IET Image Processing
DOI: 10.1049/iet-ipr.2014.0149

Linear colour correction for multiple illumination changes and non-overlapping cameras

J. Torres*[1], K. Schutte[2], H. Bouma[2], and J. M. Menéndez[1]

[1]Grupo de Aplicación de Telecomunicaciones Visuales, Universidad Politécnica de Madrid, Madrid, Spain
[2]Intelligent Imaging Group, TNO, The Hague, The Netherlands

Abstract: *Many image processing methods, such as techniques for people re-identification, assume photometric constancy between different images. This paper addresses the correction of photometric variations based upon changes in background areas to correct foreground areas. We assume a multiple light source model where all light sources can have different colors and will change over time. In training mode we learn per-location relations between foreground and background color intensities. In correction mode we apply a double linear correction model based on learned relations. This double linear correction includes a dynamic local illumination correction mapping as well as an inter-camera mapping.*

We evaluate our illumination correction by computing the similarity between two images based on the Earth Mover's Distance. We compare the results to a representative auto-exposure (AE) algorithm found in the recent literature plus a color correction one based on the inverse-intensity chromaticity. Especially in complex scenarios our method outperforms these state-of-the-art algorithms.

1 Introduction

Retrieving objects seen in earlier images is an active research topic within image processing. In very controlled environments the assumption will hold that the pixel values of an object are comparable between past and present images, but in general this is not the case due to illumination variations. Within this paper we consider conditions with multiple illumination sources, each with a different color and a space-time variant intensity. In effect, the illumination will be different for every position, time, and color channel. In addition, we consider a non-overlapping multi-camera set-up. The challenge addressed in this paper is to correct under such conditions images of the same objects such that their corresponding pixels are comparable, to allow re-identifying those objects along time and space.

Three different strategies can be distinguished in literature to improve the performance of re-identification algorithms under varying imaging conditions:

1. Control the camera parameters during acquisition, by using auto-exposure (AE) or auto-white balance control.
2. Assume that the matching algorithm handles the imaging condition changes. For instance, using intensity invariant color descriptors.
3. Correct the images using image processing after acquisition but before applying the matching algorithm.

Table 1 summarizes the main relevant approaches.

Table 1: Color correction approaches. First row refers to the first strategy and the remainders to the third strategy while the second strategy is not dealt with in this paper.

Based on	References
Acquisition control	[30, 22, 18]
Color descriptors	[2]
Color constancy	[1, 13, 8, 25, 31]
Camera Response Function	[12, 33, 21, 27]
Inter Camera Color Response	[23, 16, 20, 10, 24, 5, 19, 28, 4]

The first strategy is widely applied on current available cameras [22, 30]. These methods control the camera parameters during acquisition to compensate for the

*jta@gatv.ssr.upm.es

light variations and obtain suitable images for viewing or processing purposes. Nevertheless, with multiple light sources, each with different spatial effect, such approaches are unable to effectively correct at every position within an image because they do not allow for location dependent corrections. Liu [18] goes far beyond the classic approach because their division into regions provides better knowledge of the light behavior, making the correction more accurate.

The second strategy is often based on object descriptors able to match two images of the same object in different pose and light conditions. Such descriptors [2] may include information about color, location, structure, texture or a combination of several of these. Using invariant features may reduce the amount of information; for example in an intensity normalized image – a common descriptor – it will be very hard to separate grey clothes of people from a black background. These kinds of techniques are hence avoided in this paper.

The third strategy, including this work, may be used on its own or in combination with the previously explained. They are mainly based on knowledge about image formation and surface reflectance models and consist of designing computational color constancy methods that "*estimate the chromaticity of the light source and then correct the image to a canonical illumination using the diagonal model*" [9]. This *diagonal model*, introduced by Finlayson [7] as a Diagonal-Matrix Transform (DMT) based on the von Kries' model, consists of a linear transformation for each of the color channels. By means of the DMT, the color appearance under several light conditions can be maintained. A complete survey about color constancy algorithms can be read in [8]. Nevertheless, these methods have high computational cost and require complex calibrations using known illuminants, causing these techniques to fail when they are unknown. Furthermore, even though there are several attempts to deal with real-images, their accuracy [1] – or their application to complex scenarios [31, 13, 25] – are not sufficiently demonstrated.

Other research lines model the Camera Response Function (CRF) [12] which consists of establishing the relation between the amount of light collected by the camera sensor and the digital intensity values. Using these models, several correction algorithms may be applied by comparing current pixel intensities to reference or previous values. For instance, Withagen [33] developed a simple and fast correction method by presenting different ways to estimate a global intensity correction factor to correct the input images. This method has been recently complemented by Sayed [27], who added a local correction using the mean and standard deviation variation. Nevertheless, this calculation is very sensitive to noise and introduces blurring artifacts in the image. Inspired by Grossberg and Nayar's work [12], Parameswaran [21] implemented an illumination compensation method based on the fact that the order of the pixel intensities is maintained when the illumination changes. When the illumination change is local, several functions must be estimated, which increases the computational cost. Furthermore, for our purpose, these methods are not insufficient because they do not deal with multi-camera systems.

The classic color constancy algorithms and CRF's were the basis for the first designs of methods that correct photometric differences in non-overlapping, multi-camera architectures. Basically, the Inter Camera Color Response (ICCR) estimates [23] are based on the assumption that the appearance of an object under distinct illumination conditions will be the same when their respective cumulative histograms are equal. An initial approach from Javed [16] utilized a supervised method performing a probabilistic Principal Component Analysis using the normalized-bin histograms of the objects for obtaining the Brightness Transfer Functions (BTFs) between cameras. Several works have been derived from Javed's. They mainly focused on making the method more robust against illumination changes using several frames of the same person for the estimate [20]; defining improved BTFs [24, 5]; or implementing non-supervised methods [10, 28]. The current ones also include spatio-temporal relationships among cameras [4, 19]. However, ICCR techniques require complex calibration stages, the use of color checkers [1] or, in the non-supervised methods, a deeper knowledge of the moving objects.

