) AND GRAY | 15.74 | 99.68 | 81.33 | 94.16 |
| (H OR N OR C OR W) OR GRAY | 70.95 | 69.59 | 14.54 | 69.72 |

Sp stimulates combinations with a low False Positive Rate; Pr favors combinations with a low False Positive Rate, and PCC measures the percentage of correct classifications [17]. The overall results are summarized in Table 2. The first column shows the logic operation applied to classify foreground pixels. As an example, the combination (H OR W) AND GRAY detects the generic pixel as foreground only if either its component H or its component W belongs to a foreground pixel, and also its Gray scale data is associated to a foreground pixel.

The results presented in Table 2 show that, as expected, differently combining CIs with Gray scale data very different accuracy can be achieved in detecting foreground objects. It is worth pointing out that the number of channels used to achieve a given accuracy significantly affects the computational complexity. In Figure 2. , the average accuracy obtained with each combination is directly related to the number of channels involved. Based on numeric analysis we can see that the combination (*H* OR *N* OR *C* OR *W*) AND *GRAY* achieves the best accuracy for indoor and outdoor experimental environments, and focused on the number of channels, the set of *H AND GRAY* reaches good performance on average with the minimum number of color channels. Figure 3. shows some of the segmented images obtained with these two combinations.

The results depicted in Figure 4 show the benefits achieved by introducing Gray scale in the set of CI combination to reduce the noise and improve the accuracy.

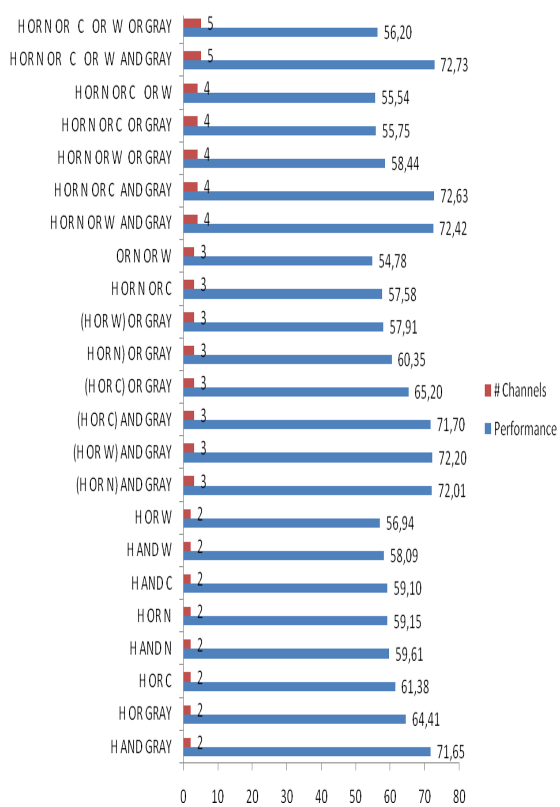


Figure 2. Analysis of the adopted combinations.

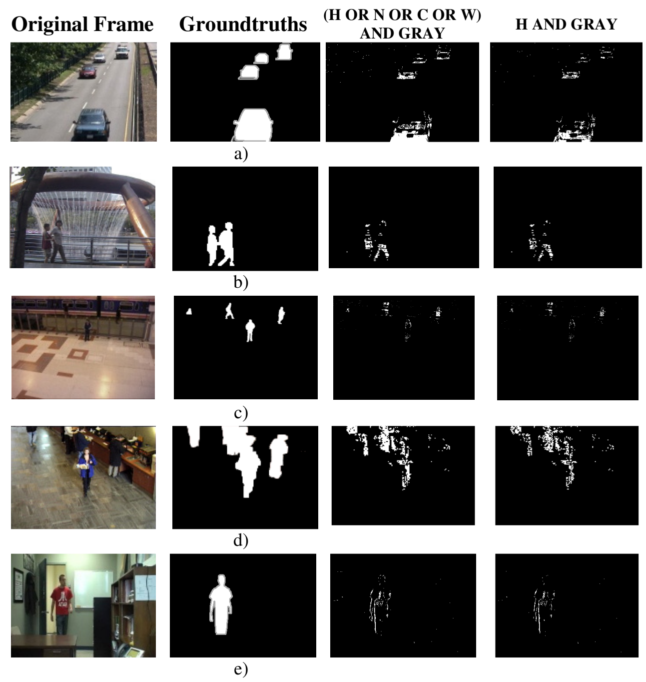


Figure 3. Results related to: a) Highway; b) Fountain; c) Pets2006; d)Bootstrap; e)Office.

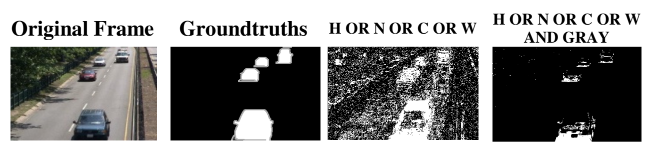


Figure 4. Results obtained introducing Gray scale information

All the above results have been obtained through the hardware system illustrated in Figure 5. The latter is based on the Raspberry Pi equipped with a camera module, able to capture RGB and grayscale images, and a Broadcom BCM2835 system on chip, consisting of an ARM1176JZF-S 700 MHz processor, a VideoCore IV GPU, 512 MB of RAM, and an SD card for long term storage and booting.

