Mapping Colour in Image Stitching Applications

David Hasler

Sabine Süsstrunk

david.hasler@bluewin.ch

sabine.susstrunk@epfl.ch

Laboratory of Audiovisual Communication (LCAV)

Swiss Federal Institute of Technology (EPFL)

Lausanne, Switzerland

Tel +41 21 693 6664

Fax +41 21 693 4312

June 4, 2003

Abstract

Digitally, panoramic pictures can be assembled from several individual, overlapping photographs. While the geometric alignment of these photographs has retained a lot of attention from the computer vision community, the mapping of colour, i.e. the correction of colour mismatches, has not been studied extensively. In this article, we analyze the colour rendering of today's digital photographic systems, and propose a method to correct for colour differences. The colour correction consists in retrieving linearized relative scene referred data from uncalibrated images by estimating the Opto-Electronic Conversion Function (OECF) and correcting for exposure, white-point, and vignetting variations between the individual pictures. Different OECF estimation methods are presented and evaluated in conjunction with motion estimation. The resulting panoramas, shown on examples using slides and digital photographs, yield much-improved visual quality compared to stitching using only motion estimation. Additionally, we show that colour correction can also improve the geometrical alignment.

Keywords – Digital panoramic photography, scene-referred image encoding, motion estimation, pose estimation, colour correction, Laguerre OECF, Polynomial OECF, white-balancing, mosaicing, vignetting.

1 Introduction

Panoramic pictures have always had a special place in photography. Being able to represent in a photograph the whole horizontal or vertical visual impression one has of a scene has another impact than the usual photographs that only depict a limited rectangular view. It is therefore not surprising that a number of specialized cameras and camera systems have been developed that allow one to capture visual angles from 120 up to 360 degrees. Using a moving slit shutter and simultaneously moving the film and/or the camera can achieve this effect. Similarly, most APS (Advanced Photo System) film cameras have a panorama setting that covers part of the film to allow taking pictures with a $1 : 3$ aspect ratio instead of the normal $2 : 3$. The disadvantages of such systems are that the specialized cameras are very expensive, and the resulting panorama from the APS system, utilizing only part of the film for exposure, results in limited resolution reproductions.

Digitally, panoramic pictures can be created by "stitching" together several individual, overlapping photographs. The advantages of being able to create a panorama from several images is that such a panorama has sufficient resolution that it can be represented on a larger output, and that a panorama can be assembled even if the minimum focal length of the camera is not short enough to capture the panoramic scene in one exposure. Also, no specialized hardware is needed, allowing the creation of panoramas to be a feature of digital capture systems with no extra hardware cost.

The geometric aspect of assembly has been thoroughly investigated by the computer vision community, but in general, little attention has been paid to colour. For successful panoramic assembly, i.e. the visual impression that the panorama was taken in one single photograph, it is necessary that individual pictures are colour matched. The boundaries between the pictures should not be visible after assembly. In other words, an object of the scene should be represented by the same colour, regardless in which picture it appears.

There would be no colour mismatch between individual pictures if the camera applies a colorimetric reproduction model, i.e. a model that attempts to reproduce absolute scene colorimetry [1]. However, current digital cameras normally apply a photographic reproduction model [1, 2]. Each image is analyzed, exposed and rendered according to its scene content, with the goal of achieving the most pleasing reproduction to a human observer. This accounts for the colour mismatch of individual pictures: the scene content varies between neighbouring pictures, and objects are rendered according to their surround in the individual scene captured by the camera, and not the overall scene recreated by the panorama.

This paper presents a set of methods that can be used to correct the colour mismatches between images that were not captured under controlled conditions, i.e. with varying white-point and exposure, and unknown colour rendering algorithms. We assume that our camera performs for each individual picture an exposure and white balancing analysis, followed by a tone mapping, i.e. Opto-Electronic Conversion Function (OECF), that does not change from one picture to the next. Depending on how the observer adapted scene illuminant is estimated, the camera adopted white-point (i.e. the white balancing parameter values) can vary across the pictures of the panorama, generating colour mismatches. In addition, differences in exposure due to different maximum and minimum scene luminance values will make objects appear either darker or lighter in the individual pictures. We will assume that any mismatches in lightness of the pictures are due to different exposure, and mismatches in the colour are due to white-balancing[1]. The OECF is estimated [3].

Three methods are presented to correct for colour mismatches and to estimate the OECF. The first uses scanned slides as input, and the OECF is estimated through a calibration procedure. The other two methods use digital camera images as input. One proposes a polynomial OECF model that leads to excellent results but tends to be numerically unstable, and the other proposes a stable but more approximate OECF estimation method.

2 From the scene to the pixels

Photography is about light. The light is generated by a light source with a given spectral power distribution, hits a particular object of the scene, whose colour can be described by its spectral reflectance factors, is reflected by the object and reaches the sensor of the camera after passing through the lens. The light is then recorded by the camera sensor, which measures three components for each pixel, red (R), green (G) and blue (B)[2]. The magnitude of each response also depends on the spectral sensitivity of the sensor and filter. Algebraically, image formation can be expressed as:

$$\mathbf{P} = \mathbf{R} \cdot \mathrm{diag}(\mathbf{E}) \cdot \mathbf{L}, \tag{1}$$

---

[1]One can avoid these effects by controlling exposure and white balancing during capture, but this may lead to under- or overexposed images due to limited dynamic range of the sensor, and colour-casted individual images.

[2]Here, we simplify the actual image formation process of most digital cameras, which use a colour filter array (CFA) and actually only capture one colour per pixel. The other colour components are interpolated during de-mosaicing.
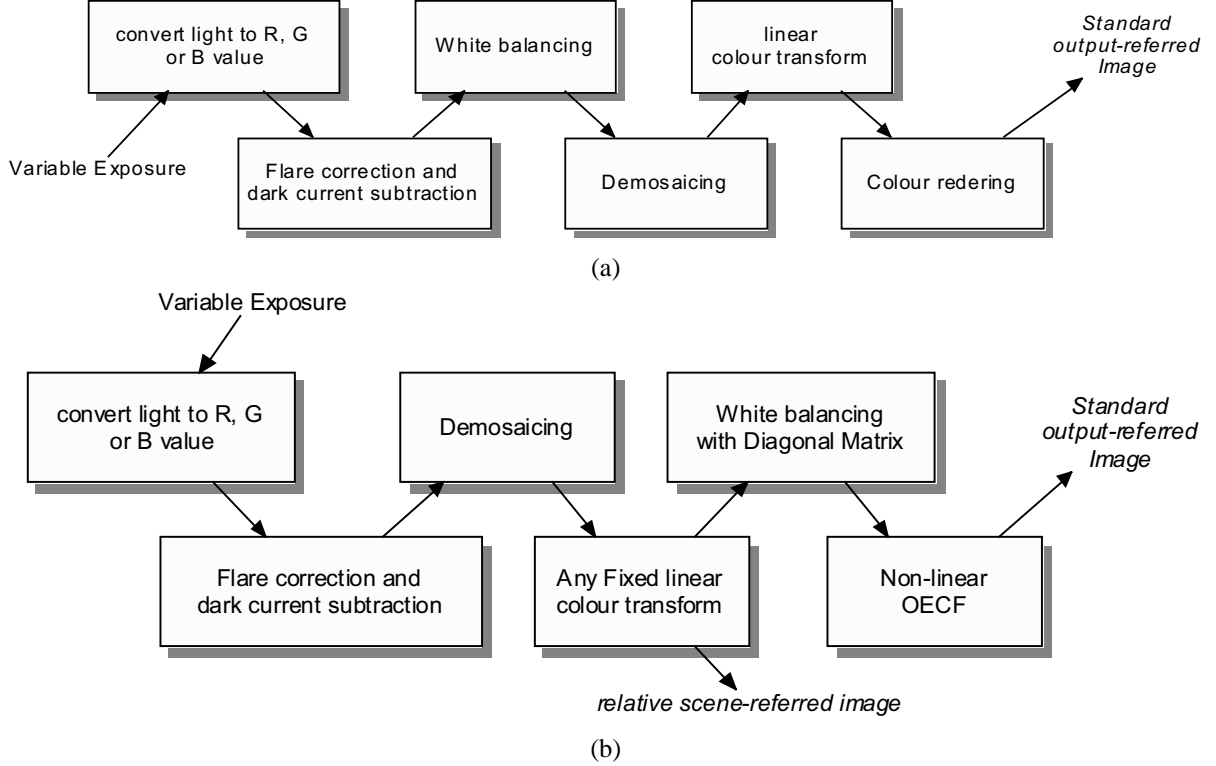
Figure 1: Image Acquisition workflow. (a) Example of real workflow (b) assumed workflow in our model.

where $\mathbf{P}$ is a $n \times 3$ matrix of camera response values R, G, B, $\mathbf{R}$ is a $n \times m$ matrix of all n sample (pixel) reflectance where m refers to the sampling of the reflectance over the visible spectrum, $\mathrm{diag}(\mathbf{E})$ is a $m \times m$ diagonal vector of the illuminant spectral power distribution and $\mathbf{L}$ is a $m \times 3$ matrix of the three device sensitivity vectors.