This paper focuses on a color correction of foreground objects (persons) based upon observed variations in the background in multi-camera scenarios. This correction is composed of a double mapping. The first mapping corrects the local illumination variations while the second compensates the photometric changes between different cameras. Separation between foreground and background is typically performed by object detection and tracking algorithms, which are out of the scope of this paper ([3] presents a complete review of these methods). The correction process is fast in run-time, and it only requires a training mode that learns inter-camera and background-foreground relations using several people samples as we describe in the next sections. It also yields worthy results with real images. Color calibration or knowledge of the illuminants are not required. Advanced matching techniques are also out of the scope. Nevertheless, we made use of the Earth Mover’s Distance (EMD) [26] as a similarity measure for the experiments.

The remainder of the paper is organized as follows. In Section 2, our linear correction model is proposed including both the inter-camera as well as the local corrections. Section 3 details the proposed technique. In Section 4 we describe our experiments which show the good performance of our algorithm in a complex photometric real scenario compared to state-of-the-art auto-exposure [18] and color correction [29] algorithms. Finally, Section 5 summarizes conclusions and suggestions for future work.

2 Linear correction model

We start providing our notation and assumptions (Section 2.1). Then, we define our model and adapt it to our requirements (Section 2.2). After that, we propose the double linear correction mapping (Section 2.3).

2.1 Notation and assumptions

In the paper, we use the notation in Table 2.
Our approach has the following assumptions:

1. **Multiple dynamic sources:** All light sources can have different colors and will change over time.
2. **Lambertian:** The objects are mainly composed by Lambertian surfaces.
3. **Static surface response:** The photometric surface-responses S of an object are not dependent of either the illuminants or the camera response; remain similar with time; and are only dependent on the surface properties.
4. **Not crowded scenes:** As the correction is based on background information, very crowded scenes – where sufficient background is not always observed – are not allowed.
5. **Linearity:** The intensity is neither under nor over saturated and only linear digital processing is applied (ignoring, e.g. gamma correction). Camera’s settings hold constant. Although not every camera obeys these assumptions, many cameras do, as is demonstrated in [32].
6. **Independent channels:** Changes in the product of the illumination E and the camera response Q is within each band R,G,B independent of the surface reflectivity S. We expect this to hold for a wide range of light sources, cameras and materials.

2.2 Image formation model

The image formation model of image I is modelled by the camera response function CRF:

$$I_k(\vec{x},t) = CRF_k(R_k(\vec{x},t)) \tag{1}$$

where R is the amount of electrons generated by each sensor pixel. In this context k refers to the three (RGB) color channels. The definition of R_k can be extended to:

$$R_k(\vec{x},t) = \int_\lambda Q_k(\lambda)E(\lambda,\vec{x},t)S(\lambda,\vec{x},t)d\lambda \tag{2}$$

where the product $E \cdot S$ is the incident light arriving at pixel coordinates $\vec{x}$ in the sensor plane and Q is the camera sensitivity (similar to the model in [17]). If we assume that the CRF is linear, and we have independent channels (Assumptions 5 and 6 of Section 2.1), Equations (1) and (2) can be simplified to:

$$I_k(\vec{x},t) = G(t)S_k(\vec{x},t)\int_\lambda Q_k(\lambda)E(\lambda,\vec{x},t)d\lambda \tag{3}$$

where G models the in-camera processing such as exposure variations and gain. In this expression we see three parts: i) the product of G and Q which depends on the camera; ii) E which depends on the source of light, location and scenario; and iii) S which depends on the observed object surface.

In Equation (2) and (3), the illumination dependent term E is a sum of all i illumination sources, and thus, $E(\lambda,\vec{x},t) = \sum_i E_i(\lambda,\vec{x},t)$. For each illumination source we define a time varying illumination gain g_i and a constant illumination term EC_i, such that $E_i(\lambda,\vec{x},t) = g_i(t)EC_i(\lambda,\vec{x})$. As such we can rewrite the model in Equation (2) and (3) to:

$$R_k(\vec{x},t) = \sum_i g_i(t)\int_\lambda Q_k(\lambda)EC_i(\lambda,\vec{x})S(\lambda,\vec{x},t)d\lambda \tag{4}$$

$$I_k(\vec{x},t) = G(t)S_k(\vec{x},t)\sum_i g_i(t)\int_\lambda Q_k(\lambda)EC_i(\lambda,\vec{x})d\lambda \tag{5}$$

2.3 Linear correction mapping

Our goal is to obtain a similar digital image of the same object when the photometric conditions and the cameras change. For background areas, the photometric surface-responses of an object in two temporal instants remains constant ($S(\vec{x},t_{ref}) \approx S(\vec{x},t) = S(\vec{x})$, Ass. 3 of Section 2.1). Therefore two images $I_{j_{ref}}$ (from camera j_{ref}) and I_j (from

Table 2: Notation used in the paper.

Symbol	Description	Symbol	Description
λ	Wavelength	b	Referred to background (as a superscript)
E	Spectral irradiance of a light source	f	Referred to foreground (as a superscript)
G	In-camera processing gain	H	Normalized cumulative histogram
I	Pixel intensity in an image	$\hat{I}$	I corrected
$\bar{I}$	Average value of I	ΔI	Ratio of two values of I
i	Illumination source index	j	Camera index
k	Color channel index	N	Number of total locations
n	Location index (range 1 to N)	Q	Camera sensor sensitivity
R	Amount of electrons generated by each sensor pixel	S	Photometric response of a surface
t	Time	V	Number of total background regions
v	Background region index (range 1 to V)	$\vec{x}$	Pixel coordinates

camera j) of the same background area S_0 having different photometric conditions can be written (Equation (3)) as:

$$\left.\begin{array}{l} I_{k,j_{ref}}(\vec{x},t_{ref}) = G_{j_{ref}}(t_{ref})S(\vec{x})\int_\lambda Q_{k,j_{ref}}E_{j_{ref}}(\lambda,\vec{x},t_{ref})d\lambda \\ I_{k,j(}\vec{x},t) = G_j(t)S(\vec{x})\int_\lambda Q_{k,j}E_j(\lambda,\vec{x},t)d\lambda \end{array}\right\} \rightarrow$$

$$I_{k,j}(\vec{x},t) = \frac{G_j(t)\int_\lambda Q_{k,j}E_j(\lambda,\vec{x},t)d\lambda}{G_{j_{ref}}(t_{ref})\int_\lambda Q_{k,j_{ref}}E_{j_{ref}}(\lambda,\vec{x},t_{ref})d\lambda}\cdot I_{k,j_{ref}}(\vec{x},t_{ref}) \quad (6)$$

$$I_{k,j}(\vec{x},t) = \frac{I_{k,j_{ref}}(\vec{x},t_{ref})}{\Delta I_{k,j}(\vec{x},t)} \quad (7)$$

