

# Colour handling in panoramic photography

David Hasler and Sabine Süsstrunk  
Audiovisual Communication Laboratory  
Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland

## ABSTRACT

Nowadays, the ability to create panoramic photographs is included with most of the commercial digital cameras. The principle is to shoot several pictures and stitch them together to build a panorama. To ensure the quality of the final image, the different pictures have to be perfectly aligned and the colours of the images should match. While the alignment of images has received a lot of attention from the computer vision community, the mismatch in colours was often ignored and handled using smooth transitions from one picture to the next to mask the mismatch. This paper presents a method to simultaneously estimate the alignment of the pictures and the colour transformation between them. By estimating the colour transformation from the scene to the pixels, the method is able to remove the mismatch in colours of the different images, and thus leads to better image quality.

**Keywords:** Digital photography, Colour correction, Mosaicking, Panoramas, OECF, Motion Estimation.

## 1. INTRODUCTION

The purpose of panoramic photography is to show a picture with a wide angle of view. The wide viewing angle corresponds better to what an observer sees of a scene. To create such wide angle pictures, one can either use specialized hardware, or can take several images with a standard camera and stitch them together on a computer. When the images are stitched together, it shouldn't be possible to see that the resulting panoramic image has been taken with several pictures. To achieve this, nothing should change in the scene during capture, the images have to be aligned properly, and the colours of the different images should match. Aligning images is a well known problem and has first been solved by aligning consecutive fields in a video sequence. On the other hand, the mapping of colours in an image has been developed by the photographic community for more than a century; their goal is to produce the best looking picture, not the one that best represents the physical properties of the scene. As a consequence, the colours in the different pictures do not always match. To construct a good panoramic image, the best approach is to retrieve the physical properties of the objects in the scene and then to render the scene as single picture. Since the physical properties of an object do not depend on the picture it appears in, this approach gets rid of the mismatch in colours. The physical properties that are needed are the Red, Green and Blue reflectance components of each entity in the scene.

The algorithm presented in this paper tries to simultaneously estimate the alignment of the pictures, and the colour transformation between them. There are two approaches that are used: The first uses scanned slides, and requires some off-line calibration data about the slide and the scanner characteristics. The only parameters that it computes are the one that can be changed on the camera during the capturing process, like for example the exposure (shutter speed and aperture). The second method uses data coming from a digital still camera and estimates the transformation induced by the camera. From a set of pixels, it computes the radiance values of the object that was originally recorded by the camera sensor (CCD), i.e. it tries to estimate the Opto-Electronic Conversion Function (OECF). This method makes strong assumptions about the shape of that function. Whether these assumptions are wise or not is controversial. The bottom line is that digital still cameras are getting more and more complex to better suit human perception and each of the assumptions made in the mosaicking process will fail sooner or later\*, so that in the not-so-far future, any good mosaicking software will have to use the knowledge of how the camera maps the pixels values and estimate the parameters that may change from one picture to the other. Nevertheless, the framework presented in this paper remains valid; only the parametrisation of the camera characteristics will change.

---

\*Even the assumption that an object is mapped to the same pixel value if the aperture, shutter speed, white balance, focus and zoom are kept constant is already wrong in some of the new cameras.

## 2. FROM THE SCENE TO THE PIXELS

Technically, photography is all about light. The light is generated by a *light source* (the sun or a lamp), hits a particular object of the scene, is reflected and transformed by the object, and reaches the sensor of the camera after passing through the lens. The way the light is transformed by an object depends on the *spectral reflectance properties* of the object. The light is then recorded by the camera sensor (CCD) that measures for each pixel three light components: the Red, Green and Blue component. A CCD has a linear behaviour, the value for each pixel is proportional to the illuminance at the sensor location. At this level, the data is called *raw data*. Then, the camera usually performs a white balancing followed by a non-linear mapping, described by the *Opto-Electronic Conversion Function* (OECF), also sometimes referred to as *gamma correction*. It also performs some transformations to get a “better looking image” - for example a gain in the colour saturation - that are ignored here. We will assume, for the rest of the paper, that the final pixel value only depends on the value read by the CCD (and not on the surrounding pixels) and that the OECF does not change from picture to picture. This assumption is not correct for high end digital cameras.

In this description, we omitted the linear transform that is used to map the image from camera RGB to another colour space - for example sRGB - because we assume it will not affect the mismatch between the colours in the different images in a significant way.