A silicon sensor has an approximately linear behaviour, therefore the value of each pixel is proportional to the illuminance at the sensor location. To be proportional to scene luminance, the camera has to correct for flare introduced by the optical system. Then, an image dependent white-balancing is performed. At this stage, the image encoding can be called *raw device RGB*. The image can then be linearly transformed (applying a 3×3 matrix) into a standard output-referred colour space encoding, such as sRGB [4]. Finally, colour rendering is applied to achieve a preferred reproduction, taking into account the characteristics of the original scene and the characteristics (gamut, viewing conditions) of the output referred encoding (see Figure 1(a)).

We will assume for the rest of this paper that the final pixel value only depends on the value read by the sensor, and not on the surrounding pixels. In other words, colour rendering is limited to applying a (non-linear) tone mapping, which can vary in each channel. We call this tone mapping the *Opto-Electronic Conversion Function* (OECF), although *gamma correction* is also an often-used term. Furthermore, we assume that the OECF does not change between the different pictures of the panorama. These two assumptions may not hold for more recent digital cameras.

2.1 White balancing

The purpose of white-balancing is twofold. First, to balance the channel responses of the camera due to differences in quantum efficiencies of the colour filters and sensor combination; and second, to account for the ability of the human visual system to discount the colour of the illuminant and to approximately preserve the appearance of an object. This visual phenomenon is called *Chromatic Adaptation*. For example, a white piece of paper appears to be white under sunlight as well as under incandescent light, even though the spectral power distributions of the two illuminants are quite different. Considering image formation as described in Eq. (1), it is clear that the camera will not record the same RGB values in the two situations. To correctly display the white piece of paper, the camera has to perform an image dependent correction, called *White-Balancing*. If white-balancing is done properly, the colour of the white piece of paper (the RGB values) will be the same whether the image was taken under sunlight or under incandescent light.

The most common chromatic adaptation model is the von Kries model [5][6, p. 125]. Given the $\mathcal{R}'_w \mathcal{G}'_w \mathcal{B}'_w$ values of the scene illuminant (or the camera adopted white point), and the $\mathcal{R}''_w \mathcal{G}''_w \mathcal{B}''_w$ of the observer adapted white point (or the white point defined by the output-referred encoding), expressed in a given colour space, a gain factor dependent on the ratios of the illuminants is applied to the three linear colour channels independently:

$$[R, G, B]^T \mapsto \mathbf{MDM}^{-1} \cdot [R, G, B]^T$$
$$\mathbf{D} \triangleq \mathrm{diag}\left(\frac{\mathcal{R}''_w}{\mathcal{R}'_w}, \frac{\mathcal{G}''_w}{\mathcal{G}'_w}, \frac{\mathcal{B}''_w}{\mathcal{B}'_w} \right).$$

Several chromatic adaptation transforms are used in practice and differ in the choice of colour space where the gain factors are applied [7]. In other words, the transforms differ in the choice of $\mathbf{M}$. In current digital cameras, we found that white balancing is often performed prior to demosaicing, i.e. directly in device RGB space ($\mathbf{M} = \mathbf{I}$).

Since we are using uncalibrated images, i.e. we do not know the linear transformation the camera applies to transform from device RGB to output-referred RGB, we will assume that the white balancing has been performed in output RGB using a diagonal matrix ($\mathbf{M} = \mathbf{I}$). Depending on the actual output-referred encoding, this assumption is reasonable [7]. For the rest of the paper, we will therefore assume that the camera imaging workflow follows the scheme of Figure 1(b), although a real workflow is much more likely to follow the one in Figure 1(a) [2]. In our models, we will estimate the white-point, OECF, and exposure parameters, and "reverse engineer" those camera processing steps. We call the resulting image encoding *relative scene-referred*. It is linear with respect to sensor illuminance (or scene luminance if flare was corrected), and within a (unknown) linear transform of sensor (or scene) colorimetric values.

2.2 Mapping the pixels

To summarize, the camera model used to obtain pixel values from the values read by the sensor is depicted in figure 1(b) and can be formalized as follows: for each colour channel, let E be the illuminance read by the sensor. The camera chooses a white point $[E_{wR}, E_{wG}, E_{wB}]^T$ (the *adopted white point*) and scales the output to this value. Finally, it applies a curve $S^{-1}(\cdot)$ to the data:

$$\begin{bmatrix} E_R \\ E_G \\ E_B \end{bmatrix} \mapsto \begin{bmatrix} S^{-1}\left(\frac{E_R}{E_{wR}}\right) \\ S^{-1}\left(\frac{E_G}{E_{wG}}\right) \\ S^{-1}\left(\frac{E_B}{E_{wB}}\right) \end{bmatrix}. \tag{2}$$

$S^{-1}(\cdot)$ is the Opto-Electronic Conversion Function (OECF). The models of the OECFs are given in the next section (3). We can ignore the colour transformation applied prior to white balancing in the model, because applying a different colour transformation is equivalent of having a camera with different spectral sensitivity curves.

3 Opto-Electronic Conversion Function models

Nowadays, an OECF is considered as highly confidential data by the camera manufacturers and is not available to the common user. Since the OECF is needed to equalise the colours in a panorama, it has to be estimated.

The term *Opto-Electronic Conversion Function* is reserved for an analog-to-digital device, such as a digital camera or a scanner. If an analog camera is involved in the picture taking process, we will use the more general term *tone mapping transfer function*. This function also relates the camera sensor illuminance to the final pixel value, but may combine more than one non-linear transform, such as the *D-logH* curve of photographic film (associated to a development process) and the scanner OECF.

The following sub-sections describe three models of a tone mapping transfer function. These models aim at the same goal, i.e. how to retrieve the sensor illuminance, or more specifically the relative scene-referred data from the pixel value, but are to be used in different contexts. The first system uses photographic slides, whose tone mapping transfer function is known. The unknowns are the exposure parameters and the scanner OECF, which can be measured using a simple procedure. The second system presents a very simple model for the OECF of a digital camera that can be used in a simultaneous motion, exposure and OECF estimation. The last model is also aimed at digital cameras, and offers more flexibility in the curve shape. Nevertheless, it can only estimate the OECF in a recursion of motion estimation and OECF estimation. In general, it delivers more precise OECFs, at the price of a worse robustness.

Conventions

We will assume that the camera outputs values range from 0 to 1. The Opto-Electronic Conversion Function inverse is designated by the letter $S(\cdot)$ and also outputs values in the range $[0, 1]$.

To simplify the notations, the computation of the OECF is described using one component per pixel (as opposed to three colour components per pixel). Nevertheless, three different OECF are computed to describe the camera, i.e. one OECF per colour channel, unless stated differently.

When referring to a colour component, we will use variable RGB. For example, E_{RGB} refers to either E_R or E_G or E_B (but not to vector $\mathbf{E}$). The summation over colour components is denoted as $\sum_{RGB}$. For example $\sum_{RGB}(E_{RGB}) = E_R + E_G + E_B$.

3.1 The transfer function of a slide and scanner system

As film photography is still the cheapest way to get high quality pictures, our first method uses scanned diapositives as input pictures. This section explains how to compute the tone mapping transfer function inverse of a slide and scanner system using one component per pixel. The extension to 3 components per pixel can be obtained by applying the results of this section individually to the Red, Green and Blue channel of the image.