$\Delta I_{k,j}(\vec{x},t)$ is the k-band relation factor between a reference camera j_{ref} and time t_{ref} with a camera j in time t.

A corrected image $\hat{I}_{k,j}$ is written as:

$$\hat{I}_{k,j}(\vec{x},t) = \Delta I_{k,j}(\vec{x},t)\cdot I_{k,j}(\vec{x},t) \quad (8)$$

If the suitable $\Delta I_{k,j}(\vec{x},t)$ is found, it is possible to correct any image $I_{k,j}(\vec{x},t)$ to make it similar to a reference one $I_{k,j_{ref}}(\vec{x},t_{ref})$.

We define a double linear correction mapping (LCM) for camera j that compensates the variations of $G\cdot Q$ and E in an acquired image $I_{k,j}$ as:

$$\Delta I_{k,j}(\vec{x},t) = a^{cam}_{k,j}(t)\cdot a^{il}_{k,j}(\vec{x},t) \quad (9)$$

Each image will be corrected by two terms: $a^{cam}_{k,j}$ is the *inter-camera correction mapping* (ICCM) that compensates the scenario and change between cameras, and $a^{il}_{k,j}$ is the *illumination correction mapping* (ICM) that corrects the change in location within the same camera and illuminant(s).

The variation of BG intensities ($\Delta I^b_{k,v}$) regarding a reference time t_{ref} is measured using the following expression:

$$\Delta I^b_{k,v}(\vec{x},t) = \frac{\bar{I}^b_{k,v}(\vec{x},t_{ref})}{\bar{I}^b_{k,v}(\vec{x},t)} \quad (10)$$

where $\bar{I}^b_{k,v}(\vec{x},t)$ is the average intensity value of BG region v in location $\vec{x}$, time t and channel k according to the selection explained in the Section 3.1.1. Note the similarities between Equation (10) and Equation (7).

Assuming camera exposure values are not modified to maintain a linear relation between light and image intensity, the variations of the BG regions model the terms $g_i(t)$ of Equation (5) with respect to t_{ref} that influence the FGs. For that purpose we estimate the ICM using:

$$a^{il}_{k,j}(\vec{x},t) = \left(1, \Delta I^b_{k,1,j}(\vec{x},t),\ldots,\Delta I^b_{k,V,j}(\vec{x},t)\right)\cdot\vec{m}_{x,j} \quad (11)$$

where $\vec{m}_{x,j}$ comes from Equation (16).

Regarding ICCM, better results are obtained when it is included an offset term capable of partially compensating for model derivations such as camera noise, intensity offsets or camera non-linearity. This term accounts for any error due to the simplifications we made regarding the image formation model (Section 2.2). Therefore, the original Equation (8) is replaced by the following:

$$\hat{I}_{k,j}(\vec{x},t) = \alpha^{cam}_{k,j}\cdot a^{il}_{k,j}(\vec{x},t)\cdot I_{k,j}(\vec{x},t) + \beta^{cam}_{k,j} \quad (12)$$

where $\alpha^{cam}_{k,j}$ and $\beta^{cam}_{k,j}$ are constants depending on camera j.

3 Method description

To estimate the relationship between the people response (FG) and the light variations in the rest of the scene (BG), we define a training mode (Section 3.1) and run-time mode (Section 3.2).

3.1 Training mode

In training mode, the relation between cameras as well as the relation between changes in BG and FG are learned. We establish these relations by using some FG objects seen in different cameras (for ICCM), different locations and illuminations (for the ICM). In Section 3.1.1 we discuss the selection of locations and the segmentation of BG regions, in Section 3.1.2 we explain how to estimate the ICCM ($a^{cam}_{k,j}$) and in Section 3.1.3 the vectors $\vec{m}_{x,j}$ for ICM which relates the FG to the BG.

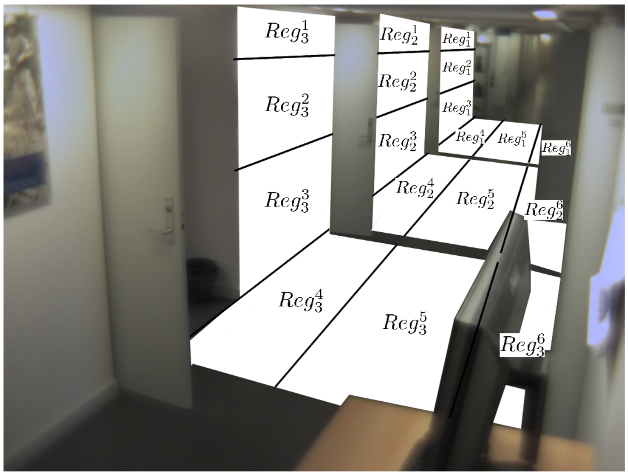

Figure 1: Indoor surveillance scenario modeling with selected background regions. Subscript refers to location and superscript to region index.