### 2.1. White balancing

The colour of an object depends on its spectral reflectance properties, on the lighting conditions and on the state of adaptation of the (human) observer. The adaptation of the human observer is a complex mechanism, but for our purpose, the only phenomenon that we are concerned with is the *white balancing*.<sup>1</sup>

The purpose of white balancing can be explained by an example: a white piece of paper appears to be white under sunlight as well as under incandescent light. From a physical point of view, the power spectrum coming from the white piece of paper is equal to the spectrum of the illuminant, and is different for the two situations. Nevertheless, it will appear white to an observer, because of *chromatic adaptation*.<sup>2</sup> Chromatic adaptation is the ability of the human visual system to discount the colour of the illuminant, and to approximately preserve the appearance of an object. If the white balancing is done properly in the camera, the colour of the white piece of paper (i.e. the RGB values) will be the same when it is taken under sunlight or under incandescent light<sup>†</sup>. Most of the current cameras perform a Von Kries adaptation: Given the  $R'_w G'_w B'_w$  values of the light source, and  $R''_w G''_w B''_w$  the value of the light used to view the image, they apply a gain factor on the three colour channels independently, while the image is still in linear space (i.e. it performs the transformation on the *raw data*):

$$\begin{aligned} R &\mapsto R \cdot \frac{R''_w}{R'_w} \\ G &\mapsto G \cdot \frac{G''_w}{G'_w} \simeq G \\ B &\mapsto B \cdot \frac{B''_w}{B'_w}. \end{aligned}$$

In practice the green channel is often left unchanged.

### 2.2. The transfer function of a slide and scanner system

Because film photography is still the best way to get high quality pictures at a decent price, we first used scanned diapositives as our input pictures. The film characteristics<sup>‡</sup> are given by a function called the *D-logH* curve that relates film density to the log Exposure. The scanner illuminates the diapositive and integrates the ratio of light that goes through the diapositive at a given photosite. The scanner is characterized by its Opto-Electronic Conversion Function (*OECF*).<sup>3</sup> The OECF can be characterized using a calibration slide containing patches of known density. It turned out that a good approximation of our scanner was given by the following formula: Let  $I$  be the pixel value

<sup>†</sup>Here, we are not considering incomplete adaptation.

<sup>‡</sup>These characteristic curves can be obtained by the film manufacturer.

normalized such that  $I \in [0..100]$ . We assume that  $I$  is related to the lightness  $L^*$  of the pixel through the relation  $I = (L^*)^{\frac{1}{\gamma}}$ , where gamma  $\gamma$  is a parameter that can be set on the scanner. Setting  $\gamma = 1$ , gives

$$I = L^* \quad (1)$$

$L^*$  is related to the relative luminance  $\frac{Y}{Y_n}$  of the original illuminated by the light source of the scanner through<sup>4</sup>

$$\begin{aligned} L^* &= 116 \cdot \sqrt[3]{\frac{Y}{Y_n}} - 16, & \frac{Y}{Y_n} > 0.008856 \\ L^* &= 903.3 \cdot \left(\frac{Y}{Y_n}\right), & \text{otherwise} \end{aligned} \quad (2)$$

Then,  $\frac{Y}{Y_n}$  is related to the film density  $D$  by

$$D = -\log_{10} \left( \frac{Y}{Y_n} \right) \quad (3)$$

and finally, the Exposure (i.e. light power) is given by the  $D$ - $\log H$  curve of the film, which enables to build the Slide and Scanner *Transfer Function Inverse*

$$H = S(I) \quad (4)$$

By assuming a constant exposure over one whole image, the illuminance  $E$  is related to the exposure  $H$  and to the shutter time  $t$  through

$$H = E \cdot t. \quad (5)$$

Now in a set of pictures, the only unknown parameter that needs to be estimated is the ratio of the parameters  $t$  between the pictures.

### 2.3. The Transfer function of a digital still camera

Here we consider pictures that are taken with a digital still camera. The digital still camera converts the light falling onto its sensor (CCD) directly into digital values, and can be characterized by its Opto-Electronic Conversion Function (OECF).<sup>5</sup> Nowadays, an OECF is considered as highly confidential data by the camera manufacturers and is not available to the common user. We will assume that the OECF of a camera is similar to the *gamma correction* curve applied in the video standards, such as the ITU-R BT.709,<sup>6</sup> and has the following form:

$$\begin{aligned} V' &= 1.055 \cdot V^{\frac{1}{2.4}} - 0.055 & \text{if } V > 0.0031308 \\ V' &= 12.92 \cdot V & \text{else} \end{aligned}$$

where  $V$  is the raw data delivered by the CCD and  $V'$  is the output pixel value. It is an exponential function extended by its tangent (when  $V \leq 0.0031308$ ) to make it pass through zero. In our model, we make the assumption that the OECF of the camera can be described by

$$V = (1 + a) \cdot V'^{\frac{1}{b}} - a \quad (6)$$

where  $a$  and  $b$  are the two unknown parameters of the model. This function is also extended by its tangent to make it pass through the origin, which is a straight forward task to accomplish numerically. We did not express it explicitly in Equation 6 because the intersection point between the exponential and its tangent that crosses the origin has no closed form. Since the algorithm needs to compute the raw data pixel value  $V$  out of the pixel value  $V'$ , the function it tries to estimate has the form

$$V = \left( \frac{V' + a}{1 + a} \right)^b \quad (7)$$

Even if it is not true for high end digital camera, we will assume that this function is the same for every picture.