Figure 5. The hardware system used for tests

## VI. CONCLUSION

This paper has empirically compared the suitability of sets of CI combinations. Some of them include Gray scale. The tests measured the performance of the combinations

 ISBN: 978-1-61208-496-1

referring to indoor and outdoor experimental environments, demonstrating that the Gray scale insertion mitigates the problem of misclassified pixel. H and Gray scale combination provides the highest performance with respect to other combinations with the benefit of include particularly only two channels. Gray color model leads to background with less noise. On the contrary, CIs increase the noise due to the transformation operations, but, the combination with Gray color space allows achieving high effectiveness in the BS. These characteristics can be efficiently introduced in the algorithms for the image segmentation.

## REFERENCES


[1] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," Computer Vision and Image Understanding, vol. 122, pp. 4-21., 2014.

[2] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," Computer Science Review, vol. 11, pp. 31 – 66, 2014.

[3] H. Zhou, Y. Chen, and R. Feng, "A novel background subtraction method based on color invariants," Computer Vision and Image Understanding, vol. 117, no 11, pp. 1589–1597, 2013.

[4] P. Kumar, K. Sengupta, and A. Lee, "A comparative study of different color spaces for foreground and detection for traffic monitoring system," In Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on. IEEE, pp. 100-105, 2002.

[5] E. Khan and E. Reinhard, "Evaluation of color spaces for edge classification in outdoor scenes," In Image Processing, 2005. ICIP 2005. IEEE International Conference on. IEEE,, pp. III-952-5, 2005.

[6] C. Benedek and T. Szirányi, "Study on color space selection for detecting cast shadows in video surveillance," International Journal of Imaging Systems and Technology, vol. 17, no 3, pp. 190-201, 2007.

[7] P. Sebastian, Y. Vooi, and R. Comley, "Colour space effect on tracking in video surveillance," International Journal on Electrical Engineering and Informatics, vol. 2, no 4, pp. 298-312, 2010.

[8] T. Gevers and A. W. M. Smeulders, "Color-based object recognition," Pattern Recognition, vol. 32, no 3, pp. 453–464, 1999.

[9] O. Schreer, I. Feldmann, U. Goelz, and P. Kauff, "Fast and robust shadow detection in videoconference applications," in Video/Image Processing and Multimedia Comuunications 4th EURASIP-IEEE Region 8 International Symposium on VIPromCom. IEEE, , pp. 371–375, 2002.

[10] T. Horprasert, D. Harwood, and L.S. Davis, "Statistical approach for real-time robust background subtraction and shadow detection," In Proceedings IEEE International Conference on Computer Vision, , pp. 1-19, 1999.

[11] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features," Computer vision and image understanding, vol. 95, no 2, pp. 238-259, 2004.

[12] H. Zhou, Y. Chen, and R. Feng, "A novel background subtraction method based on color invariants," Computer Vision and Image Understanding, vol. 117, no 11, pp. 1589–1597, 2013.

[13] L. Guachi, G. Cocorullo, P. Corsonello, F. Frustaci, and S. Perri, "A novel background subtraction method based on color invariants and grayscale levels," In Security Technology (ICCST), 2014 International Carnahan Conference on. IEEE,, pp. 1-5, 2014.

[14] J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and H. Geerts, "Color invariance," Pattern Analysis and Machine Intelligence, IEEE Transactions On, vol. 23, no 12, pp.1338-1350,2001.

[15] L. M. J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, and M.A. Viergever, "Scale and the Differential Structure of Images," Image and Vision Computing, vol. 10, no. 6, pp. 376-388, 1992.

[16] R. Gonzalez and R. E. Woods, "Digital Image Processing using Matlab, " 2004: Pearson Prentice Hall.

[17] N. Goyette, P. M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection. net: A new change detection benchmark dataset," In Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on. IEEE, pp. 1-8, 2012.

[18] Statistical Modeling of Complex Background for Foreground Object Detection. [Online]. Available from: http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html. [accessed July 2016]

[19] Test Images for Wallflower Paper. [Online]. Available from: http://research.microsoft.com/en-us/um/people/jckrumm/wallflower/testimages.htm. [accessed July 2016]

[20] ChangeDetection.NET (CDNET). [Online]. Available from: http://www.changedetection.net. [accessed July 2016]

[21] P. D. Z. Varcheie, M. Sills-Lavoie, and A. Bilodeau, "A multiscale region-based motion detection and background subtraction algorithm," Sensors, vol. 10, no 2, pp. 1041-1061, 2010.

[22] S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld, "Detection and location of people in video images using adaptive fusion of color and edge information," In Pattern Recognition, 2000. Proceedings. 15th International Conference on. IEEE, pp. 627-630, 2000.