3.1.1 Locations and background region selection

We model the 3D world geometry in the acquired image for each camera by splitting the scene in a set of $n = \{1, \dots, N\}$ pre-defined locations. Thus, each pixel $\vec{x}$ is mapped to a location n. We define the background regions $v = \{1, \dots, V\}$ associated to each location according to the geometry of the scene. The background regions (*Reg* in the example of Figure 1) are used to estimate the change in illumination to correct the foreground objects. It is desirable to select background regions that are most influenced by light changes. For example, in Figure 1 some regions are located on the wall on the left because outdoor light goes through the opposite doors. Thus, the light reflected on this surface will change in accordance with the outdoor light. In our experiments, the background regions were manually specified.

3.1.2 Camera-relation training for ICCM

In the ICCM, we obtain a linear mapping from one camera to another by using several intensity values of foreground objects in the different locations to approximate it to the reference one. A comprehensive survey on foreground-background separation is presented by Bouwmans [3]. In our experiments, we use manually segmented foreground objects. For comparing two images from two cameras, we make use of the normalized cumulative object histograms H with h values. If there are κ training objects, we have $W = h \cdot \kappa$ samples for each location and each color channel, different from the reference one. Thus, the ICCM of Equation (12) is obtained by minimizing the error between the histogram values of the reference camera j_{ref} and the estimated camera j:

$$\{\alpha^{cam}_{k,j}, \beta^{cam}_{k,j}\} = \underset{\{\alpha^{cam}_{k,j}, \beta^{cam}_{k,j}\}}{\arg\min} \sum_{w}^{W} \| H^{f}_{w,k,j_{ref}} - \left(\alpha^{cam}_{k,j} \cdot H^{f}_{w,k,j} + \beta^{cam}_{k,j}\right) \|_2 \tag{13}$$

where $H^{f}_{w,k,j}$ is the sample w of the normalized cumulative histogram between 0 and 1 of the channel k of the FG in camera j. Due to overexposed pixels, shadows, erroneous background region assignments, etc., many of these values may be outliers. Because linear least-squares estimates do not work properly when the error distribution is not normal, they are not suitable tools in this case. Instead we use a robust linear regression algorithm which is able to manage this kind of outliers. It is based on a maximum likelihood estimation [15] which performs an iterative reweighted linear squares regression using a bisquare weight function.

3.1.3 FG/BG-relation training for ICM

The normalized cumulative histograms H are also employed in the estimate of the relationship between BG and FG. In this case, we extract data from the BG and FG when the light changes. Similar to the ICCM estimate, we have $W = 3 \cdot h \cdot \kappa$ samples of data for each location. The estimation is independent of the channel because the relationship models g_i are the same in every channel. Using this data, we construct the following ratios expressing the variation for foreground (14) and background (15) histograms relative to the reference time t_{ref}:

$$\Delta H^f_{w,j}(\vec{x},t) = \frac{H^f_{w,j}(\vec{x},t_{ref})}{H^f_{w,j}(\vec{x},t)} \quad (14)$$

$$\Delta H^b_{w,v,j}(\vec{x},t) = \frac{H^b_{w,v,j}(\vec{x},t_{ref})}{H^b_{w,v,j}(\vec{x},t)} \quad (15)$$

where $H_{w,j}$ is the sample w of the normalized cumulative histogram of camera j and v refers to the background region. We fit a linear combination of the BG variations to FG using $\vec{m}_{x,j}$. This is a vector that relates the change in V background regions to that in the foreground at location $\vec{x}$, constant through time and equal for each channel k in camera j. It is estimated by:

$$\hat{m}_{x,j} = \arg\min_{\tilde{m}_{x,j}} \sum_{w}^{W} \|\Delta H^f_{w,j} - (1, \Delta H^b_{w,1,j}, \dots, \Delta H^b_{w,V,j}) \cdot \vec{m}_{x,j}\|_2 \quad (16)$$

where W is the number of samples and V is the number of BG regions, such that $v = [1, \cdots, V]$. To avoid unreliable samples, we eliminate occluded and shadowed areas of the BG. In our experiments, we tested two optimization methods: robust linear regression algorithm and linear programming. We use the second method as it yields lower errors according to indicators explained in Section 4.2. Nevertheless, the selection of the correct fitting method may depend on the data. In our experiments we observed that, to obtain good fitting, we need enough samples ($W \gg V$).

3.2 Run-time mode

In the run-time mode, the LCM of Equation (12) is applied. It first requires the estimate of the BG intensities in the image and then the computation of the mappings ICM (a^{il}) and ICCM (α^{cam}, β^{cam}). Regarding the ICCM, Equation (12) is applied using the parameters estimated by Equation (13).

4 Experimental evaluation

4.1 Experimental setup

The proposed method was tested in a dedicated experiment with multiple cameras, multiple people, multiple locations and changing lighting conditions. Although there are several datasets on the internet related to people re-identification in a multi-camera network[1] and some others

[1]QMUL underGround Re-IDentification (GRID) dataset. http://www.eecs.qmul.ac.uk/~ccloy/downloads_qmul_underground_reid.html

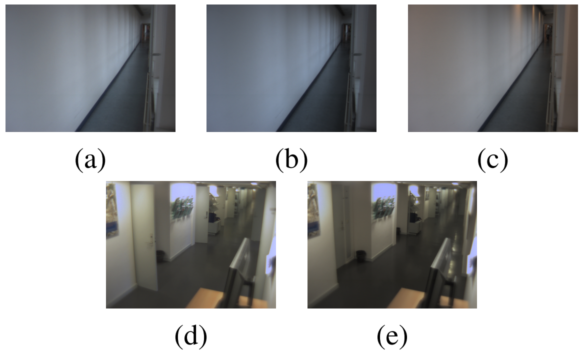

Figure 2: Examples of light changes in both experimental scenarios. Images show variations due to: natural outdoor light variations ((a) and (b)); switch the indoor lights on (c); and open (d) and close (e) multiple doors in the indoor scenario.