The film characteristics are given by a function called the *D-logH* curve that relates film density to log Exposure, as illustrated in Figure 2. The scanner illuminates the diapositive and integrates the ratio of light that goes through the diapositive at a given photosite. The scanner is characterised by its OECF, which can be measured using a calibration slide containing patches of known density [8]. This calibration delivers a relationship between the pixel values I and the film density D.

If no calibration slide is available, a fair approximation of the scanner OECF is given by the following formula: Let I be the pixel value normalised such that $I \subset [0, 1]$. We assume that I is related to the lightness L^* of the pixel through the relation $I = \frac{1}{100}(L^*)^{\frac{1}{\gamma}}$, where gamma γ is a parameter that can be set on the scanner. Setting $\gamma = 1$, gives

$$I = \frac{1}{100} \cdot L^*$$

L^* is related to the relative luminance $\frac{Y}{Y_n}$ of the slide illuminated by the light source of the scanner through [6, p. 63]

$$L^* = 116 \cdot \sqrt[3]{\frac{Y}{Y_n}} - 16, \quad \frac{Y}{Y_n} > 0.008856$$

$$L^* = 903.3 \cdot \left(\frac{Y}{Y_n}\right), \quad \text{otherwise.} \tag{3}$$

In $\frac{Y}{Y_n}$, the luminance Y is expressed relative to the luminance Y_n of the most transparent area achievable with the given slide material. $\frac{Y}{Y_n}$ is related to film density D by

$$D = \log_{10}\left(\frac{Y}{Y_n}\right),$$

and finally, the illuminance is given by the *D-logH* curve of the film, which enables to build the slide and scanner *tone mapping transfer function inverse S*

$$R = S(I).$$

R is the *relative scene-referred data*. The function $S(I)$ relates the relative scene-referred data to the output-referred data I.

3.2 The Laguerre OECF model

This OECF model has been chosen because of numerical and analytical considerations and because its shape resembles the one of many cameras. The function is given by

$$S_a(x) = x + \frac{2}{\pi} \arctan\left(\frac{a \sin(\pi x)}{1 - a \cos(\pi x)}\right), \tag{4}$$

where a is the unknown parameter; $a \subset \,]-1; 1[$. The function is monotonic and its inverse is found by changing the sign of a. It has also an almost linear behaviour with respect to a, i.e.

$$S_{a+\Delta a}(x) \simeq S_a(x) + \frac{\partial S_a(x)}{\partial a} \cdot \Delta a. \tag{5}$$
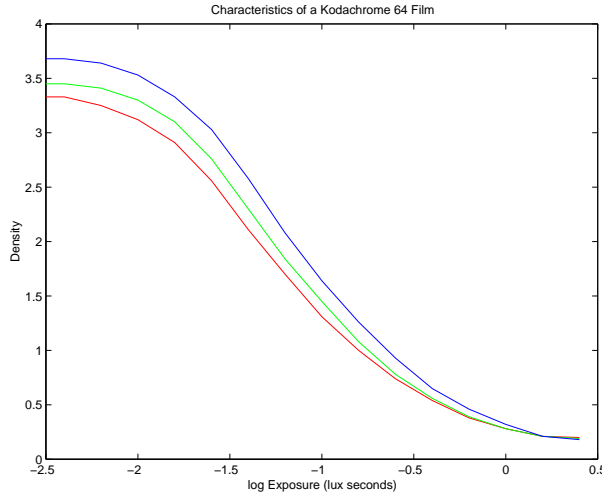
Figure 2: *D-logH* curve of the Kodachrome 64 slide film, relating film density to exposure, i.e. illuminance integrated over time. This curve can be obtained from the manufacturer Web site.

See figure 3 for an illustration of this argument.

3.3 The polynomial OECF model

In order to give a broad flexibility to the shape of the OECF, we adapt a model proposed by Mitsunaga and Nayar [9]. We will assume that the curve applied by the camera has the following shape

$$S(I) = a_1 I + a_2 I^2 + ... + a_N I^N \tag{6}$$

and the goal is to find the parameters a_i, under the constraint

$$\sum_{d=1}^{N} a_d = 1,$$

that ensures that $S(1) = 1$, and restricts the output in the $[0, 1]$ range if the function is monotonic.

4 Vignetting

Vignetting is responsible for a image being darker in the corners, and thus can cause colour mismatches in neighbouring pictures. On a high quality camera, this phenomenon is mainly due to a geometric effect: the $\cos^4$ effect [10, p. 121] [11, sec. 5.4]. Indeed, if the camera pictures a scene of constant luminance, the illuminance falling on the sensor varies according to the cosine to the fourth power of the viewing angle. Thus,

$$E_m = E_o \cdot \cos^4 \alpha$$
$$\tan(\alpha) = \frac{r_{\mathbf{p}}}{f}, \tag{7}$$

where E_m is the illuminance measured by the sensor, $r_{\mathbf{p}}$ the distance of the pixel from the optical centre of the image, f is the focal length of the camera and E_o is the illuminance that would have been measured if the light would have hit the sensor at the optical centre. By computing E_o for each pixel, we get an image without vignetting. Equivalently, vignetting can be corrected by introducing a factor $V(p)$:

$$E_o(\mathbf{p}) = E_m(\mathbf{p}) \cdot V(\mathbf{p})$$
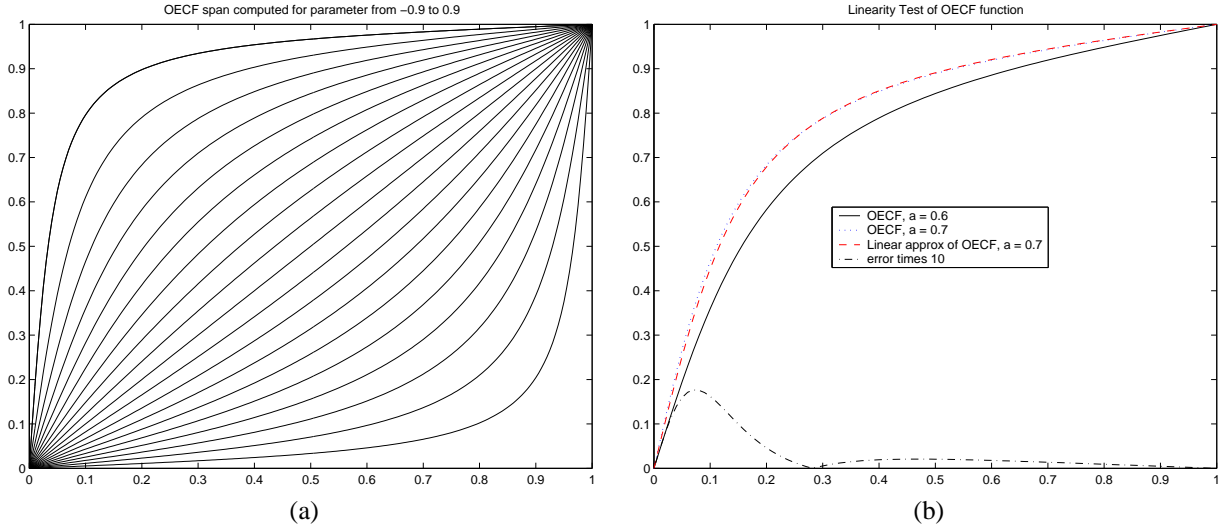$$V(\mathbf{p}) = \left[\frac{f^2 + r_{\mathbf{p}}^2}{f^2} \right]^2 \tag{8}$$

Figure 3: Plot of the Laguerre OECF. (a) Illustrates the span of the OECF defined in (4) for $a = \{-0.9, -0.8, ..., 0.9\}$. (b) Tests the linearity of the function with respect to parameter a ($a = 0.6$). The upper-dotted curves are the linear approximation of the curve superimposed to the exact curve ($a = 0.7$). The bottom curve is the error between the approximation and the exact curve, amplified by a factor 10.

$r_{\mathbf{p}}$ is given by the position $\mathbf{p}$ in the image ($r_{\mathbf{p}} = \sqrt{p_x^2 + p_y^2}$) and f is estimated by the stitching algorithm described in the next section.