In the following section, the goal is to estimate the OECF inverse using two differently exposed pictures. Since we are interested in the difference between these pictures - to correct for colour mismatch - we can consider that one image has an exposure of 1 and the other has an exposure of  $t$ , which is the ratio between the two real exposures.

Let  $S$  be the inverse of the OECF used to produce image  $I_0$ , and  $G$  be the inverse of the OECF used to produce image  $I_1$ . Both functions have the same shape, but  $G$  accounts for an exposure of  $t$  such that:

$$\begin{aligned} S(V') &= \left( \frac{V' + a}{1 + a} \right)^b \\ G(V') &= \frac{1}{t} S(V') \end{aligned}$$

The output of these two functions is the illuminance of the object of the scene measured at the sensor location (i.e. the raw data) when taken with an exposure of 1. Therefore, the values do not depend on the picture they appear in.

We want to emphasize that this section is presented as an example of what can be done, but that ideally, a good algorithm should use all the information available from the camera manufacturer when describing function  $S$ , and estimate only the parameters that may change between two pictures.

### 3. VIGNETTING

Vignetting is responsible for the image being darker in the border. On a high quality camera, this phenomenon is mainly due to a geometric effect: the  $\cos^4$  effect.<sup>7</sup> Indeed, if the camera pictures a scene of constant luminance, the illuminance hitting the sensor varies according to the cosine to the fourth power of the viewing angle. Thus,

$$\begin{aligned} E_m &= E_o \cdot \cos^4 \alpha \\ \tan(\alpha) &= \frac{r_{\mathbf{p}}}{f} \end{aligned} \tag{8}$$

where  $E_m$  is the illuminance measured by the sensor,  $r_{\mathbf{p}}$  the distance of the pixel from the optical center 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 in the optical center. By computing  $E_o$  for each pixel, we get an image without vignetting. Other phenomena related to vignetting, like its dependence on the aperture have been ignored here.

## 4. MOTION ESTIMATION ALGORITHM

### 4.1. Traditional approach: Optical flow

Traditionally, optical flow techniques have been developed by Horn and Schunk<sup>8</sup> to solve the problem of translational motion between several frames in a video sequence. It assumes that each point  $\mathbf{p}$  in image  $t$  is a translated version of the corresponding point in image  $t + \Delta t$ . The image Luminance is described as

$$I(\mathbf{p}, t) = I(\mathbf{p} + \Delta\mathbf{p}, t + \Delta t), \quad \forall(\mathbf{p}, t). \tag{9}$$

where  $\mathbf{p}$  is a two-dimensional position vector. This assumes that a particular object of the scene is mapped into the same luminance values on a frame and in the next, which is reasonable in video applications. By setting the flow velocity  $\mathbf{u}_{\mathbf{p}} = \frac{\Delta\mathbf{p}}{\Delta t}$ , and computing the Taylor expansion of equation (9) around point  $(\mathbf{p}, t)$  by keeping only the first order terms leads to the well known optical-flow equation

$$\mathbf{u}_{\mathbf{p}}^T \frac{\partial I}{\partial \mathbf{p}} + \frac{\partial I}{\partial t} \simeq 0. \tag{10}$$

This equation has two unknowns (the two components of vector  $\mathbf{u}_{\mathbf{p}}$ ), and as many equations as there are pixels in the image overlap. Now, the flow  $\mathbf{u}_{\mathbf{p}}$  might not be constant over the whole image. If the flow is allowed to have an arbitrary shape for every pixel, the system would have twice as many unknowns as equations. This problem is solved using a parametric model: It can be assumed, for example, that the scene is captured by a camera who undergoes a rotation. Then, provided that the scene is static, the whole correspondence problem can be expressed as a function of the rotation of the camera (3 angles) and its focal length (i.e. with 4 unknowns).