Figure 3: Example images of each person. The top-left is used for training.

also include light changes[2,3], we were not able to find any data set that fits all our requirements (overall, background information and camera's auto-settings off). Therefore, we used our own dataset[4]. For the experiment, we used two Firewire cameras influenced by indoor and outdoor illumination. The illumination was changed by switching the indoor lights on and off, by opening and closing doors, and by outdoor lighting variations due to changing cloud coverage (see Figure 2). The time between each capture is enough to have a noticeable change in outdoor lighting. For both cameras, the variation in the average value of images exceeds 10 %. In each camera view, we defined 3 locations (Figure 1), used 1 person for the training phase and 8 other persons for the test phase. All persons are captured walking in two directions. Each one of the training and test samples were recorded with different outdoor light conditions. In camera 2, the average background intensity during acquisition of the training samples ranges between 102 and 123, while the average background intensity during testing ranges between 97 and 112. The test persons each wore different shirts whose colors are well-distributed along the RGB space. An example picture of each subject is shown in Figure 3. People outlines are manually segmented to obtain FG information, used during training and evaluation. The shadows also are manually segmented and discarded. To demonstrate the performance of the method with less optimal segmentation, an experiment has been carried out by eroding and dilating the manually segmented outlines (Section 4.6). A total of 96 frames have been used in training and 192 frames for testing. The 96 training frames have 8 repetitions for each light-condition, location and camera (8 frames x 2 light conditions x 3 locations x 2 cameras). The 192 test frames have 2 repetitions for each person, light-condition, location and camera (8 persons x 2 directions x 2 light conditions x 3 locations x 2 cameras).

Camera's auto-exposure (aperture, exposure time and gain), auto-white balance (correction between bands) and gamma correction have been disabled.

[2]VIPeR: Viewpoint Invariant Pedestrian Recognition. `http://vision.soe.ucsc.edu/node/178`

[3]Person Re-ID (PRID) 2011 dataset. `http://lrs.icg.tugraz.at/datasets/prid/index.php`

[4]Multi-Camera Dynamic Light (MCDL) dataset is publicly available in: `https://www.researchgate.net/publication/264462014_MCDL_Dataset`

4.2 Similarity and error measure

We evaluate the intensity correction by measuring the dissimilarity between persons using their histograms. We choose the Earth Mover's Distance (EMD) [26], normalized between 0 and 1, a cross-bin algorithm which provides a minimum cost solution to transform one image distribution into another and is robust against noise and small intensity variations. We calculate the Euclidean EMD for each channel and the combined value as the average value of these 3 EMDs.

We calculated both the intra-class and inter-class distances. For intra-class, denoted with suffix ii, we compare between the same person in the different acquisitions: several locations, illumination conditions and cameras. For inter-class, denoted with suffix ij, any person is compared to the other people in the same and different photometric conditions.

We evaluate the performance of the system with an error measure that is based on normalized Cumulative Histograms Distribution (CHD) of the computed EMDs. For the intra-class CHD_{ii} closer to 0 are better while for the inter-class CHD_{ij} closer to 1 are better. Ideally our algorithm increases the distance between CHD_{ii} and CHD_{ij} as much as possible. Using the cumulative histograms, for any threshold value n_{th}: $FP(n_{th}) = CHD_{ij}(n_{th})$ and $FN(n_{th}) = 1 - CHD_{ii}(n_{th})$, we search for that threshold where the Minimal Error Criterium $MEC = FP + FN$ is minimal.

For this MEC, high error rates are expected due to pose variations, partial occlusions, inter-reflections, self-casting shadows, which all cause the persons to appear differently. The research in this paper only corrects for changes in color distributions caused by light variations.

The Receiver Operating Characteristic (ROC) curve is used to visualize the performance, where the true positive rate is $TP(r) = 1 - FN(r)$.

4.3 Implementation of the reference methods

The obtained results are compared to four cases: direct measurements without any correction, a state-of-the-art automatic exposure algorithm (AE) [18], a color correction based on the inverse-intensity chromaticity (IIC) [29] and a combination AE+IIC.

To allow a fair comparison and use the same input images for all methods, the AE algorithm estimates a correction coefficient based on the brightness of the image, without modifying the camera gain and exposure time. It is based on three parameters: the reference average value to be reached, the number of regions and the weight values of each region. The reference value is calculated by estimating the average value of the illumination reference images of the test set. Liu did not specify the method to set up the number of regions or the weight values. These choices depend on the scenario and how it responds to the light changes. We decide to split the scenario into 3 regions as it is shown in Figure 4. Regarding the weights, we choose 0.15, 0.7 & 0.15 from left to right for camera 1. We give a higher weight to the central region because people are placed there. In a similar way, the chosen weights for camera 2 are 0.2, 0.7 & 0.1. In this case the region on the left has a greater weight than the region on the right because the light coming from outside is reflected on the wall on the left. These weights have been chosen empirically after testing several variations, being the ones that perform the best.

Camera 1

Camera 2

Figure 4: Region setup used by the AE algorithm described in Section 4.3.

Similar to AE, IIC is applied to each channel using Tan’s implementation. With AE+IIC, IIC correction is applied over the AE-corrected image.

4.4 Intra camera results

(a)

(b) (c)

(d) (e)

Figure 5: Example of image correction. (a) Original image to be corrected. Corrected images for the intra-camera (b) and inter-camera (c) results. Reference images taken in the intra-camera (d) and inter-camera (e) experiment. Images (b) and (c) are more similar to images (d) and (e) in color distribution, respectively.

An example of a corrected image for the intra-camera correction is shown in Figure 5(a)(b)(d). Tables 3(a) and (b) show the numerical results of intra camera experiments. The results from camera 1 show an error reduction for our algorithm compared to the uncorrected case (MEC is 0.42 and 0.66, respectively), which implies an improvement of

37 %. The IIC algorithm do not work as well as LCM, and AE makes the uncorrected case even worse.

Table 3: Results of cameras correction. The value of the MEC besides the improvement compared to the images with no correction are shown. Each row is: uncorrected images, by using the auto-exposure algorihtm (AE), the IIC, auto-exposure plus IIC (AE+IIC) and our algorithm using three different segmentations for the foreground: i) manual –LCM–, ii) modifying the area size of the manual segmentations in -30 % and 60 % –LCM (-30,+60)–, and iii) area variations of -90 % and 200 % –LCM (-90,+200)–.