Note that vignetting can only be corrected in this way if the pixels values are linearly related to sensor illuminance. Other phenomena related to vignetting, like its dependence on the aperture, have been ignored here.

5 Estimation methods

5.1 State of the art

There exists little work on colour registration in image mosaics, probably because mosaicing techniques have been developed by the computer vision community who traditionally worked with pictures that are linear with respect to illuminance. Colour has been used to improve the registration algorithm, for example in [12, 13], but without trying to modify it. Colour is also used to identify regions of interest in image sequences to perform replacements, for example in weather news broadcast (chroma-keying), as explained in [14]. We should also mention the work of Majumder et al. [15] who addressed the opposite problem of achieving colour uniformity across multi-projector displays. Closer to our problem is the work of Rushmeier and Bernardini [16] who equalised the colours in a 3D model of a statue, in a controlled environment, by compensating for illumination changes and white point changes using light spectra techniques. The same kind of approach is used by Yu et al. [17] to construct a 3D model of the scene and estimate the specular and diffuse reflection properties of each object of the scene, using known light sources. Nevertheless, the most relevant papers for the colour correction problem are the one that compute *radiance maps* from several differently exposed photographs, mostly to acquire high dynamic range pictures. Mann and Picard [18] (followed by [19]) used an exponential model to enhance the dynamic range of an image. Debevec and Malik [20] proposed a method to estimate an OECF of arbitrary shape from pictures with known exposures. The only constraint imposed is smoothness. Robertson et al. [21] modified slightly Devebec's method to give more weight to more reliable pixels. Finally, Mitsunaga and Nayar [9] proposed a polynomial model—similar to the one of Section 3.3—to describe the OECF that can be deduced from a set of differently exposed pictures with only an approximate knowledge of the exposure time.

5.2 General overview

We present three different methods to build the panorama.

The preferred estimation method

The most efficient way to achieve colour correction consists in estimating the OECF off-line, using a well controlled image pair or a calibration procedure. The OECF is then used in a simultaneous motion, exposure and white balancing estimation. This method is the most robust and delivers the best results. However, it requires a well controlled image pair, meaning two pictures of a static scene taken with a static camera, but with different (and known) exposure parameters, which is not always available.

The global motion and OECF estimation

The next logical method consists in estimating everything using the method presented in section 5.3. For numerical reasons, this is only possible with the Laguerre OECF model.

The iterative trial-and-error method

This method consists in first performing the motion and white balancing estimation. Then, given the motion and white balancing estimates, we can estimate a new OECF. We keep the OECF estimate if it is a good estimate (i.e. if it delivers a monotonic function), and re-iterate. This method can be applied with the polynomial OECF model, but it does not converge if the exposure difference is not known a priori[3]. In addition, this method requires special care to avoid that a bad alignment of the pictures makes the OECF estimation fail (see Appendix B).

5.3 Simultaneous motion and exposure parameter estimation

The motion estimation problem is formulated to take into account the changes in the settings of the camera: Let $I_0(\mathbf{p}), I_1(\mathbf{p}')$ be an image pair. By assuming that both images can be perfectly super-imposed by a warping process, we can write

$$S_\varphi\left[I_{0,RGB}(\mathbf{p})\right] = \tau_{RGB} \cdot S_\varphi\left[I_{1,RGB}(\mathbf{p}')\right]. \tag{9}$$

where $\mathbf{p}$ and $\mathbf{p}'$ denote the matching positions in the images[4], S_φ is the OECF inverse, φ is the parameter set that controls the shape of S and τ accounts for the exposure and the white point mismatch. To account for vignetting effects, Eq. (9) has to be corrected to

$$V(\mathbf{p}) \cdot S_\varphi\left[I_{0,RGB}(\mathbf{p})\right] = \tau_{RGB} \cdot S_\varphi\left[I_{1,RGB}(\mathbf{p}')\right] \cdot V(\mathbf{p}'),$$

where $V(\mathbf{p})$ is defined in (8). Each equation of this section can be written for the red, green or blue component of the image. To correct for exposure and white point mismatch, we have to use one τ parameter per colour component:

$$\tau = [\tau_R, \tau_G, \tau_B].$$

Now, let us introduce the function

$$\mathbf{U}_\eta : \mathbf{p} \longmapsto \mathbf{p}'$$

to describe the rigid motion model that tells the system where an object of image I_0 appears in image I_1. The motion model is controlled by the set of parameters η. Consequently, Eq. (9) becomes

$$V(\mathbf{p}) \cdot S_\varphi\left[I_{0,RGB}(\mathbf{p})\right] = \tau_{RGB} \cdot S_\varphi\left[I_{1,RGB}(\mathbf{U}\eta\left[\mathbf{p}\right])\right] \cdot V(\mathbf{U}\eta\left[\mathbf{p}\right]),$$

for each colour component. All the parameters involved in the problem can be put side-by-side in one vector θ:

$$\theta = [f, \tau, \varphi, \eta]^T.$$

The vector θ contains the focal length and the parameters of the geometry, of the exposure and of the white balancing mismatch. The focal length may already be contained in the parameters of the geometry (depending on the choice of the motion model). We put it apart because it is the only parameter that influences both the alignment, and the colorimetry (see section 4 on vignetting).

When using the Laguerre OECF model, θ also contains the OECF shape. We can find an updating rule for θ with a standard descent algorithms [22] by minimizing an objective function h:

$$h(\theta) = \frac{1}{n} \sum_{i=1}^{n} \|\mathbf{r}_i(\theta)\|^2 \tag{10}$$

$$r_{i,RGB}(\theta) \triangleq S_\varphi^{-1}\left\{V(\mathbf{p}_i) \cdot S_\varphi\left[I_{0,RGB}(\mathbf{p}_i)\right] - \tau_{RGB} \cdot S_\varphi\left[I_{1,RGB}(\mathbf{U}\eta\left[\mathbf{p}_i\right])\right] \cdot V(\mathbf{U}\eta\left[\mathbf{p}_i\right])\right\} \tag{11}$$

---

[3]The exposure parameters of a pictures are often stored in the picture itself.
[4]An object of the scene that appears in image I_0 at location $\mathbf{p}$ appears at location $\mathbf{p}'$ in image I_1.

where $\mathbf{p}_i$ are the pixel locations on the overlapping parts of the images, n the number of pixels, and $r_{i,RGB}$ the error at each pixel location for one colour component, also called *residual*. From a geometrical point of view, $I_1(\mathbf{U}\eta[\mathbf{p}_i])$ is the image that has been transformed such as to resemble the other one (I_0). For implementation reasons, it is convenient to approximate the residual with

$$r_{i,RGB}(\theta) \simeq \frac{V(\mathbf{p}_i) \cdot S_\varphi[I_{0,RGB}(\mathbf{p}_i)] - \tau_{RGB} \cdot S_\varphi[I_{1,RGB}(\mathbf{U}\eta[\mathbf{p}_i])] \cdot V(\mathbf{U}\eta[\mathbf{p}_i])}{\partial S(x)/\partial x|_{x=I_{0,RGB}(\mathbf{p}_i)}}.$$

Because some pixels might get saturated, we add a factor $W_{i,RGB}$ that allows to lower the influence of potentially saturated pixels[5]:

$$h(\theta) = \frac{1}{n} \sum_{RGB} \sum_{i=1}^{n} W_{i,RGB} \cdot r_{i,RGB}^2(\theta), \tag{12}$$

$$W_{i,RGB} \triangleq W[I_{0,RGB}(\mathbf{p}_i), I_{1,RGB}(\mathbf{U}\eta[\mathbf{p}_i])].$$

The choice of $W_{i,RGB}$ is discussed in Appendix A; note that it depends only on the pixel values and not on the scene-referred values. The minimum of function $h(\theta)$ is found using, for example, the Gauss-Newton algorithm. These algorithms are based on a first- or second-order approximation of function $h(\theta)$, therefore, the functions involved in the process have to be close to linear for a successful outcome of the optimisation. This is why the property of the Laguerre OECF described in Eq. (5) is important, and this is also why the polynomial OECF cannot be estimated using this method.

When reconstructing a mosaic composed of more than two pictures, the common practice consists in combining the two first images using the techniques presented so far, and registering the remaining pictures to the former image combination iteratively. This implies that from the third picture and on, the *target* picture I_0 is already colour corrected.