## 4.2. Parametric methods

We will reformulate the problem introduced in section 4.1 in a slightly different and more general way: Let  $I_0(\mathbf{p}), I_1(\mathbf{p}')$  be an image pair. We will suppose that both images can be perfectly super-imposed by a warping process:

$$S_\theta [I_0(\mathbf{p})] = G_\theta [I_1(U_\theta [\mathbf{p}])]. \quad (11)$$

where  $S$  and  $G$  are the OECF inverses introduced in Section 2.3,  $\theta$  is a vector containing the parameters of the model (for example the three rotation angles of the camera, the focal length, the  $a$  and  $b$  parameter of the OECF and the exposure ratio  $t$ ) and  $\mathbf{p}$  is the position in the image<sup>§</sup>. This supposes that the scene is static. The function

$$\mathbf{U}_\theta : \mathbf{p} \mapsto \mathbf{p}' \quad (12)$$

describes the rigid motion model. The unknowns are the set of parameters  $\theta$ . If we suppose some smoothness in the images  $I_{0,1}$ , we can find an updating rule for  $\theta$  with a standard descent algorithm<sup>9</sup> by minimising an objective function  $h$ , which is, for example, the mean squared error function:

$$h(\theta) = \frac{1}{n} \sum_{i=1}^n \rho[r_i(\theta), \sigma] = \frac{1}{n} \sum_{i=1}^n \|r_i[\theta]\|^2 \quad (13)$$

$$r_i(\theta) = S_\theta [I_0(\mathbf{p})] - G_\theta [I_1(U_\theta [\mathbf{p}])] \quad (14)$$

where  $\mathbf{p}_i$  are the pixel locations on the overlapping parts of the images,  $n$  the number of pixels,  $\sigma$  a confidence value and  $r_i$  the error at each pixel location, also called *residual*. We will come back to the choice of the function  $\rho(\cdot)$  later on. The minimum of function  $h(\theta)$  satisfies

$$\frac{\partial h(\theta)}{\partial \theta} = \mathbf{0}. \quad (15)$$

Equation (15) is a non-linear system of equations, and is most of the time too complex to be solved analytically. Therefore, a recursive solution is applied: let's suppose that an initial parameter set  $\theta_0$  is available; the goal is to find an updating rule such that  $\rho(\theta) < \rho(\theta + \delta\theta)$ , where  $\delta\theta$  is the correction to the parameter set. By expanding function  $h(\theta)$  by Taylor series, and keeping only the second order terms, we get:

$$\begin{aligned} h(\theta + \delta\theta) &\simeq h(\theta) + \frac{\partial h(\theta)}{\partial \theta} \delta\theta + \frac{1}{2} \delta\theta^T \frac{\partial^2 h(\theta)}{\partial \theta \partial \theta^T} \delta\theta \\ &\triangleq h(\theta) + \mathbf{J}(\theta) \delta\theta + \frac{1}{2} \delta\theta^T \mathbf{H}(\theta) \delta\theta. \end{aligned} \quad (16)$$

Where  $\mathbf{J}(\theta)$  is the *jacobian* matrix and  $\mathbf{H}(\theta)$  the *hessian* matrix. According to this approximation, the correction  $\delta\theta$  is found by setting the derivative  $\frac{\partial h(\theta + \delta\theta)}{\partial \delta\theta}$  to zero, thus

$$\mathbf{J}(\theta) + \mathbf{H}(\theta) \delta\theta = \mathbf{0}$$

The correction  $\delta\theta$  to the parameter set  $\theta$  leads to a minimum iff the matrix  $\mathbf{H}$  is positive definite. Let's express the different terms of equation (16) in terms of  $I_0, I_1$  and  $\theta$ : Let  $\mathbf{U}_{\theta i} \triangleq \mathbf{U}_\theta [\mathbf{p}_i]$ , then

$$\begin{aligned} \mathbf{J}_i(\theta) &= \frac{\partial \rho(r_i, \sigma)}{\partial \theta} \\ &= \frac{\partial \rho(r_i, \sigma)}{\partial r_i} \frac{\partial r_i}{\partial \theta} \\ &= \frac{\partial \rho(r_i, \sigma)}{\partial r_i} \frac{\partial}{\partial \theta} \{S_\theta [I_0(\mathbf{p})] - G_\theta [I_1(\mathbf{U}_{\theta i})]\} \end{aligned} \quad (18)$$

$$= \frac{\partial \rho(r_i, \sigma)}{\partial r_i} \left\{ \dot{S}_\theta [I_0(\mathbf{p})] - \dot{G}_\theta [I_1(\mathbf{U}_{\theta i})] - \nabla G_\theta [I_1(\mathbf{U}_{\theta i})] \frac{\partial \mathbf{U}_{\theta i}}{\partial \theta} \right\}. \quad (19)$$

<sup>§</sup>This formulation assumes that the model doesn't require to modify the geometry of image  $I_0$ , and thus is not suitable to correct for lens distortion.