Camera 1 correction

Method	MEC	Improvement (%)
Uncorrected	0.66	–
AE [18]	0.69	-4
IIC [29]	0.64	4
AE+IIC [18, 29]	0.67	-1
LCM	0.42	37
LCM (-30,+60)	0.50	24
LCM (-90,+200)	0.55	18

(a)

Camera 2 correction

Method	MEC	Improvement (%)
Uncorrected	0.59	–
AE [18]	0.39	34
IIC [29]	0.70	-19
AE+IIC [18, 29]	0.57	4
LCM	0.40	32
LCM (-30,+60)	0.40	32
LCM (-90,+200)	0.38	36

(b)

Comparing images between both cameras

Method	MEC	Improvement (%)
Uncorrected	0.69	–
AE [18]	0.61	12
IIC [29]	0.71	-4
AE+IIC [18, 29]	0.66	5
LCM	0.50	27
LCM (-30,+60)	0.52	24
LCM (-90,+200)	0.69	-1

(c)

In the ROC curve (Figure 6), the improvement of our algorithm is very substantial while the IIC algorithm, which yields the best alternative results, hardly outperforms the original images. This is probably because our method rejects the overexposed regions of the BG, making our algorithm more robust against highlights. Furthermore, the BG-FG relationship is more reliable than the transformation values calculated by AE and IIC algorithms, as the BG-FG relationship has been trained for local rather than global illumination changes.

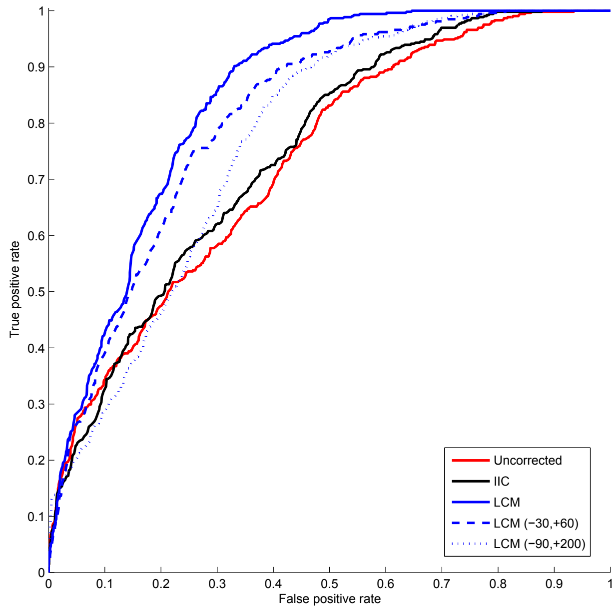

Figure 6: ROC curve for camera 1. The true positive rate against the false positive rate of the five analysis cases are shown. The red solid lines belong to the uncorrected case, the black solid lines to the best alternative performance (IIC [29]), the blue solid lines to our algorithm using a manual segmentation, the dashed ones to our algorithm with segmentation errors of -30 and +60 % of area variation, and the dotted ones to our algorithm with segmentation errors of -90 and 200 %.

The results in camera 2 confirm that LCM outperforms the others. Note that IIC does not improve the uncorrected case (in Table 3(b)) and – in comparison to (AE) alone – it also does not improve when both methods (AE+IIC) are combined. Although results of AE are positive and noteworthy (MEC is 0.39 and the improvement is 34 %), LCM error is similar (MEC is 0.40 and the improvement is 32 %).

In the case of camera 2, the most important source of light is the sun, and the indoor lights have less influence. Although indoor lights have been switched on/off, the in-

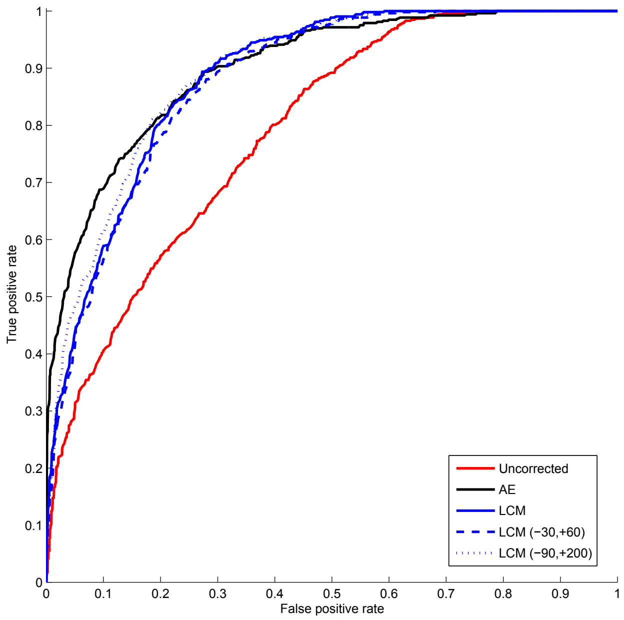

Figure 7: ROC curve for camera 2. The figure description is similar to Fig. 6. In this case, the best alternative is the AE [18].

Table 4: Estimated ICCM parameters using to correct images from camera 1. Intensity range is [0,1].

Channel	α	β
Red	1.1909	-0.0118
Green	1.2043	-0.0150
Blue	1.3264	-0.0195

fluence of the outdoor light variation produces a higher intensity and color variation. Thus, it seems that IIC is not able to estimate the correct main illuminant in every picture. In addition, the indoor lights do not over expose the image acquisition as happened in camera 1. It can be said that the multiple light sources and over-exposed BG areas in camera 1 make a more complex scenario compared to camera 2. This makes photometric variations in camera 2's FG objects simpler, so it seems reasonable that errors in uncorrected images are slightly smaller than in camera 1, as Table 3(b) shows.

To conclude, results on this dataset demonstrate that our algorithm is able to correct photometric variations in objects even in complex illumination scenarios on condition that BG areas are close enough to FG objects and the camera works on the linear region. In this context, complex scenarios may be those where there are a multitude of reflections on Lambertian surfaces, at least one illuminant is changed and the intensity of lights vary significantly.