5.4 Estimating a polynomial OECF

The polynomial model of Eq. (6) cannot be estimated using the Gauss-Newton approach of Section 5.3, because of convergence. We adapt the model proposed by Mitsunaga and Nayar [9]. The idea is to make iterations between motion estimation and OECF estimation. In the easiest case, we can consider that the images are already aligned. For each colour component, let $S_j(\mathbf{p})$ be the illuminance that falls on the camera sensor at position $\mathbf{p}$ while taking picture j. From Eq. (6), we have $S_j(\mathbf{p}) = \sum_{d=1}^{N} a_d I_j(\mathbf{p})^d$. Now, we can express the ratio τ_{RGB} of the exposure settings between image I_0 and image I_1 with the following relation:

$$\frac{S_0(\mathbf{p})}{S_1(\mathbf{p}')} = \tau_{RGB}, \tag{13}$$

where $\mathbf{p}$ and $\mathbf{p}'$ are the matching positions in the two images. Now, if the camera moves between the two pictures (i.e. $\mathbf{p} \neq \mathbf{p}'$), Eq. (13) has to take the vignetting phenomenon into account, and becomes

$$\frac{V(\mathbf{p}) \cdot S_0(\mathbf{p})}{V(\mathbf{p}') \cdot S_1(\mathbf{p}')} = \tau_{RGB},$$

where $V(\mathbf{p})$ is the vignetting correction factor defined in (8).

The OECF (parameters $\{a_1, ..., a_N\}$) is found by computing the mis-registration h between the images in the overlap area

$$h(a_1, ..., a_N) = \frac{1}{n} \sum_{i=1}^{n} W_i \cdot \|r_i[a_1, ..., a_N]\|^2,$$

$$r_i(a_1, ..., a_N) \triangleq \{V(\mathbf{p}_i) \cdot S_0(\mathbf{p}_i) - \tau_{RGB} \cdot S_1(\mathbf{p}'_i) \cdot V(\mathbf{p}'_i)\}, \tag{14}$$

and by setting its derivative to zero:

$$\frac{\partial h([a_1, ..., a_N]^T)}{\partial[a_1, ..., a_N]^T} = \mathbf{0}. \tag{15}$$

---

[5]In this paper, we do not consider the resistance of the system to outliers.

The factor W_i allows to put more weight on reliable pixels, and is computed according to Appendix B. There are n pixels in the overlap area $\{\mathbf{p}_i, \mathbf{p}'_i\}$. The system of equation in (15) is a linear system that is solved under the constraint that function $S(\cdot)$ outputs values in the range $[0, 1]$, that is

$$\sum_{d=1}^{N} a_d = 1,$$

if the function is monotonic[6]. Once the OECF is found, the factors τ_{RGB} are found using a motion estimation iteration described in Section 5.3.

Note that the error function (14) differs from the one in (11). The error function (14) has the advantage of being less complex, and leads to a simple OECF estimation method, but has the drawback of having a degenerate solution:

$$S(\mathbf{p}) = \text{const}, \quad \tau_{RGB} = 1.$$

By imposing that S ranges from 0 to 1, one may think to avoid this degenerateness, but in fact, the solution $S(x) \simeq \text{const}$, and $\tau_{RGB} = 1$ in the regions of the histogram that contain most of the data still delivers a very small error. The red curve obtained in Figure 12(d) illustrates this argument. This is why the method should be used only if the exposure ratio τ is known, at least when working with only two images at a time.

In Figure 12, the nominal exposure difference between the two picture is known, and is equal to τ_G. τ_R and τ_B are unknown because they can be altered by the white balancing process. In the first iteration, only the green OECF can be computed, using the green channel. Then, by assuming the OECF is equal for the three colour channels, τ_R and τ_B can be derived, followed by the computation of the individual OECFs for each channel.

6 Results

In this section, the colour correction methods are evaluated on image pairs. The image pairs are blended using a checkerboard technique. The overlap area can be considered as a checkerboard; the black cells of the checkerboard contain the pixels from image I_0 and the white cells of the checkerboard contain the pixels from image I_1 (see Figure 4). In terms of quality, the checkerboard blending is about the worst that can be done to render the final mosaic and is solely aimed at emphasising the mismatch in the registration.

6.1 Preferred estimation method

6.1.1 The analog system

In the first example, we use two pictures that have been scanned from slides. The slides are on Kodachrome 64 film, whose characteristics are given by the D-LogH curve of Figure 2, which relates film density to log-exposure. The slide is passed through a scanner whose OECF has been computed by the characterisation process of Section 3.1. The result of the mosaicing is shown in Figure 5. The pictures have been taken using a fixed exposure and focal length. Figure 5(a) shows the original mosaic without any colour correction. The pixel values are the one originally delivered by the scanner. This results shows what any standard mosaicing algorithm would do. In this example, it is not clear to us whether the change in exposure occurred in the scanner or because of the camera shutter imperfections. Figure 5(b) shows the mosaic processed with colour correction, but without taking vignetting into account. Figure 5(c) shows the final mosaic, with vignetting correction. The only difference between 5(b) and 5(c) is the vignetting correction. We used both a calibration slide and the approximation of the scanner by Eq. (3), and, for our scanner[7], the results were very close (we could not distinguish them visually).

6.1.2 The polynomial OECF

The images are taken with a digital still camera. We assume that the camera OECF follows the polynomial model of Section 3.3 with 5 coefficients: $S(I) = a_1 I + a_2 I^2 + a_3 I^3 + a_4 I^4 + a_5 I^5$. The characteristics of the camera (the OECF inverse) has been computed using the two images of Figure 6. The result is shown in Figure 7.

The next example shows an old building, taken with the camera set on automatic mode. The camera used a shutter speed of $1/60$ sec for the lower building and $1/200$ sec for the top. The aperture was kept constant, and the adopted white point changed (the camera was set on automatic white balancing). Figure 8(a) shows the mosaic of

---

[6]An OECF is always monotonic.
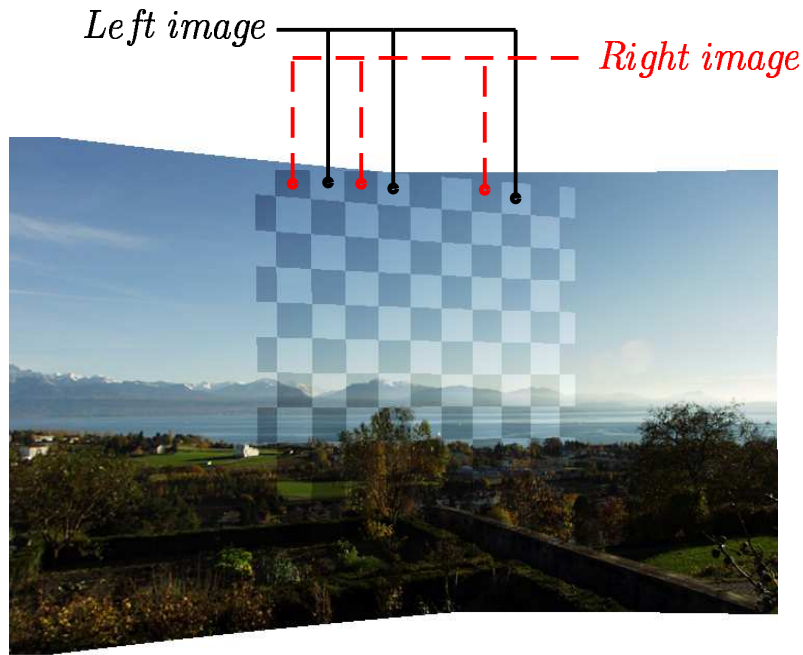
[7]We used a Nikon-LS2000 slide scanner.