The Hessian can be found similarly, and is approximated by<sup>¶</sup>

$$\mathbf{H}_i(\theta) \simeq \frac{\partial \rho(r_i, \sigma)}{r_i \cdot \partial r_i} \mathbf{A} \cdot \mathbf{A}^T \quad (20)$$

$$\mathbf{A} \triangleq \dot{S}_\theta [I_0(\mathbf{p})] - \dot{G}_\theta [I_1(\mathbf{U}_{\theta i})] - \nabla G_\theta [I_1(\mathbf{U}_{\theta i})] \frac{\partial \mathbf{U}_{\theta i}}{\partial \theta}. \quad (21)$$

The complete Hessian is found by summing up the terms for every pixel:  $\mathbf{H}_i(\theta) = \frac{1}{n} \sum_{i=0}^n \mathbf{H}_i(\theta)$ . Finally, the recursion algorithm works as follow:

$$\begin{aligned} \theta_{i+1} &= \theta_i + \alpha \delta \theta \\ \mathbf{H}(\theta) \delta \theta &= -\mathbf{J}(\theta). \end{aligned}$$

By replacing the *Jacobian* and the *Hessian* by their approximate values, the recursion becomes

$$\begin{aligned} \theta_{i+1} &= \theta_i + \alpha \delta \theta \\ \frac{\partial \rho(r_i, \sigma)}{\partial r_i} \cdot \mathbf{A} \cdot \mathbf{A}^T \delta \theta &= \frac{\partial \rho(r_i, \sigma)}{\partial r_i} \cdot \mathbf{A} \cdot r_i. \end{aligned} \quad (22)$$

where only  $i^{th}$  equation of the equation system is shown here; the system has as many equations as there are pixels in the image overlap.  $\mathbf{A}$  is defined in Equation 21. Thus, at each iteration step, a linear system of equations has to be solved, by taking the minimum mean squared error solution (since the system has more equations than unknowns). Then, a *line search* is performed in order to determine the factor  $\alpha$  that minimises equation (13), by starting with  $\alpha = 1$ . In Equation 22 it is important not to factor out the  $\mathbf{A}$  and  $\frac{\partial \rho(r_i, \sigma)}{\partial r_i}$  terms, since it would affect the respective influence of each pixel in the equation system.

In other words, the terms involved in Equation 22 are:

- $\frac{\partial \rho(r_i, \sigma)}{\partial r_i}$  : The derivative of the objective function. In the case of a Minimum Mean Squared Error,  $\frac{\partial \rho(r_i, \sigma)}{\partial r_i} = \frac{2 \cdot r_i}{\sigma}$
- $\dot{S}_\theta [I_0(\mathbf{p})]$  : The derivative of the OECF inverse, computed for each pixel value of image  $I_0$  (in the overlap area of the two images). This term contains the derivative with respect to parameters  $a$  and  $b$  of Section 2.3. For the system using slides, this term is equal to 0.
- $\dot{G}_\theta [I_1(\mathbf{U}_{\theta i})]$  : The derivative of the OECF inverse, computed for each pixel value of image  $I_1$  (in the overlap area of the two images). This term contains the derivative with respect to parameters  $a$  and  $b$  and  $t$  of Section 2.3.
- $\nabla G_\theta [I_1(\mathbf{U}_{\theta i})]$  : The Gradient of the raw image 1, warped according to the motion parameter set. The Gradient is the derivative along the lines and columns of the image. The image is first warped and transformed into raw data prior to the derivative computation along the lines and columns.
- $r_i$  : The difference between the estimate of the radiance of each pixel of the scene that appear in both images, i.e. the difference between the raw data of image  $I_0$  and  $I_1$ .

### 4.3. Robustness of the estimation

Here, the choice of function  $\rho(\cdot)$  is discussed: Technically, this function modifies the influence that a particular pixel has in the alignment process. Its goal is to sort out the “good” pixels from the “bad” ones. There are two different kinds of “bad” pixels: The ones that contain moving elements in the scene, and the ones that are saturated or noisy. Being able to sort out pixels in the parts of the image that contain moving objects is a well known topic in motion estimation and we refer the reader to the robust motion estimation literature<sup>10,11,12</sup>. Here we are concerned with saturating and noisy pixels: A pixel that has a value of 255 on a 8 bits per sample camera contains only little information about the illuminance that hit the sensor at that location, because it will only give a lowerbound on the illuminance.

<sup>¶</sup>The approximation ensures that the matrix is positive definite.