4.5 Inter camera results

An example of a corrected image for the inter-camera correction is shown in Figure 5(a)(c)(e). The estimates of the ICCM are shown in Table 4.

Table 3(c) shows the results of inter-camera correction. Compared to intra-camera results the MEC are slightly worse than the corresponding mean of camera 1 and camera 2 MEC. Deterioration of LCM is probably due to different illumination and shadow geometries at both cameras. Still our algorithm significantly outperforms the others. As Table 3(c) shows, the MEC improvement of IIC correction is negative and using the AE+IIC algorithm is fairly close to the uncorrected case (a difference of 5 %) due to the poor performance of camera 1. The inter-camera results show that LCM obtains a large improvement of 27 %, which it is noteworthy and more than twice the AE algorithm's improvement. Figure 8 confirms the LCM improvement: although AE curve is over the uncorrected curve for almost every false positive rate value, the LCM curve is even closer to 1 and, hence, much better than AE's.

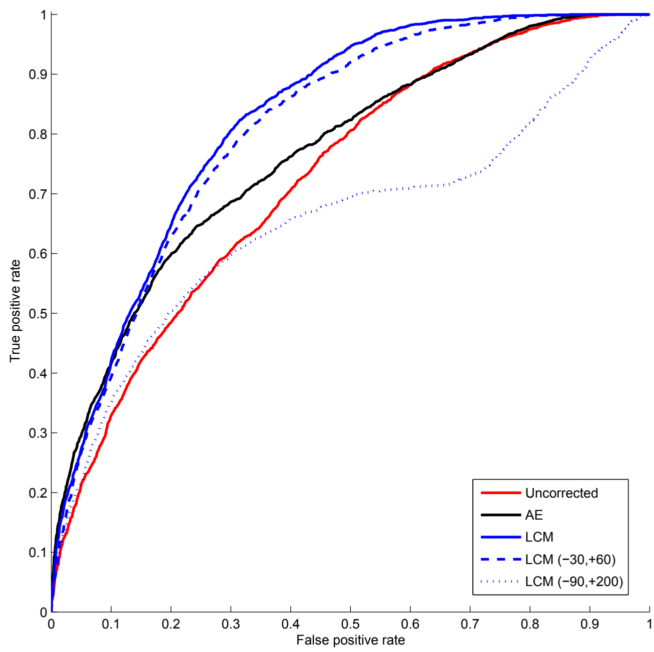

Figure 8: ROC curve for both cameras. The figure description is similar to Fig. 7. In this case, the best alternative is the AE [18].

These results show that in a non-overlapping multi-camera system – although scenarios' illumination is different (indoor and outdoor) – our algorithm is able to estimate a correction that helps to maintain the color appearance of objects in such scenarios.

4.6 Robustness against segmentation errors

To evaluate the robustness of our method against segmentation errors in the foreground extraction an additional experiment was carried out.

The size of each foreground object was extended or reduced. The values of -30/60 % and -90/200 % were chosen to reflect small and large errors in the FG segmentation, where a negative value refers to a reduction and a positive value refers to an extension. For each image, the decision about the extension or reduction was random.

Note that the scope of this paper is on color correction. For color correction, the foreground segmentation is only used during training mode whereas the correction during run-time is based on background intensities only. Foreground segmentation is during run-time only used for evaluation. To determine the sensitivity of the color correction method, only the segmentation during training mode is modified by erosions and dilations, and not the segmentation used in evaluation within the EMD matching.

The assessment was the same as the exposed above. The results are presented in Table 3 and Figures 6–8 for comparison with the rest of the methods. These demonstrate that our method maintains similar error rates when the size variation is reduced by 30 % and enlarged by 60 %. In this case it still outperforms the AE, IIC and AE+IIC methods. The deterioration of the algorithm becomes noticeable when the segmentation errors are always large (in the case of -90 % and 200 %).

5 Conclusions

In this paper, we describe a computationally efficient method to correct photometric variations between scenarios. For each image, these variations are estimated by simple vector operations and the correction is based on a double linear mapping. This is linked with the fact that, contrary to other color constancy algorithms, the illuminants do not have to be estimated. Therefore, the scenarios are easy to calibrate and the use of canonical, known illuminants or color-checker patterns are not required.

The main conclusions of this paper are the following: Our algorithm performs much better than uncorrected images and state-of-the-art auto-exposure together with a representative color correction algorithm. Furthermore, it is able to work properly in complex environments with several illuminants because, unlike most of inter-camera and other correction methods, local photometric changes are taken into account and the estimate of the BG-FG relation makes it easy to adapt changes in any object using BG information. It is also robust against foreground segmentation errors.

Our algorithm is able to correct changes in the settings of the camera as long as the change in the camera response is linear. Gain, aperture and exposure time variations mostly produce linear variations in the image if the post-processing is also linear. In future work we will investigate a combination of non-linear in-camera processing as present in many cameras with our algorithm. This shall include automatic characterization of such non-linear processing [11, 6, 14] and an adequate compensation.

Future work will also involve the development of a segmentation algorithm that automatically creates a segmentation of background regions that are classified by their response to the illumination changes.