Figure 4: Illustration of the checkerboard blending technique. The images are mixed like on a checkerboard, as the goal is to best visualise the colour mismatch.

this building without any colour correction. Figure 8(b) shows a correction that assumes a fixed white point and a varying exposure, while figure 8(c) assumes a varying white point. The images were aligned using a rotational model with lens distortion correction [23, 11], and the alignment parameters used to render these images are the ones found by computing image 8(c). It is worth mentioning that the colour correction improves the alignment of the pictures, as illustrated in Figure 9, where one picture has been computed by assuming a fixed white point and the other by assuming a varying white point. We should also mention that a multi-resolution pyramid has been used and the complexity of the model is adapted to the resolution. This means, for example, that to get the image of Figure 8(c), we first assume a fixed white point, and do not adapt the focal length nor the distortion parameter at the lowest resolution. Then, while the resolution of the pictures increases, we allow the white point, the focal length and the distortion to vary. If we had adapted every parameter from the start, the system would have ended in a local minimum, and the result would not have been the one expected.

Another mosaic estimated with the exact same procedure is depicted in Figure 10. The focal length used is about half as long (i.e. 45mm versus 110mm), making the alignment more difficult. Aligning pictures taken with a wide angle lens is more difficult because of the increased distortion of the image—especially when using a lens with a zoom—and because of the greater impact that a bad alignment of the optical centre can have on the reconstructed image. Figure 10(a) shows the mosaic of the original pictures, without any colour adjustments, 10(b) shows the mosaic adjusted with the nominal shutter speed given by the camera and 10(c) shows the white point corrected mosaic. In the original pictures shown in Figure 11, there is a big difference in the white point. The mosaic has been rendered with the motion parameters of Figure 10(c). There is a gradual improvement with the complexity of the model.

6.2 Global motion and OECF estimation method

The global estimation method is applied using the Laguerre OECF model. The OECF shape, as well as the exposure and white point settings are estimated along with the picture alignment. The results are shown in figure 8 and 10. In figure 8, the Laguerre OECF model was used to output figure 8(d) and delivers equivalent results to the polynomial model used to get figure 8(c). In figure 10, the Laguerre model (fig. 10(d)) shows some limitations compared to the polynomial model (fig. 10(c)). Nevertheless, it is more reliable than the polynomial model and requires less trial-and-error. We also have to mention that the cameras used in our experiment had functions that

(a)

(b)

(c)

Figure 5: Colour correction using slides with prior calibration. (a) Original mosaic (b) Exposure correction (c) Exposure and vignetting correction. The figure shows a step-wise improvement with the model complexity.

Figure 6: Images used to fit a polynomial OECF model of the camera. The images have the same white point but different exposures.
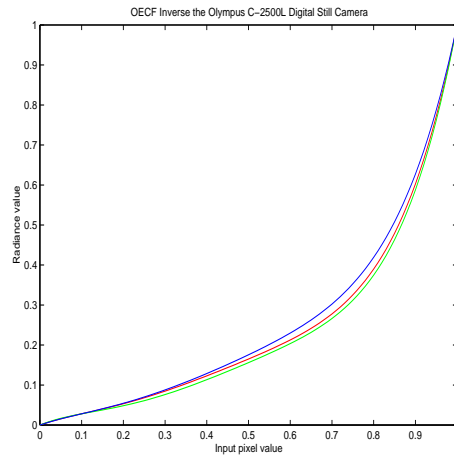

Figure 7: OECF Inverse of the Olympus C-2500L camera obtained by fitting a polynomial model of order 5 to the images of Figure 6.

Figure 8: Illustration of the colour correction process. (a) original mosaic without correction. (b) Shutter speed compensation—polynomial OECF (c) shutter speed and white point compensation—polynomial OECF (d) shutter speed and white point compensation—Laguerre OECF.

(a)	(b)

Figure 9: Improving the alignment using colour correction. (a) shows the alignment of the picture for a fixed white point assumption; notice the mis-alignment of the window in the middle-right. (b) Alignment using a varying white point. This example illustrates that the colour correction helps the motion estimation algorithm on the ill-posed problem of the focal length and the lens distortion parameter estimation.

are well modelled by the Laguerre OECF, which is not necessarily the case for some other cameras.

6.3 Iterative trial-and-error estimation method

The polynomial OECF model, and its estimation method has the particularity of either working quite well or failing totally. An example of failed estimation is given in Figure 12(b). In all the examples we tested, when the estimation failed, it delivered a non monotonic curve. This makes a failure easy to detect and in the case when only one or two curves failed (out of the three colour channels) it is possible to use the curve that succeeded and apply it to the other channels. This has been done to compute Figure 12(c) where the red curve has been computed from the blue and green. The curves found for 12(b) are depicted in Figure 12(d). To be robust, the system has to verify after each OECF estimation if it succeeded or not: in practice the estimate is retained only if it delivers at least one monotonic curve and if the resulting registration error is lowered.

7 Conclusions

This paper presents a set of techniques that enable to recover the relative scene-referred data—i.e. an image linearly related to the scene colours—from a colour rendered image and renders the final scene as a single picture. From the relative scene-referred data, the exposure differences, the white point and the vignetting can be conveniently corrected, resulting in a colour consistent mosaic. Because relative scene referred data cannot be displayed as is, we rendered it using the same transformation applied by the camera to produce the original pictures. The purpose of this last step enables to show the results in a visually sensible way, but could have been performed in many other ways.

We want to emphasise that this paper presents examples of parametric models that are used to retrieve the raw data from the pictures. The important matter are not the models themselves, but rather the approach that is used. Ideally, a good colour correction algorithm should use all the information available from the camera manufacturer, and estimate only the parameters that may change between the pictures.
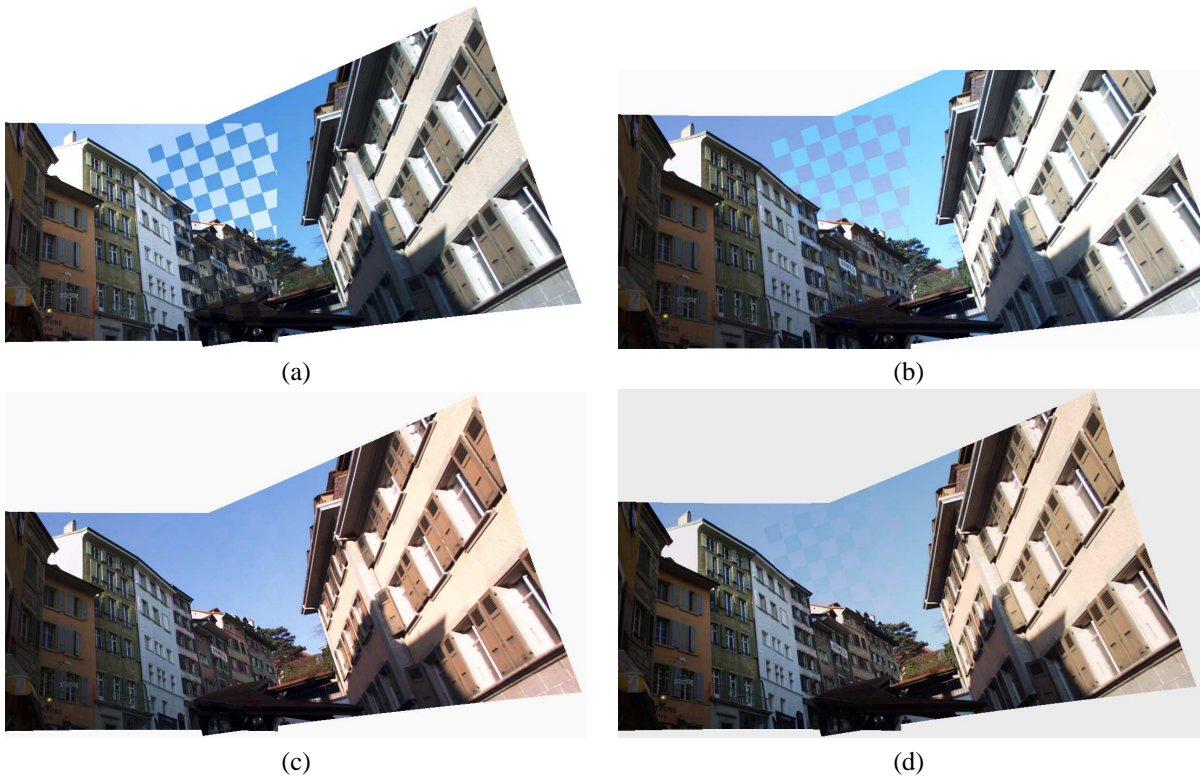
Figure 10: Illustration of the colour correction process. (a) original mosaic without correction. (b) Shutter speed compensation with parameters given by the camera—polynomial OECF (c) shutter speed and white point compensation—polynomial OECF. (d) shutter speed and white point compensation—Laguerre OECF.