Let  $i$  be a pixel value,  $X_i$  be the illuminance that generates the value  $i$  in the CCD and  $t$  be the exposure parameter of the picture. The exposure parameter determines the illuminance present at the camera sensor given the radiance power of the object being pictured.  $X_i$  is assumed to be a normally distributed random variable with variance  $t^2 \cdot \sigma_i^2$ :

$$X_i \sim N(t \cdot i, t^2 \cdot \sigma_i^2). \quad (23)$$

Now, let's assume that an object of the scene generates values  $i$  and  $j$  in two pictures taken with different exposure parameters  $t_1$  and  $t_2$ . The radiance  $X$  that this object may have can be computed using either one of the random variable associated to the pixel value:

$$\begin{aligned} X_i &= t_1 \cdot (i + n_1) \\ X_j &= t_2 \cdot (j + n_2) \end{aligned} \quad (24)$$

where  $n_1$  and  $n_2$  are the acquisition noise for the pixels in both images,  $N_1 \sim N(0, \sigma_i^2)$  and  $N_2 \sim N(0, \sigma_j^2)$ . let  $R_{ij}$  be a new random variable defined as

$$R_{ij} = X_i - X_j. \quad (25)$$

Since  $X_i$  and  $X_j$  are normally distributed, and by assuming that the noise terms in Equations (24) are uncorrelated, we get

$$R_{ij} \sim N(0, t_1^2 \cdot \sigma_i^2 + t_2^2 \cdot \sigma_j^2). \quad (26)$$

The  $R_{ij}$  is the random variable associated to the variable  $r_i$  of equation (14). Now, ideally,  $\sigma_i$  should be measured for each camera during a calibration process, but as an approximation, we can take an arbitrary function such that  $\frac{1}{\sigma_i}$  tends toward 0 for pixel values close to the maximum and minimum of what the camera can deliver (typically, a camera delivers values from 0 to 255).

To use this value in the motion estimation process, the value of  $\sigma_i$  has to be computed for each pixel pair, and the (weighted) sum  $\sigma_s$  of equation (26) is used in the weighting process in the following way:

$$\frac{\partial \rho(r_i, \sigma)}{\partial r_i} = \frac{\dot{\rho}\left(\frac{r_i}{\sigma_s}\right)}{\sigma_s} \quad (27)$$

where  $\dot{\rho}$  is determined by the choice of the error model for robust estimation. In the standard case of a Squared Error Measure,

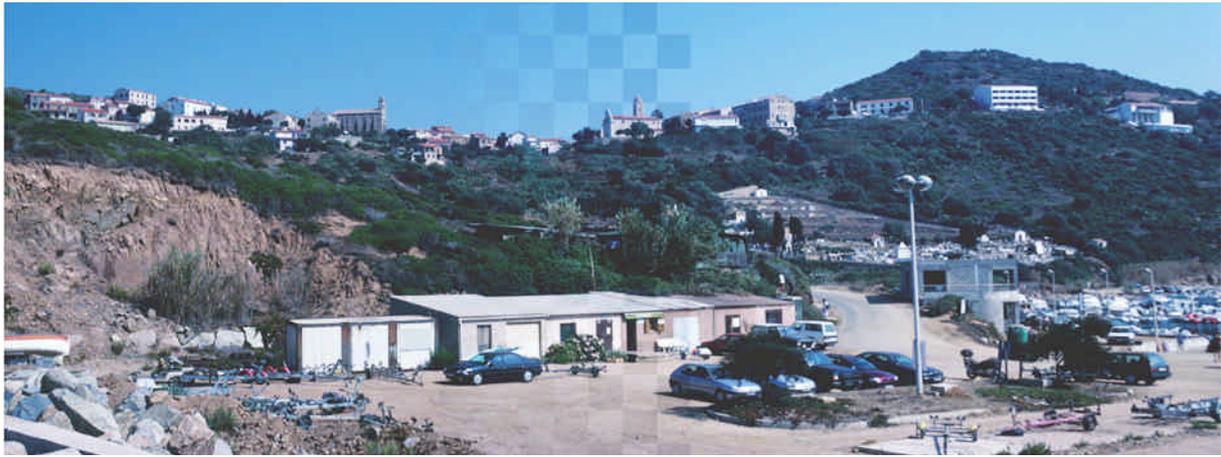
$$\frac{\partial \rho(r_i, \sigma)}{\partial r_i} = \frac{2 \cdot r_i}{\sigma_s^2}.$$