REFERENCES

[1] Barnard, K., Martin, L., Funt, B.V.: Colour by correlation in a three-dimensional colour space. In: ECCV (1), pp. 375–389 (2000)

[2] Bouma, H., Borsboom, S., den Hollander, R., Landsmeer, S., Worring, M.: Re-identification of persons in multicamera surveillance under varying viewpoints and illumination. In: Proc. SPIE, vol. 8359 (2012)

[3] Bouwmans, T.: Recent advanced statistical background modeling for foreground detection-a systematic survey. Recent Patents on Computer Science **4**(3), 147–176 (2011)

[4] Chen, K.W., Lai, C.C., Lee, P.J., Chen, C.S., Hung, Y.P.: Adaptive learning for target tracking and true linking discovering across multiple non-overlapping cameras. IEEE Trans. Multimedia **13**(4), 625–638 (2011)

[5] D’Orazio, T., Mazzeo, P., Spagnolo, P.: Color brightness transfer function evaluation for non overlapping multi camera tracking. In: ACM/IEEE Int. Conf. Distributed Smart Cameras, vol. 2, pp. 1–6 (2009)

[6] Farid, H.: Blind inverse gamma correction. IEEE Trans. Image Processing **10**(10), 1428–1433 (2001)

[7] Finlayson, G.D., Drew, M.S., Funt, B.V.: Spectral sharpening: sensor transformations for improved color constancy. J. Opt. Soc. Am. A **11**(5), 1553–1563 (1994)

[8] Gijsenij, A., Gevers, T., van de Weijer, J.: Computational color constancy: survey and experiments. IEEE Trans. Image Processing **20**(9), 2475–2489 (2011)

[9] Gijsenij, A., Lu, R., Gevers, T.: Color constancy for multiple light sources. IEEE Trans. Image Processing **21**(2), 697–707 (2012)

[10] Gilbert, A., Bowden, R.: Incremental, scalable tracking of objects inter camera. Computer Vision and Image Understanding **111**(1), 43–58 (2008)

[11] Goldman, D.: Vignette and Exposure Calibration and Compensation. IEEE Trans. Pattern Analysis and Machine Intelligence **32**(12), 2276–2288 (2010)

[12] Grossberg, M., Nayar, S.: Modeling the space of camera response functions. IEEE Trans. Pattern Analysis and Machine Intelligence **26**(10), 1272–1282 (2004)

[13] Hsu, E., Mertens, T., Paris, S., Avidan, S., Durand, F.: Light mixture estimation for spatially varying white balance. In: ACM SIGGRAPH, pp. 70:1–7 (2008)

[14] Huang, S.C., Cheng, F.C., Chiu, Y.S.: Efficient contrast enhancement using adaptive gamma correction with weighting distribution. IEEE Trans. Image Processing **22**(3), 1032–1041 (2013)

[15] Huber, P.J.: Robust Statistics. Wiley Series in Probability and Statistics. Wiley-Interscience (1981)

[16] Javed, O., Shafique, K., Shah, M.: Appearance modeling for tracking in multiple non-overlapping cameras. In: IEEE Computer Vision and Pattern Recognition, vol. 2, pp. 26–33 (2005)

[17] Kim, S., Lin, H., Lu, Z., Sustrunk, S., Lin, S., Brown, M.: A new in-camera imaging model for color computer vision and its application. IEEE Trans. Pattern Analysis and Machine Intelligence **34**(12), 2289 –2302 (2012)

[18] Liu, J., Ren, D., Zou, J., Wu, Y., Li, S.: Study of automatic exposure algorithm based on HD IP camera. In: Int. Conf. Advanced Intelligence and Awareness Internet, pp. 265–268 (2010)

[19] Loy, C.C., Xiang, T., Gong, S.: Multi-camera activity correlation analysis. In: IEEE Computer Vision and Pattern Recognition, pp. 1988–1995. IEEE (2009)

[20] Madden, C., Cheng, E., Piccardi, M.: Tracking people across disjoint camera views by an illumination-tolerant appearance representation. Machine Vision and Applications **18**(3), 233–247 (2007)

[21] Parameswaran, V., Singh, M., Ramesh, V.: Illumination compensation based change detection using order consistency. In: IEEE Computer Vision and Pattern Recognition, pp. 1982 –1989 (2010)

[22] Park, S., Kim, G., Jeon, J.: The method of auto exposure control for low-end digital camera. In: Int. Conf. Adv. Communication Techn., vol. 03, pp. 1712–1714 (2009)

[23] Porikli, F.: Inter-camera color calibration by correlation model function. In: Int. Conf. Image Processing, vol. 2, pp. 133–6 (2003)

[24] Prosser, B., Gong, S., Xiang, T.: Multi-camera matching using bi-directional cumulative brightness transfer functions. In: British Machine Vision Conf. (2008)

[25] Riess, C., Eibenberger, E., Angelopoulou, E.: Illuminant color estimation for real-world mixed-illuminant scenes. In: IEEE Int. Conf. Computer Vision Workshops, pp. 782–789 (2011)

[26] Rubner, Y., Tomasi, C., Guibas, L.J.: The Earth Mover’s Distance as a metric for image retrieval. Int. J. Comput. Vision **40**(2), 99–121 (2000)

[27] Sayed, M.S., Delva, J.G.R.: An efficient intensity correction algorithm for high definition video surveillance applications. IEEE Trans. Circuits and Systems for Video Technology **21**(11), 1622–1630 (2011)

[28] Siebler, C., Bernardin, K., Stiefelhagen, R.: Adaptive color transformation for person re-identification in camera networks. In: ACM/IEEE Int. Conf. Distributed Smart Cameras, pp. 199–205 (2010)

[29] Tan, R.T., Nishino, K., Ikeuchi, K.: Illumination chromaticity estimation using inverse-intensity chromaticity space. In: IEEE Computer Vision and Pattern Recognition, pp. I:673–680. IEEE Comput. Soc (2003)

[30] Wang, Y.K., Wang, C.F.: Face detection with automatic white balance for digital still camera. In: Int. Conf. Intelligent Information Hiding and Multimedia Signal Processing, pp. 173–178 (2008)

[31] Wannous, H., Lucas, Y., Treuillet, S., Mansouri, A., Voisin, Y.: Improving color correction across camera and illumination changes by contextual sample selection. J. Electronic Imaging **21**(2), 023,015–1–023,015–14 (2012)

[32] Withagen, P.J., Groen, F.C.A., Schutte, K.: CCD color camera characterization for image measurements. IEEE Trans. Instrumentation and Measurement **56**(1), 199–203 (2007)

[33] Withagen, P.J., Schutte, K., Groen, F.C.A.: Global intensity correction in dynamic scenes. Int. J. Computer Vision **86**(1), 33–47 (2010)