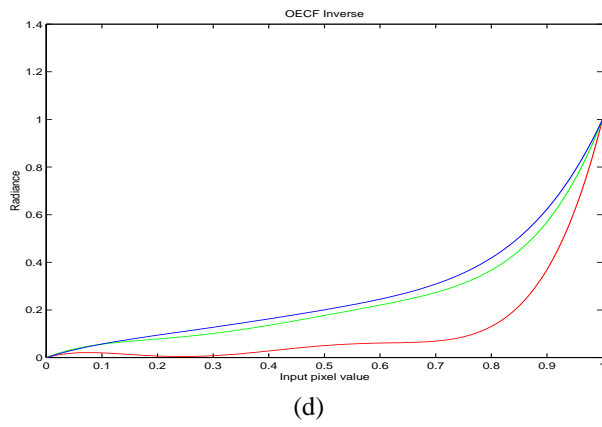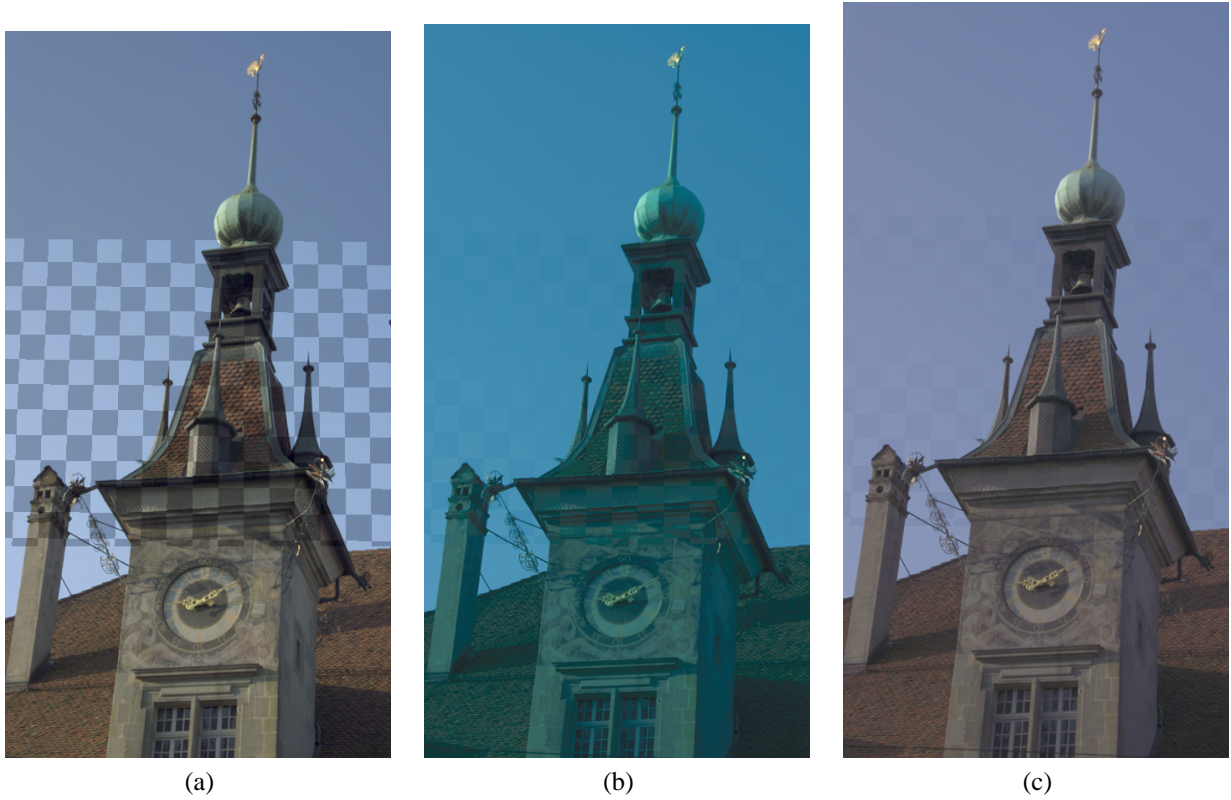Figure 11: original pictures used to build the mosaic of Figure 10.

Figure 12: Simultaneous polynomial OECF and motion estimation. The OECF is estimated on the image overlap. (a) Original mosaic. (b) Correction with OECF estimation—the estimation of the red curve failed. (c) The red curve is computed from the blue and green. (d) OECF plot of image (b).
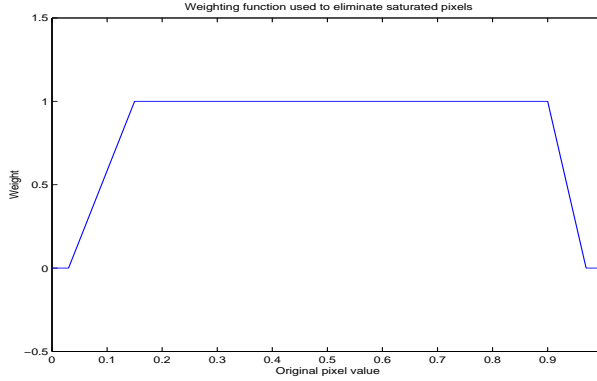
Figure 13: Weight values used to discount saturated pixels. The influence of each pixel in the registration process is decreased in the border of the image range. Possible pixel value range from 0 to 1.

Among the different techniques used to correct the colours, we preferred the one that uses a set of still pictures to compute the camera characteristics, like in Figure 8. This techniques gives excellent results, and once the OECF is estimated, the colour correction is robust enough to be performed blindly on the pictures. When assembling a set of picture taken with an unknown camera, then the Laguerre OECF is the best choice: it delivers excellent results in most situations. For achieving an even better registration, one can use the polynomial model, but at the price of verifying if the model succeeded at each iteration step and also with the risk that the model never succeeds, in which case the Laguerre OECF has to be used.

A   Avoiding saturated pixels

The following presents some implementation details related to the OECF and exposure estimation method. We are interested in knowing how much information is contained in a pixel pair about the lighting conditions, about the camera settings, etc. If a pixel appears saturated in one of the images[8], it will only give upper- or lower-bound information about the exposure difference of the image pair. If this pixel is used in the same way than the "good" ones, that is, the pixels that have values in the middle range of the camera, then the estimation will fail. By writing the motion estimation Eq. (12), the reliability of a pixel pair is embodied by the factor W_i, who multiplies each line of the equation system by a different factor. This factor depends on the two values delivered by the camera: for each image j of the pair, in each pixel location $\mathbf{p}_i$, we can compute a weight value $w_{j,i}$ that tends toward 0 in the border of the dynamic range of the camera, and take values of 1 in the middle range of the camera, thus giving less weight to potentially noisy pixels. Since we need one single weight value for each pixel of the image pair, the two weights $w_{0,i}$ and $w_{1,i}$ have to be combined into a single one. The resulting weight for the pixel pair is computed as

$$\frac{1}{W_i} = \frac{1}{w_{0,i}} + \frac{1}{w_{1,i}}. \tag{16}$$

The underlying assumption is that $w_{\cdot,i}$ is the inverse of variance of a Gaussian random variable. Indeed, if the pixel value is a random variable distributed as $\mathcal{N}\left(\mu_{\cdot,i}, 1/w_{\cdot,i}\right)$,—where $1/w_{\cdot,i}$ is the *variance*—then the residual is also a Gaussian random variable distributed as $\mathcal{N}\left\{\mu_{0,i} - \mu_{1,i}, \frac{1}{w_{0,i}} + \frac{1}{w_{1,i}}\right\}$, leading to Eq. (16).

An example of the function we used is depicted in Figure 13[9]. This weight is computed independently for each colour channel.

Remark – Although the function of Figure 13 seems somewhat arbitrary, an attempt to precisely model the noise as the variance of a gaussian random variable by just using a set of differently exposed photographs has been tried, but did not deliver useful results. The modelling involves the use of the exposure parameters of each picture, and we could not use the approximation $\frac{\partial W_i}{\partial \theta} \simeq 0$ to solve to optimisation in Section 5.3. The added complexity of the motion estimation makes the approach computationally more expensive and, unless precise estimates of the noise variance are available, is not worth the effort.