## 5. RESULTS

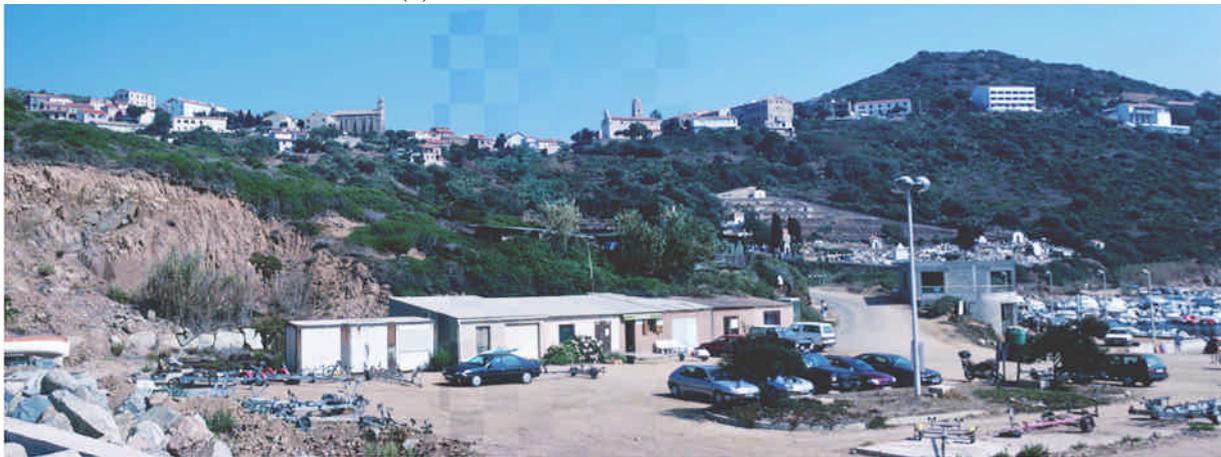
### 5.1. Stitching images from slides

In the first example, we use two pictures that have been scanned from slides. The slides are on Kodachrome 64 film, whose characteristics can be found on the manufacturer Web site. The slide characteristics is given by a D-LogH curve, which relates the film density to the log-exposure. Then, the slide is passed through a scanner whose OECF has been characterized by a prior characterization process, as presented in Section 2.2. The result of the mosaicking is shown in Figure 1. The picture have been taken using the same settings (exposure and focal length) for both pictures, and the resulting mosaic is shown using a chessboard-like blending technique: The overlap area can be considered as a chessboard; the black cells of the chessboard contain the pixels from the left image and the white cells of the chessboard contain the pixels from the right image. The chessboard-like blending is about the worst for the final representation and is solely aimed at emphasizing the mismatch in alignment and colours. Figure 1(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 mosaicking algorithm would do. Figure 1(b) shows the mosaic processed with colour correction, but without taking vignetting into account. The estimated photometric parameters are the exposure difference between the images, measured independently in the Red, Green and Blue channel. This accounts for the difference in white balancing and gain that may occur in the scanner<sup>||</sup>. Figure 1(c) shows the final mosaic, with vignetting correction. The only difference between 1(b) and 1(c) is the vignetting correction.

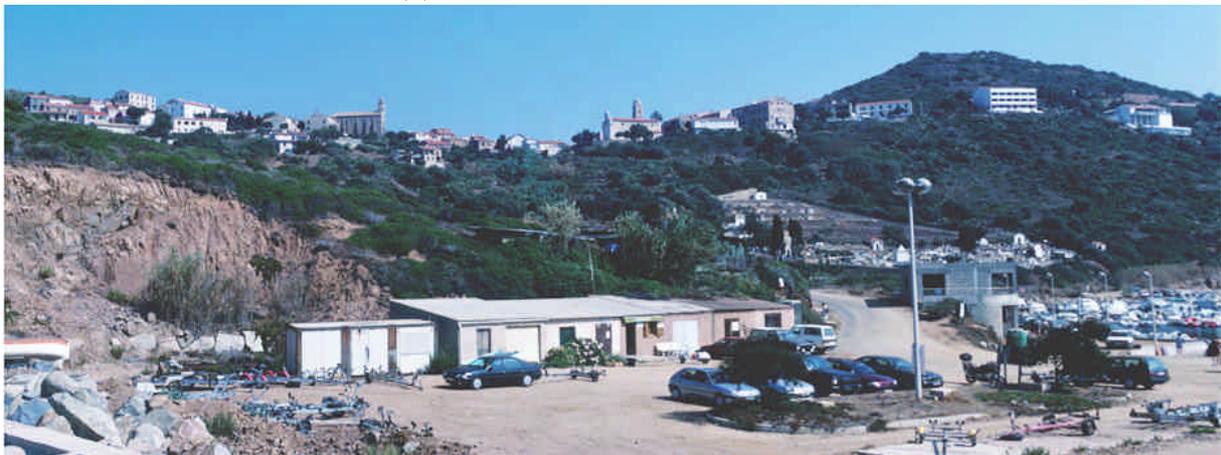
<sup>||</sup>It is not clear to us whether the change in exposure occurred in the scanner or because of the camera shutter imperfections.



(a)



(b)



(c)

**Figure 1.** Colour correction using slides with prior calibration. (a) Original mosaic (b) Exposure correction (c) Exposure and vignetting correction. To emphasize the colour mismatch, the images have been blended using a chessboard-like technique.