---

[8]A pixel is saturated if it has a value of 0 or 255 in a 8 bits per channel camera.

[9]In practice, the weight never reaches 0, but a value of 10^{-5} is used instead in order to avoid rejecting every pixel of the image pair if the pair has very different exposure parameters.

B   Detecting mis-alignments

The algorithm proposed by this paper performs a simultaneous motion and camera parameter estimation. As a consequence, it cannot assume that the images are properly aligned when it computes a camera OECF. Thus, the parts of the overlap that are well aligned, or the parts that are not sensitive to bad alignments should be used to estimate the OECF. Regions with fine details are very sensitive to mis-alignments, whereas smooth regions are only sensitive to colour mismatches. To discount the influence of a colour mismatch, we first filter the image with a highpass filter. Then, to annihilate the influence of the difference in exposure, we normalise the dynamic range of the two pictures. Finally, the comparison of the high-pass images provide a confidence measure that determines which part of the image is sensitive to mis-alignments or, conversely, which part of the image should be used to perform the estimation.

In practice the images are filtered with the derivative of a gaussian filter (with $\sigma = 1.25$ pixels) and normalised such that the variance of the images in the overlap area are equal. The alignment weight is computed as

$$W_i^{(a)} = \frac{1}{|f(I_0(\mathbf{p})) - f'(I_1(\mathbf{p}'))| + Q},$$

where $f(I_0(\mathbf{p}))$ is the filtered image I_0 and $f'(I_1(\mathbf{p}'))$ is the filtered image I_1 with normalised variance. Q is a small positive number equal to the quantisation step of the image ($Q = 1/255$ for an 8 bit image). The quantisation step is used to prevent the weight to go to infinity and also because an error of zero can be generated by pixel values separated by (at most) an amount of Q. The mis-alignment weight is combined to the one used in Appendix A, using the same approach, that is, (by borrowing the notations of Appendix A)

$$\frac{1}{W_i} = \frac{1}{W_i^{(a)}} + \frac{1}{w_{0,i}} + \frac{1}{w_{1,i}}.$$

We want to emphasise that this weight is only used to estimate the OECF in Section 5.4, and is not used in the context of Section 5.3. The reason is, in the last case, that the algorithm needs the mis-aligned pixels to get a better alignment.

Following the argument of this section, a side question appears naturally: Why not use the high-pass images to perform the motion estimation and do the colour correction once the pictures are aligned? The reason we did not proceed in that way is related to the shape of the error surface: The motion algorithm performs a gradient descent that, given an error surface (in n-dimensional space), finds a path that follows the surface downwards. If the surface is irregular, and contains local minima, the algorithm is likely to fail. By filtering the image with a high-pass filter, the error surface gets also high-pass filtered and becomes irregular ; consequently, the system requires to have a better initial estimate to converge to the right solution (see [11, sec. 2.8] for more information).

References

[1] R.W.G. Hunt. *The Reproduction of Colour in Photography, Printing and Television*. Fountain Press, 5 edition, 1995.

[2] Jack Holm, Ingeborg Tastl, Larry Hanlon, and Paul Hubel. *Colour Engineering: Achieving Device Independent Colour*, chapter Color Processing for Digital Photography, pages 179–217. John Wiley & Sons, 2002.

[3] ISO 14524. *Photography - Electronic Still Picture Cameras- Methods for Measuring Opto-Electronic Conversion Functions (OECFs) (Item 192)*. 1999.

[4] IEC 61966-2-1. *Multimedia systems and equipment* – Colour measurment and management – Part 2.1: Colour management – Default RGB colour space – sRBG. 1999.

[5] J. Von Kries. Chromatic adaptation. In *Festschrift der Albrecht-Ludwigs-Universität*, 1902. Translation : D.L. MacAdam, "Colorimetry Fundamentals", SPIE Milestone series, Vol. MS 77, 1993.

[6] R.W.G. Hunt. *Measuring Colour*. Fountain Press, England, 3 edition, 1998.

[7] Sabine Süsstrunk, Jack Holm, and Graham D. Finlayson. Chromatic adaptation performance of different RGB sensors. In *IS&T/SPIE Electronic Imaging*, volume 4300, pages 172–183, San Jose, CA, 2001.

[8] ISO16067 WD : 1999. *Electronic Scanners for Photographic Images - Spatial Resolution Measurements, Part 1: Scanners for Reflective Media.* 1999.

[9] Tomoo Mitsunaga and Shree K. Nayar. Radiometric self calibration. *Proc. Of Computer Vision and Pattern Recognition*, 1:374–380, June 1999.

[10] R. Kingslake. *Optical Systems Design*. New York, 1983.

[11] David Hasler. *Perspectives on Panoramic Photography*. PhD thesis, Swiss Federal Institute of Technology (EPFL), 2001.

[12] A.E. Johnson and Sing Bing Kang. Registration and integration of textured 3-d data. In *International Conference on Recent Advances in 3-D Digital Imaging and Modeling, 1997.*, pages 234–241, 1997.

[13] C. Guestrin, F. Cozman, and E. Krotkov. Fast software image stabilization with color registration. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 19–24, vol. 1, 1998.

[14] K.J. Hanna, H.S. Sawhney, R. Kumar, Y. Guo, and S. Samarasekara. Annotation of video by alignment to reference imagery. In *IEEE International Conference on Multimedia Computing and Systems*, pages 38–43, vol.1, 1999.

[15] A. Majumder, Zhu He, H. Towles, and G. Welch. Achieving color uniformity across multi-projector displays. In *Visualization 2000*, pages 117–124, 2000.

[16] H. Rushmeier and F. Bernardini. Computing consistent normals and colors from photometric data. In *Second International Conference on 3-D Digital Imaging and Modeling*, pages 99–108, 1999.

[17] Y. Yu, P. J. Debevec, J. Malik, and T. Hawkins. Recovering high dynamic radiance maps from images. In *SIGGRAPH 99, Computer Graphics Proceedings*, pages 215–224, Los Angeles, 1999.

[18] Steve Mann and Rosalind W. Piccard. Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *IS&T 46th Annual Conference*, pages 422–428, May 1995.

[19] Steve Mann. Comparametric equations with practical applications in quantigraphic image processing. *IEEE Transactions on Image Processing*, 9(8):1389–1406, 2000.

[20] P. J. Debevec and J. Malik. Recovering high dynamic radiance maps from images. In *SIGGRAPH 97*, pages 369–378, 1997.

[21] Mark A. Robertson, Sean Borman, and Robert S. Stevenson. Dynamic range improvement through multiple exposures. In *IEEE International Conference On Image Processing, ICIP*, pages 159–163, 1999.

[22] G.A.F. Seber and C.J. Wild. *Nonlinear Regression*. John Wiley and Sons, New York, 1989.

[23] H. S. Sawhney. True multi-image alignment and its application to mosaicing and lens distortion correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(3):235–243, March 1999.

Figure 14: David Hasler graduated in 1996 and got his Ph.D. in 2001, both from the Swiss federal institute of technology (EPFL) in Lausanne. During his undergraduate studies, his was exchange student at Carnegie Mellon University in Pittsburgh, USA, and did his master's thesis at Daimler-Chrisler research, in Ulm, Germany. He also did two summer internships, one at NASA Ames, the other at Sony research, both in California. His main research area is vision and digital photography. Early 2002, he worked in a startup company—Genimedia SA—on video quality. He joined now GretagMacbeth and is active in colour management.

Figure 15: Sabine Süsstrunk is Assistant Professor for Imaging and Visual Representation in the Audiovisual Communications Laboratory at the Swiss Federal Institute of Technology (EPFL) in Lausanne, Switzerland. Her main research areas are color imaging, image quality metrics, and digital image archiving. She received a BS in Scientific Photography from the Swiss Federal Institute of Technology (ETHZ) and a MS in Electronic Publishing from the Rochester Institute of Technology (RIT). She is currently working part time on her Ph.D. degree at the School of Information Systems, University of East Anglia, Norwich, Great Britain.