## 5.2. Stitching digital still camera images

Here, the images are taken from a digital still camera set on automatic exposure and automatic white balancing. The OECF of the camera is unknown, but the nominal shutter speed, aperture and focal length are known and are used as initial estimate (we assumed constant exposure and constant white balancing). The estimation proceeds in a coarse-to-fine manner: it starts by estimating the parameters on a low resolution version of the image, and then adds up the number of estimated parameters and increases the resolution of the image it uses. In practice, it starts by estimating only the camera rotation, given an estimate of the OECF (it uses the standard sRGB curve) and an estimate of the exposure. If there is no estimate for the exposure difference, the exposure difference is computed using the median of the error in the overlap area of the two image; the median being a measure that is quiet robust to mis-alignment. Then, once the images are roughly aligned, the exposure is added to the estimation parameters list. Once this has converged, the OECF (the 2 parameters of Section 2.3) is added to the parameter list. Finally, 3 exposure differences are estimated between the images - one for each colour channel - in order to account for a white balancing difference in the images. The result is shown in Figure 2. The first image is obtained using a traditional technique, without touching the colours. The corrected image is shown on Figure 2(b). On the second image, the system did a better focal length estimation. Indeed, a focal length is an ill-posed parameter in motion estimation, but by considering the vignetting effect, this parameter becomes easier to estimate (this occurs because, geometrically, a change in focal length causes a tiny change in the image overlap - provided that the rotation parameters are changed accordingly - whereas with vignetting correction this change becomes more important). The gains variation of the different colour channel (i.e. the white balancing) was of the order of 10%.

## 5.3. Discussion

If we compare the results obtained with the slides and with the digital cameras, we can say that it is better to use a prior calibration. The drawback of calibration is the effort and equipment needed to accomplish it. Also, the results might be slightly biased because of the respective quality of the images used in both experiments. Scanned slides provide images of excellent quality, whereas the images coming from the digital still camera were of very poor quality and contain a lot of noise. This noise is mainly quantization noise of the JPEG compression in the camera. Noise makes the estimation process much more difficult, and may cause the estimation not to converge towards the best solution.

Section 2.3 presented a particular form of an OECF by using 2 parameters. This function was used here because the traditional gamma correction of video is still used in many of the cameras today. In the ideal case, the only parameters that have to be estimated should be the one that may change from one picture to the other, an nothing more. The results of Section 4 are then applicable with any form of parametric function for the OECF.

Once the pictures have been transformed into “raw data” and stitched together, the rendering has been performed by scaling the image into the interval  $\{0, 1\}$  followed by a gamma correction. Now if two estimations resulted in a different set of exposure corrections, the dynamic range of the final picture is different, and so are the colours in the different mosaics. This explains why the colours are not the same in Figure 2(a) and 2(b). Rendering issues were left out in this work.

## 6. CONCLUSION

This paper presented a technique to correct for the colour mismatch between the pictures of a panoramic image. The principle is to compute the raw data for each image by estimating the settings of the camera, stitch the images in raw data space, and render the final mosaic as a single picture. The results show two examples: one of a calibrated system, where the only unknowns are the parameters that can vary during capture, and the other of a completely uncalibrated system. The result are quiet good for the calibrated system, but are less convincing for the uncalibrated one. Nevertheless, each of the method used outperforms the traditional approaches used in mosaicking nowadays.

## REFERENCES

1. M. D. Fairchild, *Color Appearance Models*, Addison, Wesley, Reading, MA, 1998.
2. 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.
3. ISO16067 : 1999, *Electronic Scanners for Photographic Images - Spatial Resolution Measurements, Part 1: Scanners for Reflective Media*.



(a)



(b)

**Figure 2.** Colour correction of an image taken with a digital still camera. (a) original image, no correction (b) Exposure, vignetting and white point correction.

4. R. Hunt, *Measuring Colour*, Fountain Press, England, 3 ed., 1998.
5. ISO 14524:1999, *Electronic still picture cameras- Methods for measuring opto-electronic conversion functions (OECFs)*, 1999.
6. ITU-R Recommendation BT.709-3, *Parameter Values for the HDTV Standards for Production and International Programme Exchange.*, 1998.
7. R. Kingslake, *Optical Systems Design*, New York, 1983.
8. B. K. Horn, *Robot Vision*, McGraw-Hill, New York, 1986.
9. G. Seber and C. Wild, *Nonlinear Regression*, John Wiley and Sons, New York, 1989.
10. P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*, Applied probability and statistics, John Wiley and Sons, Inc., New York, 1987.
11. H. S. Sawhney, S. Hsu, and R. Kumar, "Robust video mosaicing through topology inference and local to global alignment," in *ECCV98*, Springer, ed., pp. 103–119, (Freiburg, Germany), 1998.
12. S. Ayer, *Sequential and competitive Methods for Estimation of Multiple Motions*. PhD thesis, Swiss Federal Institute of Technology (EPFL), 1995.